

# Identification of Cell cycle Gene Signatures Predicting Survival in Patients with Lung Squamous Cell Carcinoma

Lei Zhang

First Affiliated Hospital of China Medical University

Zhe Zhang

First Affiliated Hospital of China Medical University

Zhenglun Yu (✉ [alan5915@126.com](mailto:alan5915@126.com))

China Medical University Hospital

---

## Research article

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

**Posted Date:** November 14th, 2019

**DOI:** <https://doi.org/10.21203/rs.2.17290/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** Lung cancer (LC) is one of the most important and common malignant tumors, and its incidence and mortality are increasing annually. Lung squamous cell carcinoma (LUSC) is the common pathological type of lung cancer. A small part of biomarkers have been confirmed to be related to the prognosis and survival by data excavation. However, the moderate forecast effect of a single gene biomarker is not accurate. Thus, we aimed to identify new gene signatures to better predict Lung squamous cell carcinoma (LUSC). **Methods :** Using the mRNA-mining approach, we performed mRNA expression profiling in large lung squamous cell carcinoma cohorts (n= from The Cancer Genome Atlas (TCGA) database. Gene Set Enrichment Analysis (GSEA) and Gene Set Variation Analysis (GSVA) were accomplished, and connections between genes and cell cycle were found in the Cox proportional regression model. **Results :** We have confirmed a set of four genes (CDKN1A, CHEK2, E2F4 and RAD21) that were importantly associated with overall survival (OS) in the test series. Based on the research of the four-gene signature, the patients in the test series could be divided into high-risk and low-risk teams. Additionally, multivariate Cox regression analysis revealed that the prognostic power of the four-gene signature is independent of the clinical factors. **Conclusion :** Our study demonstrated the connections between the four-gene signature and cell cycle. Novel insights into the research mechanisms of cell cycle was also revealed regarding the biomarkers of a poor prognosis for lung squamous cell carcinoma patients.

## Background

In recent years, the incidence and mortality rates of lung cancer are the highest all over the world, and it is the leading cause of cancer 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 is the one type of Non-small cell lung cancer which accounts for approximately 25%–30% of all lung cancer patients [3]. Although the incidence of LUSC can be controlled in a way by restricting tobacco smoking [4]. The identification of targeted treatment strategies for LUSC has remained a complicated goal [5]. An adding amount of evidence demonstrates that the detection and adhibition of molecular biomarkers will improve the prognostic evaluation and identification of potential high-risk patients with LUSC. Therefore we should find more biomarkers which affect the mechanism of tumorigenesis and progression of LUSC. So, the development of novel biomarkers which can predict the prognosis value for LUSC is of great significance. Recent studies have provided insights into the mechanisms that contribute to DNA repair in specific cell-cycle phases and have highlighted the mechanisms that ensure cell-cycle progression or arrest in normal and cancerous cells. Some gene Bind to and inhibits cyclin-dependent kinase activity, preventing phosphorylation of critical cyclin-dependent kinase substrates and blocking cell cycle progression and promote 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], which have been considered new biomarkers of cancer with infinite clinical value [7]. Scientists researched and detected many relating genes, and opened up a new

approach for the further research and effective treatment method. Some biomarkers have been confirmed that they are related to the prognostic factor of lung squamous cell carcinoma. miR-324-3p and miR-1285 as diagnostic and prognostic biomarkers for early stage lung squamous cell carcinoma[8].

Minichromosome maintenance protein (MCM) has been found to be a relevant marker for prognosis in many human cancers, which may act as a potential therapeutic target for LUSC patients[9]. With the development of high throughput sequencing and magnificent data analysis techniques, numerous databases have enabled detailed understanding of genomic changes. Disinterment of cancer related genes and the identification of gather models can lean on bioinformatics and data integration[10]. However, a single gene biomarker cannot produce good prediction effects, and some studies have found that the gene signature can be a better alternative to predict the prognosis and survival [11]. The method of the Kaplan-Meier plots and Cox proportional hazards regression have been applied diffusely in biomarker screening[12,13]. Multigene prognostic signatures can be used in clinical treatment because of the original cancer biopsies. However, not all pathways have been explored to identify new biomarkers of LUSC. Therefore, more efforts are needed to find more sensitive biomarkers of LUSC.

The major public project "The Cancer Genome Atlas" (TCGA) [14] can collect, select and analyze human tissue for genome changes effectively. The TCGA database provides analysis of high-throughput data for varieties of genomic changes, including exchange of cell cycle. The information of patient diagnostics, treatment, and tumor pathology can be combined to the bioinformatics analysis. We used Gene Set Enrichment Analysis (GSEA) and Gene Set Variation Analysis (GSVA) to search for some genes and perform further analysis. The priority of GSEA does not require a distinct differential gene threshold. The algorithm is also special such that several genes can be identified whose expression is based on the entire trend of effective data and overall level, even without any prior experience. The priority of GSVA were performed to explore the biological function and functional annotation of the differences. We conducted various of bioinformatics and survival analyses, screened for one pathway associated with cell cycle, and built a prognostic risk model to predict the prognosis of LUSC patients. The results of analysis could bring new insight into the molecular mechanisms based on cell cycle in LUSC.

In our study, we have drawn hallmark gene sets from lung squamous cell carcinoma cases with completed mRNA expression data set from the TCGA database. We have confirmed the key mRNAs related to cell cycle and have built a four-gene risks signature that can accurately predict the patients' prognosis results. Surprisingly, in fifteen pathways, the cell cycle-related risk signature can predict patients who are in the high-risk group with a poor prognosis.

## Methods

### Patient Clinical Data and mRNA Expression Dataset

The mRNA expression profiles and clinical data concerning lung squamous cell carcinoma patients were extracted from the TCGA database (<https://cancergenome.nih.gov/>) and were divided into two groups, one with lung squamous cell carcinoma tissue and one with adjacent noncancerous tissue. Additionally,

the following clinical information was recorded: gender, age, size of tumor, pathological stage, new tumor events after the initial treatment, status of distant organ metastasis, neoplasm cancer status, and residual tumor. Finally, patients were classified. The general clinical features are detailed in Table 1.

## Gene Set Enrichment Analysis and Gene Set Variation Analysis

GSEA (<http://www.broadinstitute.org/gsea/index.jsp>) was performed to explore whether the identified gene sets showed significant differences between the groups [15]. The expression levels of mRNAs in lung squamous cell carcinoma samples and adjacent noncancerous tissue were analyzed. Normalized p values ( $p < 0.05$ ) were used to determine which functions can be used for further investigation. Gene Set Variation Analysis (GSVA), which was a GSE method that evaluates variation of pathway activity over a specimen population in an unsupervised method. We showed the robustness of GSVA in a comparison with present state of the art sample-wise enrichment technique.

## Statistical Analysis

The expression profiles of 24,991 mRNAs were shown as raw data, and each mRNA was normalized by log<sub>2</sub> for further analysis. We used univariate Cox regression to analyze and identify genes with obvious relationships to overall survival with p values  $< 0.05$ . Next, we used multivariate Cox proportional hazards regression to analyze and further confirm the prognostic genes from the previous steps. The filtered mRNAs were classified into risk (hazard ratio (HR)  $> 1$ ) and protective ( $0 < HR < 1$ ) types. Thereafter, a prognostic risky 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. Risky score = expression of gene 1  $\times \beta_1$  + expression of gene 2  $\times \beta_2$  + ... + expression of gene n  $\times \beta_n$ . We separated patients into high-risk and low-risk subgroups by the median value using the median risky score as the cutoff. Next, we used Kaplan-Meier curves and log-rank methods to validate the prognostic importance of the risky score. Subsequently, we examined the differential expression of optimal genes in lung squamous cell carcinoma samples and adjacent noncancerous tissue. We classified them into high-risk and low-risk teams by the median risk score and used the KM method (multiplication of the positive limit) to predict the accuracy of the survival state and survival time. The survival function was constructed by the KM method, and the ROC curve was drawn. The chosen gene alterations were shown in specific cancer types online (<http://www.cbioportal.org/>). All the statistical analyses were performed using SPSS 16.0 and Graph Pad Prism8 software.

# Results

## Initial Screening of Genes using GSEA and GSVA

We obtained the clinical features from patients with lung squamous cell carcinoma, along with an expression data set for 56392 mRNAs from the TCGA database. The KEGG gene sets had expressed signatures derived by concentrating multiple gene sets from the Molecular Signatures Database (MSigDB) to represent well-defined biological statuses or courses. GSEA was performed using the above

detailed data to detect whether the identified gene sets showed statistically important differences between normal tissue and LUSC tissues patients. GSEA estimates variation of gene set enrichment over the samples independently of any class label. therefore, we got fifteen pathway (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 signaling pathway, Basal transcription factors, RNA degradation, RNA polymerase and Oocyte meiosis, and we selected cell cycle pathway based on the number of gene and the value of NES.

### **Identification of Cell cycle-related mRNAs Associated with the Survival of Patients**

First, we employed univariate Cox regression analysis of the 125 genes for preliminary screening and obtained 20 genes with p values <0.1. Additionally, multivariate Cox regression analysis was used to further examine the relationship between the expression profiles of 20 mRNAs and the patient survival rate. Subsequently, 4 mRNAs (CDKN1A, CHEK2, E2F4 and RAD21) were verified as independent poor prognostic indicators. The filtered mRNAs were classified into a risky 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 made Pearson correlation coefficient among the 4 mRNAs on the basis of Table 2, and we found the correlation between E2F4 and CHEK2, between E2F4 and RAD21, between CHEK2 and RAD21, and the value of R is greater than 0.3 (Fig.2).

### **Construction of a four-mRNA Signature to Predict Patients' Result**

A prognostic risky score formula was established based on a linear combination of the expression levels weighted with the regression coefficients derived from multivariate Cox regression analysis. The risky score =  $0.2551 \times \text{expression of HMMR} + 0.2160 \times \text{expression of B4GALT1} - 0.1570 \times \text{expression of SLC16A3} + 0.1238 \times \text{expression of ANGPTL4} + 0.2381 \times \text{expression of EXT1} + 0.1027 \times \text{expression of GPC1} + 0.1820 \times \text{expression of RBCK1} + 0.1874 \times \text{expression of SOD1} + 0.2226 \times \text{expression of AGRN}$ . Each patient of LUSC had only one risky score. We calculated the scores and ranked them, and then classified the patients into high- and low-risk groups by the median value (Fig.3A). The survival length of 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. Additionally, a heatmap (Fig.3C) was revealed to display the expression profiles of the four mRNAs. Then, we compared the risk score to the prognosis of the 4-mRNA, and it was showed that the different expression of four gene between cancer tissue and adjacent normal tissue in Fig.4A, and the different expression of four gene in each stage was displayed in Figure 4B, and the graph of survival curve of between risk score and four gene were showed in Figure 4C, therefore, we can see that the value of P of risk score is dominant. With the increasing risk score of patients with LUSC, the expression of high-risk types of mRNAs (CDKN1A, E2F4 and RAD21) was obviously upregulated. By contrast, the expression of the protective type of mRNAs (CHEK2) was downregulated.

### **Generation of Risk Score from the Four mRNA Signatures as an Indicator of Prognosis**

The prognostic values of the risk scores were compared with the clinicopathological information by univariate and multivariate analyses. Samples with completed clinical data were used for analysis. The median age of the 504 patients with lung squamous cell carcinoma was 68 years and included 373 male patients and 131 female patients. Among 391 patients, 79 (20.2%) had a positive tumor during the follow-up visit. Among 439 patients, 41 (9.3%) had residual tumors. Among 503 patients, 184 (36.6%) patients had lymph node metastasis and 86 (17.3%) had distant metastases among 497 patients with lung squamous cell carcinoma. Among 154 patients, 15 (9.7%) received radiation therapy. Additionally, we found that the risk score, new event, tobacco smoking history, and neoplasm cancer status were independent prognostic indicators because they showed important differences in univariate analysis with  $p$  values  $< 0.05$  (Table 3). In the subsequent multivariate analysis (Table 3), we found that the risk score, new event, neoplasm cancer status and tobacco smoking history showed statistical significance in univariate and multivariate analyses ( $P < 0.05$ ). Whether univariate or multivariate analysis, the risky score had prominent prognostic values, with  $p$  values  $< 0.05$  (HR = 1.566, 95% CI (confidence interval) = 1.073-2.288). Additionally, the most obvious clinical parameter to predict patient survival was "neoplasm cancer status", and patients with tumors are probably 4.871 times more likely to be exposed to death than those who were tumor free. From the value of  $P$ , we can draw that the risk score is more dominant than TNM classification. And the 4-mRNA expression-based survival risk score was used to assign patients into a low-risk or high-risk group using the median risk score as the cut-off. The ROC curve analysis score was 0.661 (Fig. 5A), indicating good sensitivity and specificity of the 4-mRNA signature in predicting survival in LUSC. We also made ROC curve of important clinical parameter (Fig. 5B-H) and found ROC curve of risky score was obviously higher than ROC curve of other clinical parameter, and superior to clinical parameter for prognosis indicator.

### **Validation of Four mRNA Markers for Survival Prediction by Kaplan-Meier Curve Analysis**

Kaplan-Meier curves and the log-rank method showed a poor prognosis in patients with high-risk scores ( $p < 0.0010$ ) (Fig. 6A). Univariate Cox regression analysis of the overall survival showed that several clinicopathological data were effective in predicting the survival rate of lung squamous cell carcinoma, including Age, Sex, T classification, N classification, M classification, new event, neoplasm cancer status, tobacco smoking history, radiation therapy and residual tumor. The K-M method was then adopted to confirm the above results. According to the curve, patients with age older than sixty-eight, with tumor after treatment, T classification greater than T1, distant organ metastasis, stage greater than stage I, a residual tumor, tobacco smoking history, or a positive tumor finding during the follow-up visit were correlated with poor prognosis (Fig. 6B, 6C, 6E, 6G, 6K, 6M, 6N, 6O). These results provided further confirmation of the accuracy of our analysis. Hence, further stratified analysis was performed for data mining. And the risk score is superior to other clinical indicator from results.

### **Validation of differential gene between high risk team and low risk team**

We classified the patients of LUSC into two groups followed by risk score, and enriched them into pathway of high risk team and low risk team by GSEA and GSVA. And got differential gene, then put them into

mass survival and got related genes. We made Pearson correlation coefficient between related genes and four genes, and found the correlation between CDKN1A and KLK5, between CDKN1A and KLK7, and the value of R is greater than 0.3 (Figure 7B).

## Discussion

In recent years, some researchers have verified miRNAs could play an important role in tumorigenesis and progression of lung cancer. Accumulating evidence reveals the potential of miRNAs as biomarkers for evaluating and predicting the prognosis for lung cancer [16], indicating their considerable clinical significance in research [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 drew that *MiR-421 is Overexpressed and Promotes Cell Proliferation in Non-Small Cell Lung Cancer* [19]. However, patient survival could not be predicted by these genes because various factors can modulate a single gene, resulting in an inaccurate predictive effect. Thus, a gene signature comprising various genes has been built from the statistical model to predict cancer outcomes. The results highlight that the predictive effects of each gene involved can offer a more accurate prediction than a single biomarker [20].

With the development of gene signatures, the prognosis of some types of cancer has been predicted by statistical models. MiR-98-5p, miR-152-3p, miR-326 and miR-4289 Expression could serve as Biomarker for Prostate Cancer Diagnosis [21]. We also performed GSEA analysis using the mRNA expression data of the 503 LUSC patients and found that 4 exhibit significant differences, with a P value <0.05, and the minimum P value could be used for further analysis. We explored specific functions to identify genes by GSEA that could predict the survival of LUAD patients. Furthermore, we identified a combination of 4 genes with prognostic value for LUAD patients instead of a single gene by univariate Cox regression and multivariate Cox regression analyses. Subsequently, through comparison with some known prognostic biomarkers, we found that our identified risk signature may strongly support clinical results. We analyzed cell-related genes using the LUSC data set in TCGA and then compared one with lung squamous cell carcinoma tissue to one with adjacent noncancerous tissue data of LUAD patients. Kaplan-Meier analysis showed that a high-risk score was associated with metastasis and a poor prognosis. Among our 4 genes, 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 were significantly associated with unfavourable OS in all gastric cancer patients* [23]. *RAD21* were also increased in bladder cancer tissues and cell lines, and it could be used as potential therapeutic targets in bladder cancer [24]. Conventional prognostic systems generally make inaccurate predictions for risk stratification and estimations of clinical outcomes because of the heterogeneity between patients. To the best of our knowledge, compared with a single common biomarker, the 4-mRNA signature can better predict the prognosis of LUSC.

The 4 genes were found to be correlated with cell cycle. CDKN1A is a protein, which played multiple roles not only in the DNA damage response, but also in many cellular processes during cell growth. The

function of CDKN1A is to arrest cell cycle progression by inhibiting the movement of cyclin-dependent kinases[25].And it is found to participate in other significant processes, such as DNA replication/repair,cell death,gene transcription, and cell motility[26]. CDKN1A can inhibit the CDK activity, by interacting with their N-terminal domain or by interfering with the phosphorylation of CDK1 and CDK2[27]. Moreover, it contributes to the G1 arrest, by inhibiting the cyclin E and cyclin A/CDK2 activity[28].CHEK2 is Serine/threonine-protein kinase which is required for checkpoint-mediated cell cycle arrest, activation of DNA repair and apoptosis in response to the presence of DNA double-strand breaks. CHEK2 then reacts with 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. And CHEK2 regulates apoptosis through the phosphorylation of p53/TP53, MDM4 and PML.E2F4 is transcription activator that binds DNA cooperatively with DP proteins through the E2 recognition site, 5'-TTTC[CG]CGC-3' found in the promoter region of a number of genes whose products are involved in cell cycle regulation or in DNA replication.RAD21 is involved in sister chromatid cohesion from the time of DNA replication in S phase to their segregation in mitosis, a function that is essential for proper chromosome segregation, post-replicative DNA repair, and the prevention of inappropriate recombination between repetitive regions.It acts as a target gene for multiple genes or as a regulatory gene for other genes to participate extensively in the physiological and pathological processes of the human body[30-32]. RAD21 is reported that expressed highly in breast cancer'squamous cell carcinoma of the oral cavity and hepatocellular carcinoma tissuesand promotes the proliferation and metastasis[33-35].By Pearson correlation analysis,we also find related gene KLK5 and KLK7 ,which is correlate to CDKN1A,expressed in lung cancer[36].Therefore,we can see the relation between the 4 genes and cell cycle, and the cell cycle arrest and apoptosis induced the progression of lung cancer.

In conclusion, this work is the first to report a four-gene risk signature related to cell cycle that can help predict survival and prognosis in LUSC patients. A higher risk score indicates a poorer prognosis. This finding will help future researchers in their efforts to identify new treatments for LUSC and to provide more gene targets to cure LUSC in patients.

## Conclusion

We used a four-gene signature (CDKN1A, CHEK2, E2F4 and RAD21) to predict and evaluate LUSC via tissue or blood samples and examined whether mutations in these genes can promote the development of LUSC.Furthermore, we identified treatments related to cell circle to successfully target these genes; this signature could also be used to develop new targeted treatments to cure LUSC patients. Finally, we confirmed the relationship between the four-gene risk signature related to cell circle and LUSC.

## 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**

### **Authors' contributions**

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

### **Acknowledgments**

We thank all the participants who were involved in this study and Giulia Stella for providing editorial assistance.

### **Funding**

Not applicable

### **Availability of data and materials**

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

### **Consent for publication**

Not applicable.

### **Ethics approval and consent to participate**

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.

### **Author details**

<sup>1</sup> Department of Breast Surgery, The First Hospital Affiliated China Medical University, No.

155 Nanjing Street, Heping District, Shenyang, Liaoning Province, 110001, China;

<sup>2</sup> Department of Thoracic Surgery, The First Hospital Affiliated China Medical University,

No. 155 Nanjing Street, Heping District, Shenyang, Liaoning Province, 110001, China.

## 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. 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, et al. 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. *Proc Natl Acad Sci U S A*. 2005;102:15545-50. <https://doi.org/10.1073/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/10.1111/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 and neck 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 J, Schubert 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, et al. DNA topoisomerase II $\alpha$  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](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). <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, Reyat 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>.

## Tables

**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 <sub>x</sub>	6	1.1
M classification		
M0	411	82.7
M1	7	1.4
M <sub>x</sub>	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

**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%CI)	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

**Table 3** Univariable and multivariable analyses for each clinical feature

Clinical feature	Univariate analysis			Multivariate analysis		
	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						
New event	3.485	2.493-4.871	0.000	2.199	1.364-3.547	0.001
Radiation therapy	2.231	1.700-2.926	0.000	1.680	1.056-2.673	0.029
Residual tumor	1.875	0.914-3.848	0.087			
	1.884	1.120-3.168	0.017	1.022	0.486-2.148	0.955

## Figures

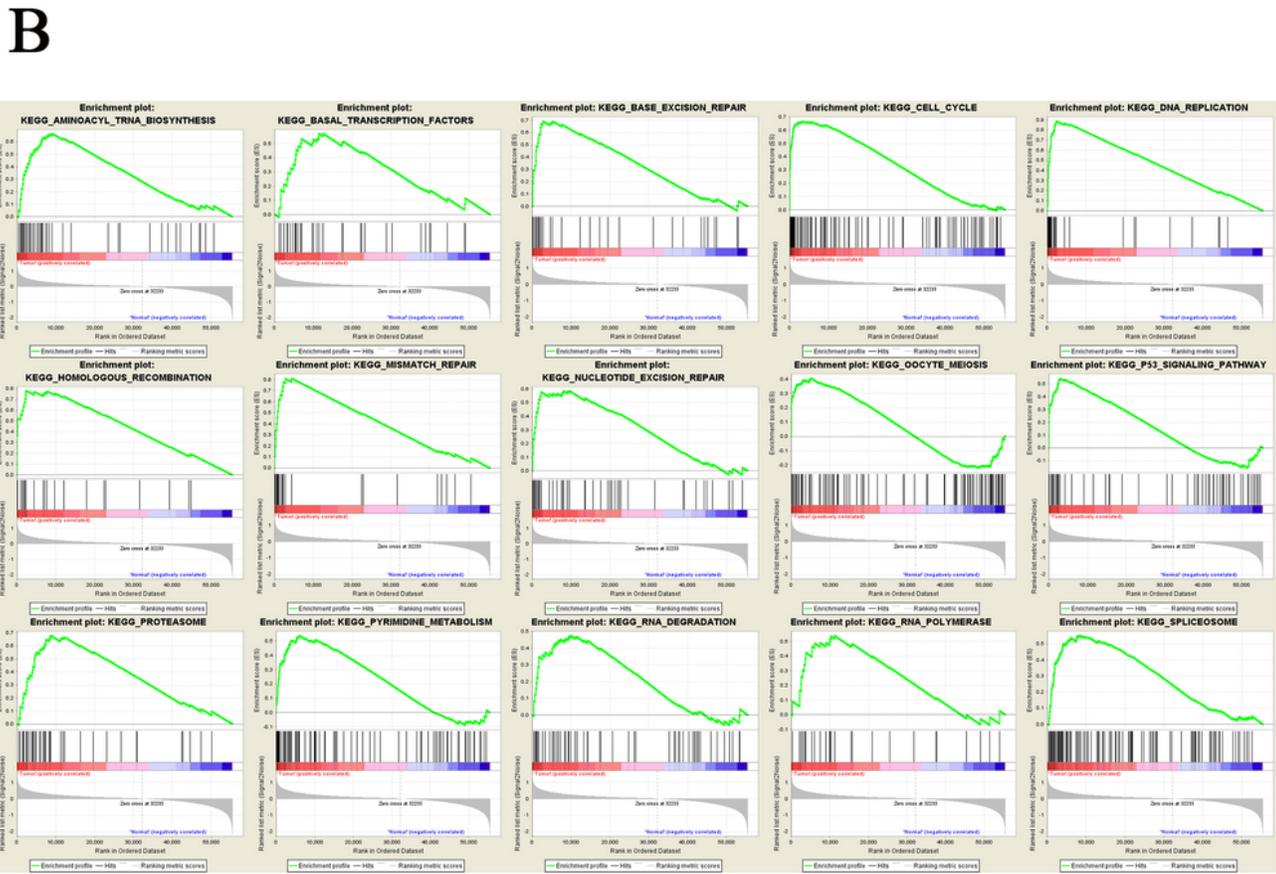
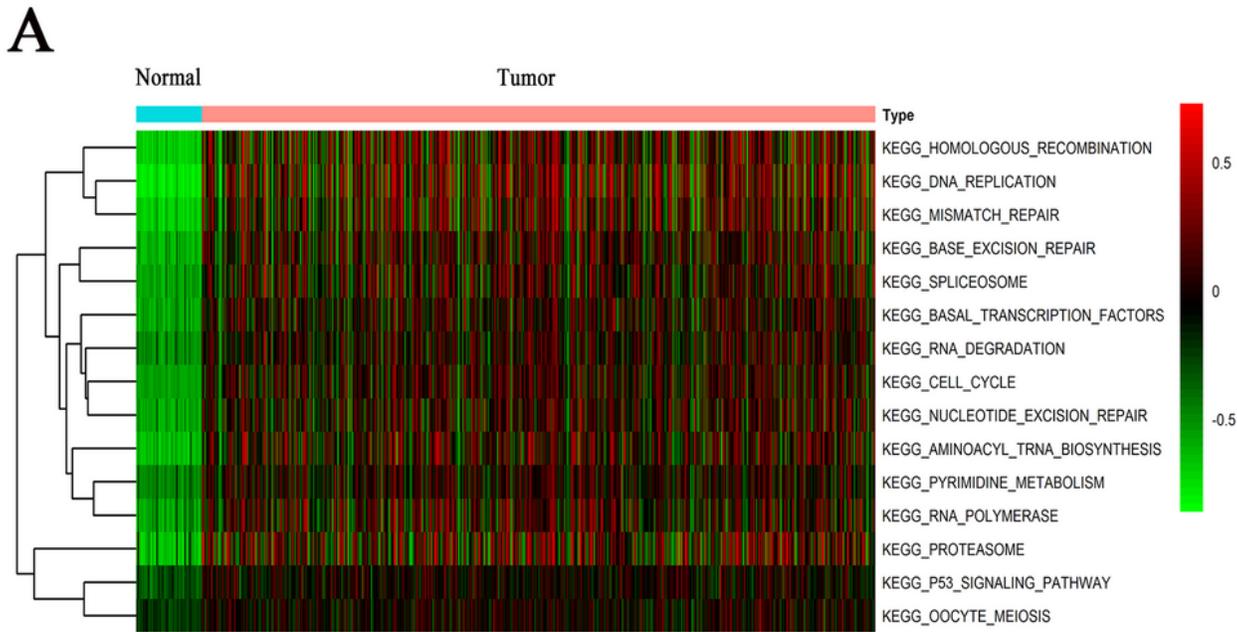


Figure 1

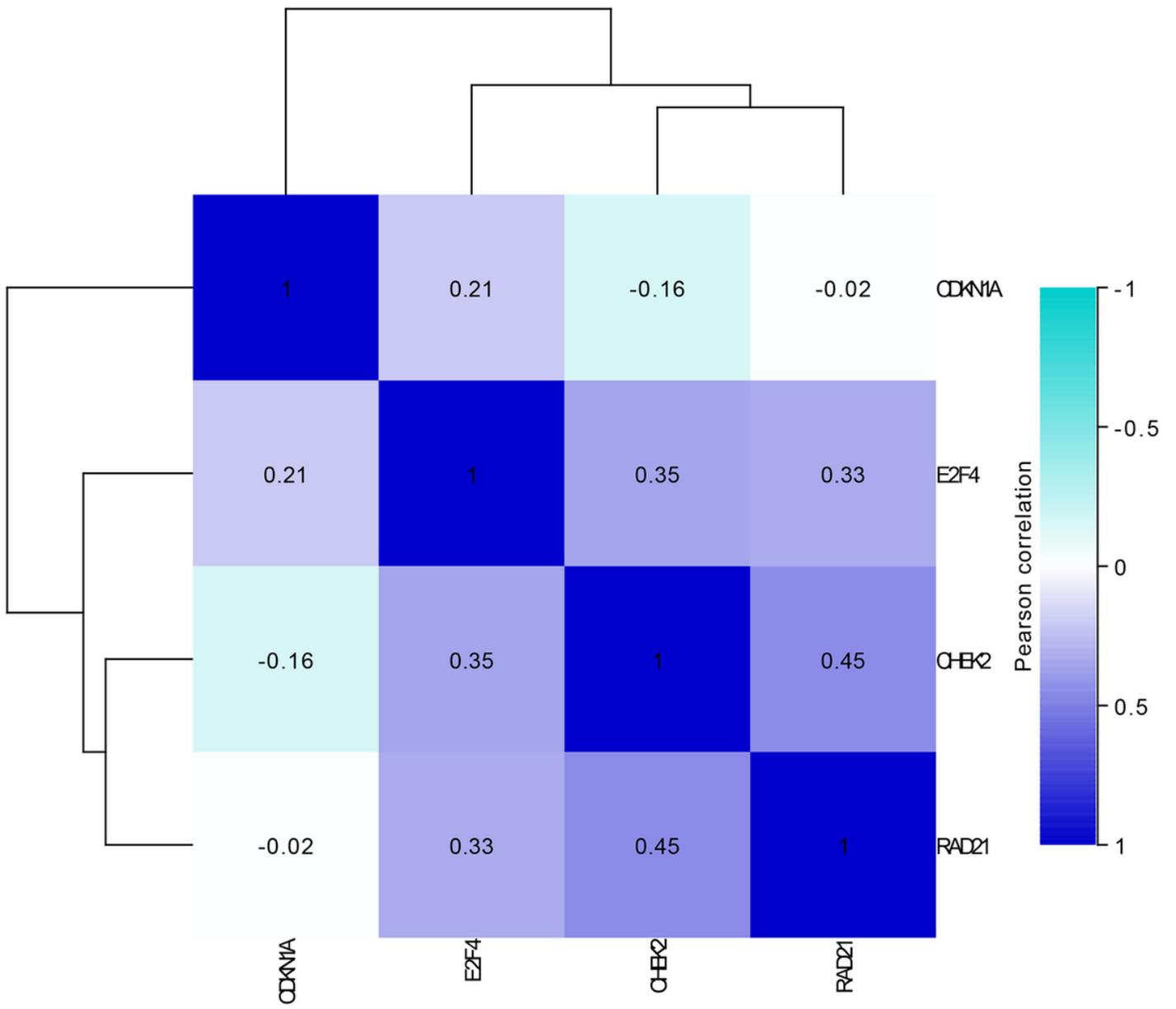
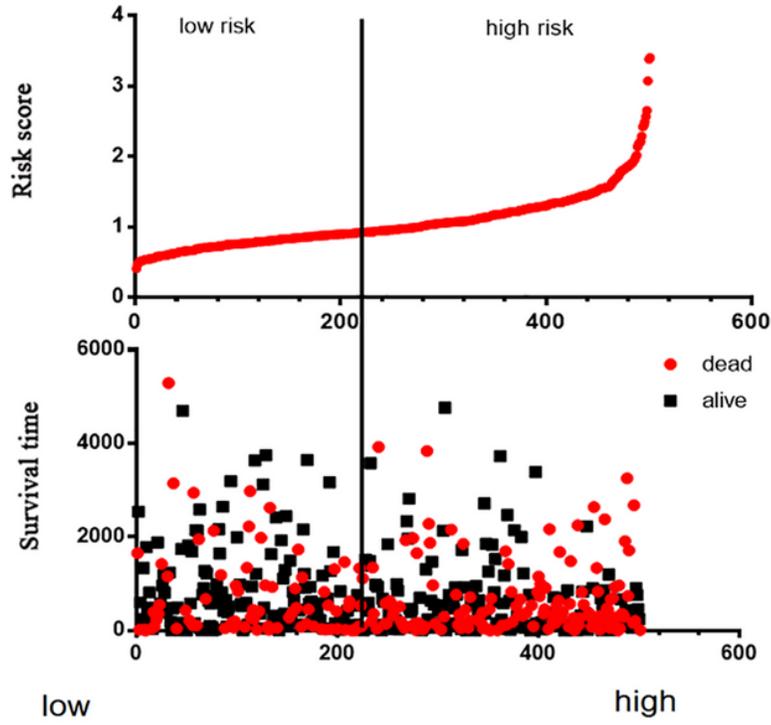
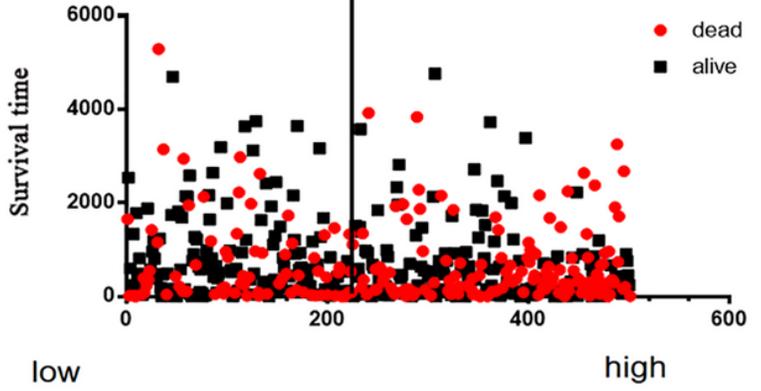


Figure 2

A



B



C

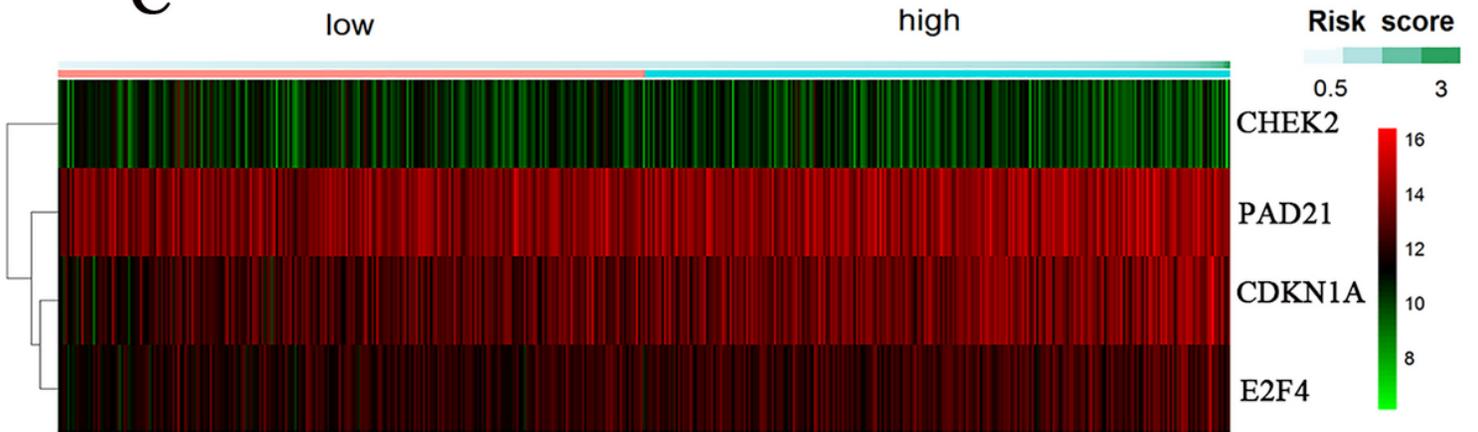


Figure 3

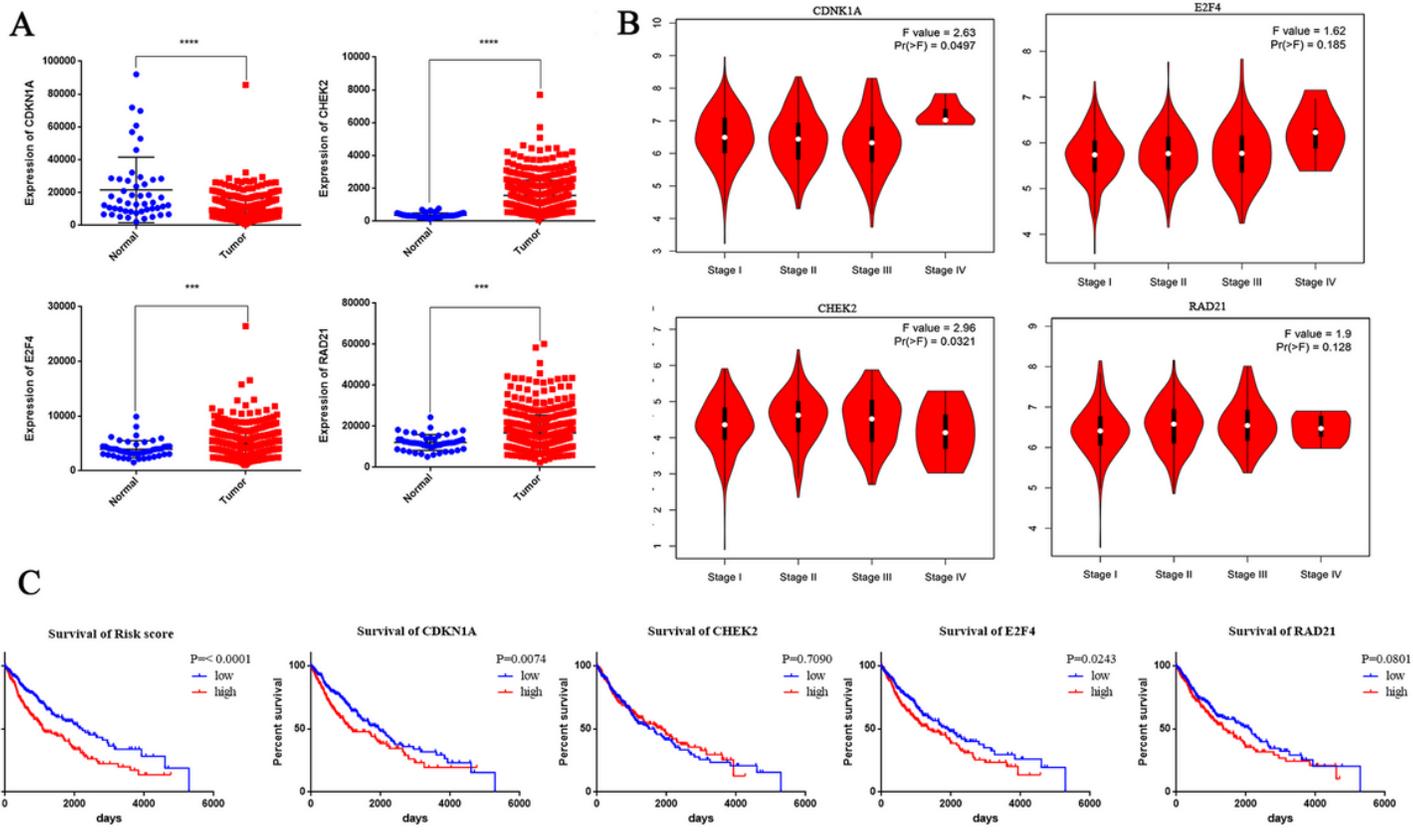


Figure 4

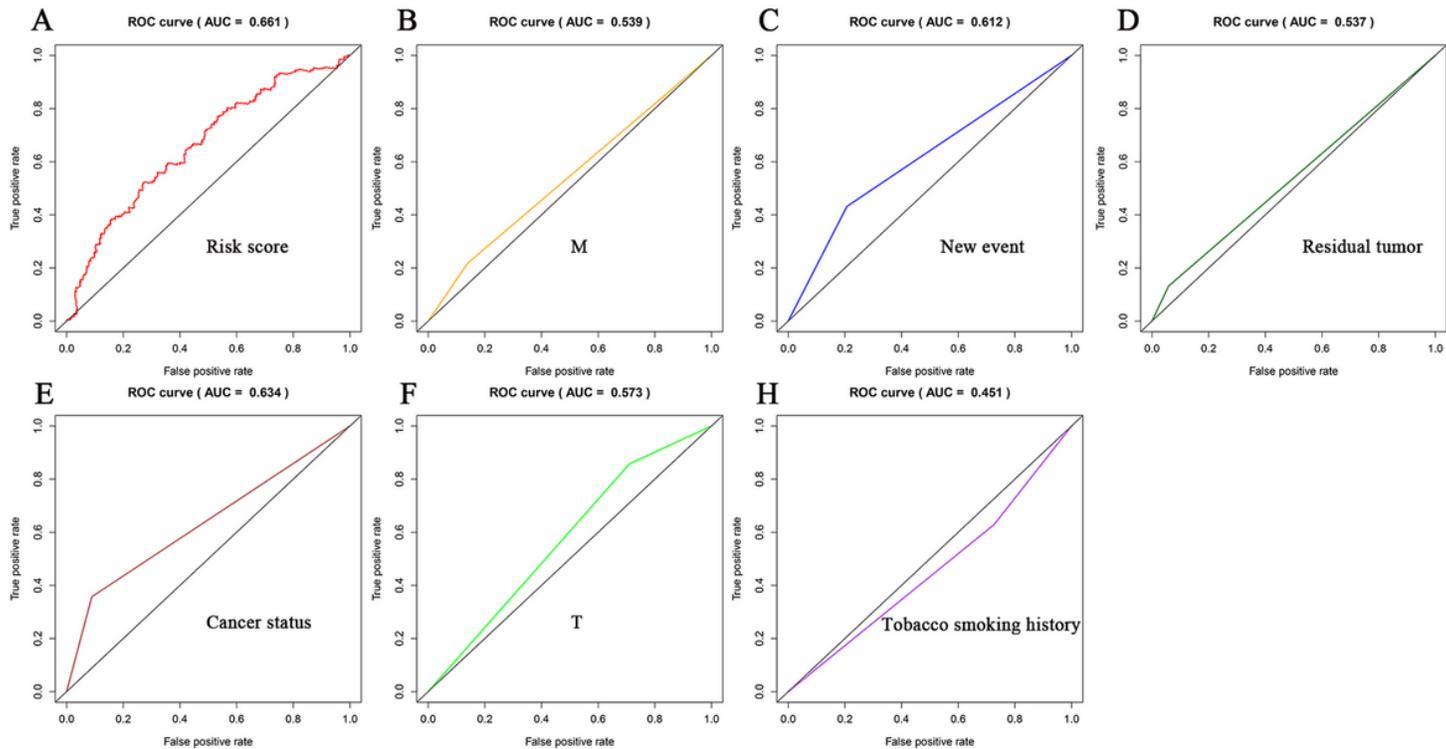


Figure 5

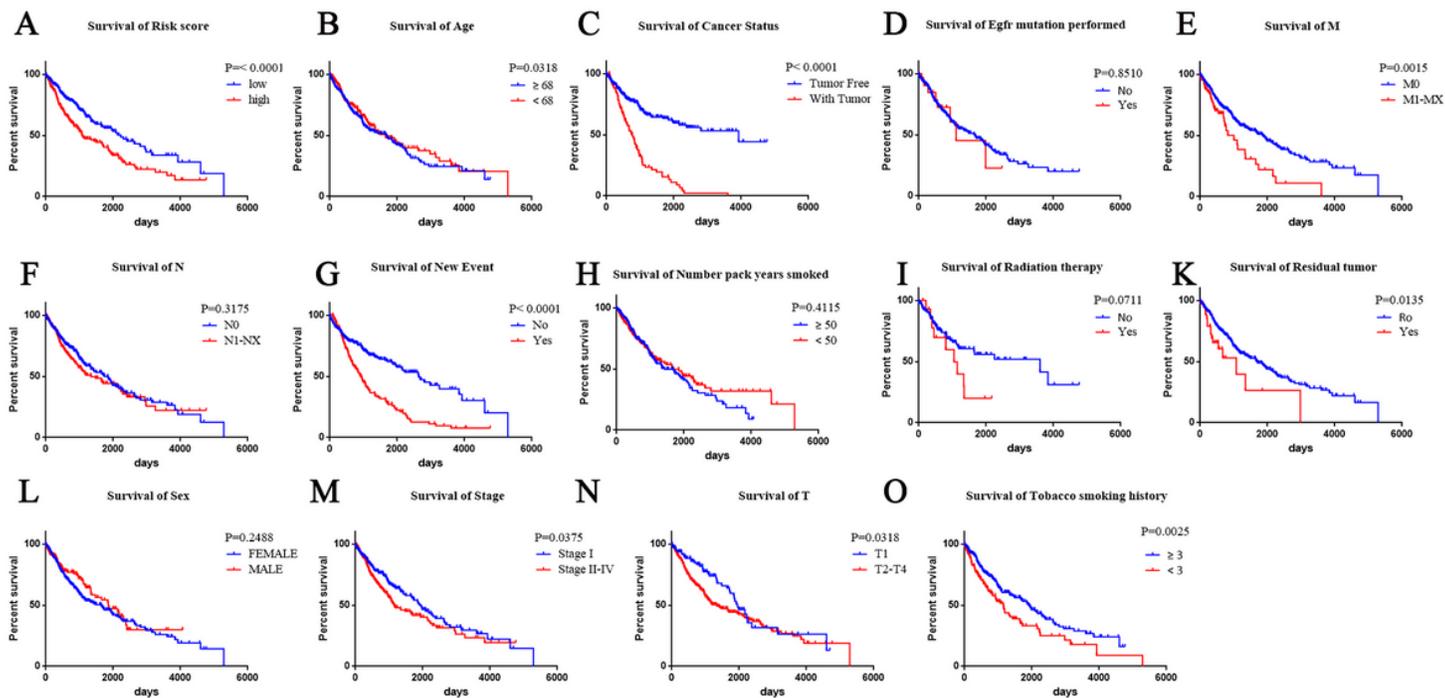


Figure 6

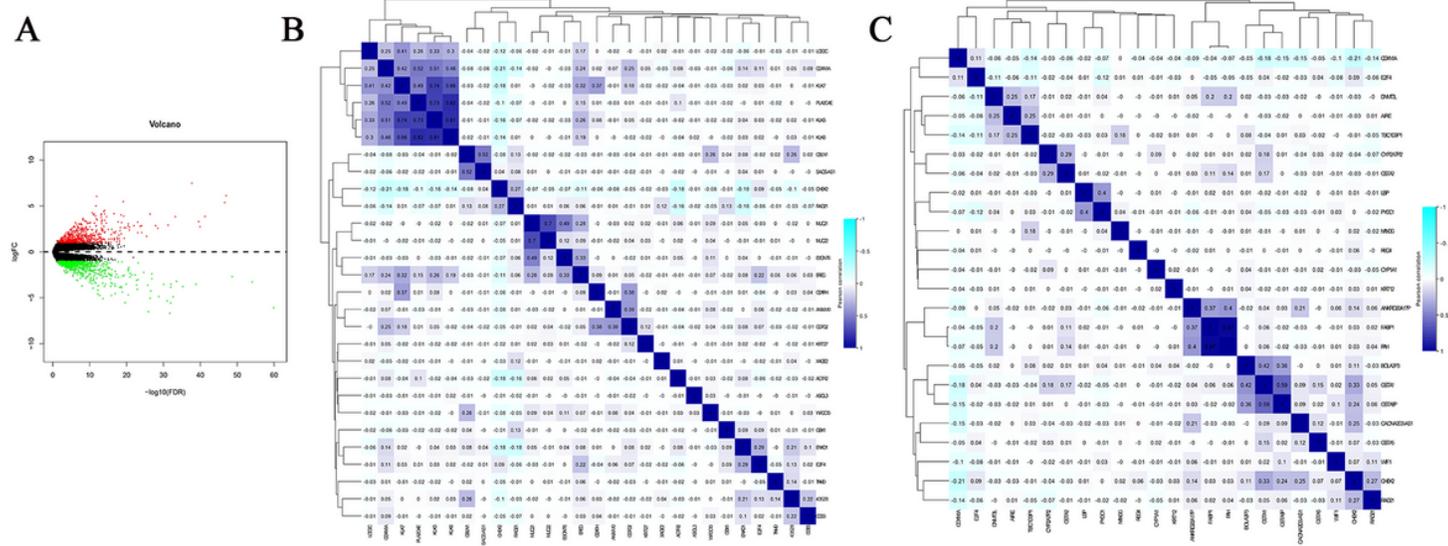


Figure 7