

80526\_Auto\_Edited.docx

**Name of Journal:** *World Journal of Clinical Oncology*

**Manuscript NO:** 80526

**Manuscript Type:** ORIGINAL ARTICLE

### *Basic Study*

4

## **A 5-mRNA-based prognostic signature of survival in lung adenocarcinoma**

Qianlin Xia, Xiaomeng He, Yan Ma, Qiuyue Li, Yuzhen Du, Jin Wang

4

### **Abstract**

#### **BACKGROUND**

Lung adenocarcinoma (LUAD) is the most common non-small-cell lung cancer, with a high incidence and a poor prognosis.

#### **AIM**

To construct effective predictive models to evaluate the prognosis of LUAD patients.

#### **METHODS**

In this study, we thoroughly mined LUAD genomic data from GEO (GSE43458, GSE32863, and GSE27262) and TCGA datasets, including 698 LUAD and 172 healthy (or adjacent normal) lung tissue samples. Univariate regression and LASSO regression analyses were used to screen DEGs related to patient prognosis, and multivariate Cox regression analysis was applied to establish the risk score equation and construct the survival prognosis model. Receiver operating characteristic (ROC) curve and Kaplan-Meier (KM) survival analyses with clinically independent prognostic parameters were performed to verify the predictive power of the model and further establish a prognostic nomogram.

2

### **RESULTS**

A total of 380 DEGs were identified in LUAD tissues through GEO and TCGA datasets, and 5 DEGs (TCN1, CENPF, MAOB, CRTAC1 and PLEK2) were screened out by multivariate Cox regression analysis, indicating that the prognostic risk model could be used as an independent prognostic factor (HR = 1.520,  $P < 0.001$ ). Internal and external validation of the model confirmed that the prediction model had good sensitivity and specificity (AUC = 0.754, 0.737). Combining genetic models and clinical prognostic factors, nomograms can also predict overall survival more effectively.

## CONCLUSION

A 5-mRNA-based model was constructed to predict the prognosis of lung adenocarcinoma, which may provide clinicians with reliable prognostic assessment tools and help clinical treatment decisions.

## INTRODUCTION

Lung adenocarcinoma (LUAD) is a common histological type of lung cancer that is a malignant tumor that seriously threatens human health, accounting for approximately 40% of lung cancers [1]. In recent years, some progress has been made in diagnostic and treatment strategies of clinical and experimental oncology for lung cancer. However, LUAD patients with localized or locally advanced disease still have a high risk of death, and their 5-year overall survival rate is still less than 15% [2]. Assessing the patient's prognosis can help choose effective treatments to balance side effects with treatment benefits and decide whether to give more aggressive treatment. Although tumor-node-metastasis (TNM) classification plays an important role in the prognosis assessment of LUAD patients, the prognosis of some patients is significantly different even if the stages are similar. Therefore, the identification of reliable prognostic biomarkers to predict clinical outcomes and help make accurate clinical treatment decisions is clearly critical. The rapid development of gene chips and high-throughput sequencing have facilitated the development of new predictive tools based on prognostic genes for lung cancer. These relevant studies involved in prognostic genes of lung cancer have

identified several prognostic models that have predicted the overall survival rate of LUAD patients (Table 1) [3-13]. For example, a six-gene model (RRAGB, RSPH9, RPS6KL1, RXFP1, RTL1 and RRM2) based on the weighted gene coexpression network predicted the overall survival rate of non-small-cell lung cancer patients [12]. A 3-gene prognostic model (URKA, CDC20, and TPX2A) also accurately predicted overall survival in smoking-related lung adenocarcinoma [13]. In addition, through analysis of TCGA data, the risk score of the 12-mRNA signature was correlated with poor prognosis in patients with lung adenocarcinoma [3]. Therefore, the in-depth exploration of public databases such as the GEO and TCGA databases, discovery of other genes related to the prognosis of LUAD and development of a comprehensive prognosis assessment system including multiple biomarkers may be effective ways to predict the prognosis of lung adenocarcinoma and individual treatment.

Here, we first integrated three lung adenocarcinoma datasets from the GEO database to screen for DEGs. Then, the TCGA-LUAD data set was used to identify DEGs. Univariate Cox and LASSO regression analyses were further used to determine the DEGs associated with overall survival. The risk score was calculated by multiplying multiple Cox coefficients by gene expression. The prognostic model was also combined with clinical parameters to construct a prognostic nomogram to predict overall survival. Finally, Gene Set Enrichment Analysis (GSEA) was performed to identify the potential biological pathways of the five genes in the model.

## **MATERIALS AND METHODS**

### **Gene expression profile data collection from the GEO and TCGA databases.**

The GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) was used for the mRNA expression and clinical data of lung adenocarcinoma that needed to meet the following criteria: (1) human lung adenocarcinoma tissue samples; (2) tumor and nontumor lung control tissue samples; and (3)  $\geq 50$  samples. Finally, three gene expression microarray data sets (GSE43458, GSE32863 and GSE27262), which included 163 LUAD tumor tissue samples and 113 adjacent normal tissue samples, were downloaded for DEG analysis.

On the other hand, the original count data and corresponding clinical data of LUAD patients in the training set and test set, which includes 535 LUAD patient samples and 59 control samples, were downloaded from The Cancer Genome Atlas (TCGA) project (<https://tcga-data.nci.nih.gov/tcga/>). Complete survival information and gene expression profile data of 494 patients were obtained from the TCGA database after excluding samples that could not be assessed for tumor histological grade (GX) or had no OS information. The model was validated using transcriptome analysis of 90 LUAD patients from the GSE11969 dataset. The workflow of our LUAD biomarker analysis process is shown in Figure 1.

### **Screening and verification of differentially expressed genes (DEGs) in lung adenocarcinoma tissue.**

To identify DEGs between LUAD and lung tissues, GE02R was used for differential expression analysis of the GSE43458, GSE32863, and GSE27262 data sets. The DEGs of the TCGA-LUAD dataset were analyzed using the "limma" software package of R software, and the threshold of DEG screening was  $|\log FC| > 2$  and  $P < 0.05$  according to our previous study<sup>[14]</sup>. Human protein mapping (<https://www.proteinatlas.org/>) evaluates lung adenocarcinoma and DEG protein expression in normal lung tissue<sup>[15]</sup>. Mutation data in lung adenocarcinoma patients were obtained from cBioPortal (<https://www.cbioportal.org/>)<sup>[16]</sup>.

### **Identification of prognostic differential genes and establishment of prognostic models.**

TCGA-LUAD data were randomly divided into a training set ( $n = 346$ ) and a test set ( $n = 148$ ). In the test set, we performed univariate Cox regression analysis for DEGs determined by a comprehensive analysis of the GEO data set to determine the relationship between patient survival and gene expression.  $P < 0.01$  was considered statistically significant and was included in subsequent analysis. Next, we applied LASSO regression to further reduce the number of DEGs in the selected panel with the

best predictive performance by 10-fold cross-validation of the R-based glmnet package. Finally, multivariate Cox regression analysis was performed to obtain the five optimal prognostic gene regression coefficients from the multivariate Cox proportional hazard regression model. A prognostic risk score for the five genes was then established based on the multivariate Cox regression model regression coefficient multiplied by a linear combination of its mRNA expression level.

### Identification of prognostic models and related genes.

The Lung Adenocarcinoma (TCGA, PanCancer Atlas) database in cBioPortal was used to analyze the genetic mutation model. We used data from the TCGA to analyze model-related gene expression. The THPA (<http://www.proteinatlas.org>) database was used to analyze the protein expression of model-related genes<sup>[15]</sup>. Patients in the training set were divided into high-risk and low-risk groups according to the median risk score as the cutoff point. Kaplan-Meier survival curves and Wilcoxon tests combined with the R package "survival" were used to compare the survival differences between the high-risk and low-risk groups. Time-dependent ROC curve analysis was conducted using the R software package "survivalROC" to assess the prediction model's forecasting capacity.  $P < 0.05$  was considered statistically significant. The test cohort and the entire cohort were used for internal validation, the GSE11969 dataset was downloaded from the GEO database for external validation, and the risk score of each patient was calculated using the same model based on the prognostic gene signature to further verify the predictive value of the prognostic gene signature.

### Establish and verify the forecast nomograms.

To provide clinicians with a quantitative method for predicting 1-year, 3-year, and 5-year overall survival in LUAD patients, we used a combined model of all independent prognostic factors selected by multivariate Cox regression analysis to construct a nomogram. Kaplan-Meier analysis, AUC, consistency index (C-index), and comparison



of predicted and observed overall survival were used to evaluate the prognostic nomographs' performance<sup>[17]</sup>.

### Functional enrichment analysis of model genes.

<sup>1</sup> Gene set enrichment analysis (GSEA) was used to analyze the signaling pathways of relevant genes involved in the development of lung adenocarcinoma to clarify the molecular mechanism of the prognostic gene signature. GSEA software (GSEA 4.0.3) <sup>6</sup> was downloaded from the Broad Institute website (<http://software.broadinstitute.org/gsea/index.jsp>), and the analyzed access was from the c2.cp.kegg.v7.0.symbols.gmt data set in the Molecular Signature Database (MsigDB). The enrichment analysis was carried out by the weighted enrichment method, and the number of random combinations was set as 1,000. All other parameters were set as default values. Gene sets with  $P < 0.05$  were regarded as significantly enriched gene sets.

### Statistical analysis.

Statistical analysis and corresponding graph drawing were performed using R3.6.3 software, <sup>9</sup> and Cox regression analysis of the hazard ratio (HR) and 95% confidence interval (CI) was used to evaluate the association between DEG expression and prognosis. <sup>21</sup> A t-test of paired samples or a nonparametric Wilcoxon rank sum test of unpaired samples <sup>12</sup> were used for analysis of continuous variables, and categorical variables were tested by the chi-square test or Fisher's exact test.  $P < 0.05$  indicates a significant difference.

## RESULTS

### Screening and identification of differentially expressed genes in the TCGA-LUAD database.

We researched the results as described in the flowchart (Supplemental Figure S1). This study analyzed three GEO datasets (GSE43458, GSE32863, and GSE27262), and 886,

1270, and 1921 DEGs were found, respectively. Then, we found 380 common DEGs by Venn diagram analysis. DEGs were verified in the TCGA-LUAD database (535 Lung adenocarcinoma tissues and 59 Lung cancer tissues), further confirming the differential expression of these 380 genes in lung adenocarcinoma and normal lung tissues (Figure 1A-D).

#### **Screening of prognostic differential genes and establishment of prognostic models.**

Univariate Cox regression analysis was performed on 380 DEGs in the training set. A total of 30 DEGs were related to the survival of patients with lung adenocarcinoma ( $P < 0.05$ ) and further screened by LASSO regression (Figure 1e). Cross-validation was used to establish the model, as shown in Figure 1f. A total of 5 mRNAs (TCN1, CENPF, MAOB, CRTAC1, and PLEK2) were included in the model. Multivariate Cox regression analysis was performed for the above 5 mRNAs, and the risk scoring equation was established according to the corresponding regression coefficient. Risk score (RS) =  $(0.00288 * \text{TCN1 EXP}) + (0.0387 * \text{CENPF EXP}) + (-0.0291 * \text{MAOB EXP}) + (-0.0198 * \text{CRTAC1 EXP}) + (0.0214 * \text{PLEK2 EXP})$ .

#### **Verification of mRNA expression and genetic changes associated with 5 prognostic genes.**

Among the 566 patients included in the cBioPortal for Lung Adenocarcinoma (TCGA, PanCancer Atlas) database, 93 patients (16.4%) showed genetic changes in these 5 genes, among which missense mutations were the most common mutation type (Figure 2A). In the TCGA LUAD cohort, the mRNA expression levels of TCN1, CENPF, and PLEK2 were significantly increased in lung adenocarcinoma tissues compared with those in normal lung tissues, while MAOB and CRTAC1 were significantly decreased in lung adenocarcinoma tissues (Figure 2B). A human protein mapping database (<http://www.proteinatlas.org>) was used to explore the protein expression level. Immunohistochemical (IHC) results of four genes (TCN1 was not included in the database) in lung cancer and normal lung gland tissues are shown in Figure 3c.



Consistent with the mRNA results, IHC results showed that CENPF and PLEK2 had significantly higher mean expression levels in lung adenocarcinoma tissue than in normal lung tissue. In contrast, the CRTAC1 expression level was higher in normal lung tissue than in lung adenocarcinoma tissue. MAOB showed no difference between normal and lung adenocarcinoma tissues (Figure 2C).

### Evaluation of five-mRNA prognostic model.

Each patient's risk score in the training group was calculated based on the above risk score function. The "SurvMiner" R software package was used to obtain the median critical point, and the patients were divided into a high-risk group and a low-risk group (Figure 3A). As the RS score increased, the patients' survival time was shortened, and the number of deaths increased significantly (Figure 3B). Figure 3C shows the heat map of 5 prognostic genes in the high- and low-risk groups. The KM survival curve indicated a lower overall survival in the high-risk group than in the low-risk group ( $P < 0.001$ ) (Figure 3D). To further verify the prognostic assessment model's accuracy, we used the R "survival ROC" package to draw the ROC curve (Figure 3E). The results showed that the AUC values of the risk score model for predicting the overall survival at 1, 3 and 5 years in patients with lung adenocarcinoma were 0.711, 0.668 and 0.728, respectively, indicating that the multigene model had a good predictive ability for the OS of patients with lung adenocarcinoma. Multiple Cox regression analysis showed that RS, along with patient age and stage, could be independent prognostic factors for lung adenocarcinoma patients (Figure 3f).

### Internal and external validation of the five-mRNA prognostic signature.

To verify the predictive value of the 5-mRNA prognostic signature, we used the same formula to calculate risk scores for patients with the internal validation set ( $n = 160$ ), entire validation set ( $n = 535$ ), and external validation set (GSE11969,  $n = 90$ ). Consistent with the training group results, the OS of LUAD patients in the high-risk group was significantly lower than that in the low-risk group (Figure 4A-C). The Kaplan-Meier

<sup>3</sup> survival analysis of the prognostic signature showed that the AUC values of the 1-year, 3-year, and 5-year OS of the internal validation set, the overall validation set and external validation set were 0.754, 0.630, 0.684, and 0.737, 0.701, 0.680, and 0.779, 0.752, 0.715, respectively (Figure 4D-F). Taken together, our results suggest that this 5-gene signature performs well in predicting overall survival in patients with lung adenocarcinoma.

### <sup>19</sup> Establish and verify the nomogram.

To establish clinically applicable methods for predicting survival in patients with lung adenocarcinoma, we established a nomogram using three independent prognostic factors (including age, stage, <sup>16</sup> and risk score) to predict 1-year, 3-year, and 5-year OS in patients with lung adenocarcinoma (Supplemental Figure S2A). The calibration diagram shows that the nomogram performs well (Supplemental Figure S2B). <sup>1</sup> The AUC values of the 1-year, 2-year, and 3-year overall survival predictions of the nomograph were 0.760, 0.712, and 0.709, respectively (Supplemental Figure S3A). The Kaplan-Meier chart effectively distinguishes the various risks of these categories, with people with higher scores having significantly poorer overall survival ( $P < 0.001$ ) (Supplemental Figure S3B). The C-index (95%CI) of the age, stage, and risk score and combination models was 0.501 (0.480 - 0.522), 0.684 (0.662 - 0.706), 0.625 (0.604 - 0.646), and 0.726 (0.702 - 0.750), respectively (Supplementary Table 1). Thus, the nomogram performs well in predicting overall survival in patients with lung adenocarcinoma, which may be useful for patient counseling and clinical decision-making.

### Biological pathways of the five prognostic model genes were identified.

<sup>3</sup> GSEA was performed to identify the potential biological processes of the 5 prognostic genes and showed that the samples with highly expressed TCN1, CENPF, and PLEK2 were enriched with focal adhesion, the p53 signaling pathway, and Toll-like sensors (TLRs), respectively. MAOB and CRTAC1 samples were mediated in the cell cycle and ubiquitin-mediated proteolysis (Supplemental Figure S4), respectively.

## DISCUSSION

Recently, the tumor prognosis model based on the abnormal gene mRNAs has shown great potential due to its high prediction accuracy. Traditional clinicopathological parameters, such as tumor stage, have been used to reflect and predict disease progression. However, a single clinical parameter has poor predictive ability for prognosis [3, 6, 7, 10]. In this study, we identified 380 reliable lung adenocarcinoma differential genes by comprehensive analysis of three GEO datasets combined with data from the TCGA-LUAD. <sup>22</sup> Univariate, LASSO and multivariate Cox analyses of DEGs were performed to establish a prognostic risk model for lung adenocarcinoma based on 5 mRNAs (TCN1, CENPF, MAOB, CRTAC1, and PLEK2). These five new genes were significantly correlated with the prognosis of LUAD patients. MAOB and CRTAC1 were negative prognostic genes, while TCN1, CENPF and PLEK2 were positive prognostic genes. Recently, several studies have revealed the important role of these five genes in cancer progression. Monoamine oxidase B (MAOB) is an enzyme located on the outer mitochondrial membrane. It is responsible for catalyzing monoamine oxidation to produce hydrogen peroxide and is mainly involved in the metabolism of neurotransmitters [18]. The relationship between MAOB and tumors is less reported. It has been reported that MAOB mRNA is <sup>13</sup> significantly lower in the saliva of oral squamous cell carcinoma patients than in that of healthy controls [19]. Currently, this gene is rarely reported in tumors. <sup>10</sup> TCN1 is a member of the vitamin B12-binding protein family. Xu *et al* found that MAOB is a key DNA methylation driver gene for prostate cancer and plays an important role in the DNA methylation of prostate cancer patients through a comprehensive analysis of the TCGA methylation data [20]. There is no previous report about MAOB in lung adenocarcinoma. CRTAC1 encodes human chondrogenic acid protein 1, which can be used as a marker of chondrocytes to distinguish human chondrocytes from osteoblasts and mesenchymal stem cells in cultures [21]. Currently, this gene is rarely reported in tumors. TCN1 is a member of the vitamin B12-binding protein family and is a 60-70 kDa molecular weight protein. High

levels of TCN1 are primarily related to abnormal granulocyte proliferation. TCN1 is overexpressed in a variety of malignancies, such as pancreas, breast, and colon cancer, and is associated with tumor progression and metastasis [22-24]. TCN1 was significantly associated with advanced colorectal cancer<sup>[23]</sup> and laryngeal cancer<sup>[25]</sup>. Centromere protein F (CENPF), as an important member of the centromere protein family, is a component of the centromere complex and plays an important regulatory role in mitosis [26]. CENPF expression is abnormally increased in a variety of malignant tumors and is associated with the prognosis of patients [27, 28]. Using bioinformatics and immunohistochemical analysis, CENPF overexpression was associated with poor prognosis of breast cancer and tumor bone metastasis [29]. Through comprehensive analysis of three GEO databases, CENPF was identified as a key gene with prognostic value in lung adenocarcinoma, which was consistent with our research results [30].

<sup>7</sup> Pleckstrin-2 (PLEK2) is a 353 amino acid protein encoded by the PLEK2 gene in the human genome and is widely expressed in various tissues. Its overexpression contributes to the formation of large apolipoproteins, thereby promoting cell proliferation [31]. PLEK2 has been found to be related to the invasion and metastasis of multiple tumors. <sup>7</sup> In gallbladder cancer (GBC), PLEK2 overexpression enhances the epithelial-mesenchymal transformation (EMT) process in GBC cells, leading to subsequent higher rates of cell migration, invasion, and liver metastasis [32]. The overexpression of PLEK2 also significantly promoted the EMT and migration of non-small cell lung cancer and destroyed the vascular endothelial barrier [33]. After identifying the five prognostic gene markers, we also conducted internal and external validation to confirm their predictive value and revealed that the prognostic signatures had good prognostic diagnostic value. To improve the prognostic predictive power of the five prognostic gene markers, a predictive nomogram combining risk scores and conventional clinical prognostic parameters (including age and tumor stage) was constructed to enable <sup>5</sup> clinicians to determine the prognosis of each patient. Its graphical scoring system is easy to understand and helps customize treatment and medical decisions. The prognostic models and nomograms associated with five-gene



characteristics have not been reported. Hence, our study may be useful prognostic and diagnostic classification tools for lung adenocarcinoma. Our study still has some limitations. First, the study only focuses on transcriptome sequencing data. If other omics techniques, such as DNA methylation and single nucleotide polymorphisms (SNPs), can be analyzed together, more favorable results may be obtained. Second, our research is limited to the bioinformatics analysis of the TCGA and GEO databases. Although we have verified the accuracy of the models internally and externally, the verification of large samples in the clinical diagnosis and treatment process will further enhance their diagnostic accuracy and clinical value.

## **CONCLUSION**

In summary, <sup>5</sup> our study identified a 5-gene model and prognostic nomogram that combined gene models and clinical prognostic factors to predict the overall survival rate of lung adenocarcinoma patients, and this nomogram may be of great significance for the selection of personalized treatment options and clinical medical decisions in patients with lung adenocarcinoma.

## **ARTICLE HIGHLIGHTS**

### ***Research background***

Lung adenocarcinoma patients with localized or locally advanced disease have a high risk of death, and their 5-year overall survival rate is less than 15%.

### ***Research motivation***

<sup>4</sup> To evaluate the prognosis of LUAD patients and optimize treatment, effective clinical research prediction models .

### ***Research objectives***

To identify reliable prognostic biomarkers to predict clinical outcomes and to help clinicians to make accurate clinical treatment decisions.

### ***Research methods***

TCGA and GEO were used to screen for differential genes for lung adenocarcinoma. Univariate regression analysis combined with LASSO regression analysis was used to screen for prognostic-related genes. Multivariate Cox regression analysis was applied to establish the risk score equation and construct the survival prognosis model.

### ***Research results***

We establish a prognostic risk model for lung adenocarcinoma based on 5 mRNAs (TCN1, CENPF, MAOB, CRTAC1, and PLEK2). These five new genes were significantly correlated with the prognosis of LUAD patients. To improve the prognostic predictive power of the five prognostic gene markers, a predictive nomogram combining risk scores and conventional clinical prognostic parameters (including age and tumor stage) was constructed to enable clinicians to determine the prognosis of each patient.

### ***Research conclusions***

A 5-mRNA-based model was constructed to predict the prognosis of lung adenocarcinoma, which may provide clinicians with reliable prognostic assessment tools and help clinical treatment decisions.

### ***Research perspectives***

Our study identified a 5-gene model and constructed a nomogram which may have important implications for clinical medical decision and personalized treatment of patients with lung adenocarcinoma



23%

SIMILARITY INDEX

---

PRIMARY SOURCES

---

- |  |  |                |
|--|--|----------------|
| <div style="background-color: red; color: white; width: 40px; height: 40px; display: flex; align-items: center; justify-content: center; margin: 5px 0;">1</div>     | <a href="http://www.frontiersin.org" style="color: red;">www.frontiersin.org</a><br><small>Internet</small>  | 139 words — 4% |
| <hr/>  |  |                |
| <div style="background-color: magenta; color: white; width: 40px; height: 40px; display: flex; align-items: center; justify-content: center; margin: 5px 0;">2</div> | <a href="#">qianlin xia, Weimo Yu, Qiuyue Li, Jin Wang, Yuzhen Du. "A 5-mRNA-Based Prognostic Signature of Survival in Lung Adenocarcinoma", Research Square Platform LLC, 2021</a><br><small>Crossref Posted Content</small>                                      | 119 words — 3% |
| <hr/>  |  |                |
| <div style="background-color: purple; color: white; width: 40px; height: 40px; display: flex; align-items: center; justify-content: center; margin: 5px 0;">3</div>  | <a href="http://cancerci.biomedcentral.com" style="color: purple;">cancerci.biomedcentral.com</a><br><small>Internet</small>   | 108 words — 3% |
| <hr/>  |  |                |
| <div style="background-color: teal; color: white; width: 40px; height: 40px; display: flex; align-items: center; justify-content: center; margin: 5px 0;">4</div>    | <a href="http://assets.researchsquare.com" style="color: teal;">assets.researchsquare.com</a><br><small>Internet</small>   | 94 words — 2%  |
| <hr/>  |  |                |
| <div style="background-color: green; color: white; width: 40px; height: 40px; display: flex; align-items: center; justify-content: center; margin: 5px 0;">5</div>   | <a href="#">FengLing Shao, Zhenni Wang, Shan Wang. "Identification of -Related Gene as a Potential Biomarker for Neuroblastoma Prognostic Model by Integrated Analysis and Quantitative Real-Time PCR ", DNA and Cell Biology, 2021</a><br><small>Crossref</small> | 77 words — 2%  |
| <hr/>  |  |                |
| <div style="background-color: brown; color: white; width: 40px; height: 40px; display: flex; align-items: center; justify-content: center; margin: 5px 0;">6</div>   | <a href="#">Bo Wang, Wei Wei, Shengrong Long, Lesheng Wang et al. "CENPA acts as a prognostic factor that relates to immune infiltrates in gliomas", Frontiers in Neurology, 2022</a><br><small>Crossref</small>   | 52 words — 1%  |

7 Wenqian Zhang, Tong Li, Bin Hu, Hui Li. " Gene Upregulation Might Independently Predict Shorter Progression-Free Survival in Lung Adenocarcinoma ", Technology in Cancer Research & Treatment, 2020 42 words — 1%

Crossref

8 Wang Jianfeng, Wang Yutao, Bi Jianbin. "Long non-coding RNAs correlate with genomic stability in prostate cancer: A clinical outcome and survival analysis", Genomics, 2021 18 words — < 1%

Crossref

9 [ir.csmu.edu.tw:8080](http://ir.csmu.edu.tw:8080) 17 words — < 1%

Internet

10 [wjso.biomedcentral.com](http://wjso.biomedcentral.com) 17 words — < 1%

Internet

11 [www.researchgate.net](http://www.researchgate.net) 17 words — < 1%

Internet

12 [www.wjgnet.com](http://www.wjgnet.com) 17 words — < 1%

Internet

13 Hikaru Nakashima, Ryoji Yoshida, Akiyuki Hirose, Kenta Kawahara et al. "Circulating miRNA-1290 as a potential biomarker for response to chemoradiotherapy and prognosis of patients with advanced oral squamous cell carcinoma: A single-center retrospective study", Tumor Biology, 2019 15 words — < 1%

Crossref

14 Peng Song, Xiaobin Ma, Dongliang Yang. "Construction of a Prognostic Model of Non-small Cell Lung Cancer Based on TCGA and ImmPort Databases and 15 words — < 1%

# Analysis of Its Correlation With Tumor Microenvironment", Research Square Platform LLC, 2021

Crossref Posted Content

---

15 Xiaoming Li, Yingli Liang, Huixin Wang. 15 words — < 1 %  
"Comprehensive Analysis of Epigenetic  
Alternations of MicroRNA and DNA Methylation Contribute to  
Gene Expression and Prognosis in Lung Squamous Cell  
Carcinoma", Research Square Platform LLC, 2022

Crossref Posted Content

---

16 Jie Luo, Xiaotian Du. "A promising prognostic 14 words — < 1 %  
signature for lung adenocarcinoma (LUAD)  
patients basing on 6 hypoxia-related genes", Medicine, 2021

Crossref

---

17 [www.dovepress.com](http://www.dovepress.com) 14 words — < 1 %  
Internet

---

18 [www.longdom.org](http://www.longdom.org) 14 words — < 1 %  
Internet

---

19 Junyu Long, Lei Zhang, Xueshuai Wan, Jianzhen 13 words — < 1 %  
Lin, Yi Bai, Weiyu Xu, Jianping Xiong, Haitao Zhao.  
"A four-gene-based prognostic model predicts overall survival  
in patients with hepatocellular carcinoma", Journal of Cellular  
and Molecular Medicine, 2018

Crossref

---

20 [www.researchsquare.com](http://www.researchsquare.com) 13 words — < 1 %  
Internet

---

21 Gao-Min Liu, Hua-Dong Zeng, Cai-Yun Zhang, Ji- 12 words — < 1 %  
Wei Xu. "Identification of a six-gene signature  
predicting overall survival for hepatocellular carcinoma",  
Cancer Cell International, 2019

Crossref

22	academic-accelerator.com Internet	12 words — < 1%
23	bioinformatics.com.cn Internet	12 words — < 1%
24	fjfsdata01prod.blob.core.windows.net Internet	12 words — < 1%
25	www.nature.com Internet	12 words — < 1%

EXCLUDE QUOTES ON

EXCLUDE BIBLIOGRAPHY ON

EXCLUDE SOURCES

EXCLUDE MATCHES

OFF

< 12 WORDS