

## Research Paper

# Development and validation of a 4-gene combination for the prognostication in lung adenocarcinoma patients

Xiao-Hong Yin<sup>1,2</sup>, Li-Ping Yu<sup>2</sup>, Xiao-Hong Zhao<sup>2</sup>, Qin-Mei Li<sup>3</sup>, Xiao-Ping Liu<sup>4</sup>✉, Li He<sup>1</sup>✉

1. Department of Hematology, Zhongnan Hospital of Wuhan University, Wuhan, Hubei province, China
2. Wuhan University School of Health Sciences, Wuhan, Hubei province, China
3. Department of Epidemiology, Department of Epidemiology, School of Health Sciences, Wuhan University, Wuhan, Hubei, China
4. Department of Urology, Zhongnan Hospital of Wuhan University, Wuhan, Hubei province, China

✉ Corresponding authors: Li He, Department of Hematology, Zhongnan Hospital of Wuhan University, Donghu road 169, Wuhan, Hubei province, China, 430071. Email: lihe126@hotmail.com Xiao-Ping Liu, Department of Urology, Zhongnan Hospital of Wuhan University, Donghu road 169, Wuhan, Hubei province, China, 430071. Email: liuxiaoping@whu.edu.cn

© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>). See <http://ivyspring.com/terms> for full terms and conditions.

Received: 2019.10.08; Accepted: 2019.12.07; Published: 2020.01.29

## Abstract

**Objective:** To identify a multi-gene prognostic factor in patients with lung adenocarcinoma (LUAD).

**Materials and methods** Prognosis-related genes were screened in the TCGA-LUAD cohort. Then, patients in this cohort were randomly separated into training set and test set. Least absolute shrinkage and selection operator (LASSO) regression was performed to the penalized the Cox proportional hazards regression (CPH) model on the training set, and a prognostication combination based on the result of LASSO analysis was developed. By performing Kaplan-Meier curve analysis, univariate and multivariable CPH model on the overall survival (OS) as well as recurrence free survival (RFS), the prognostication performance of the multigene combination were evaluated. Moreover, we constructed a nomogram and performed decision curve analysis to evaluate the clinical application of the multigene combination.

**Results** We obtained 99 prognosis-related genes and screened out a 4-gene combination (including CIDEA, ZFP3, DKK1, and USP4) according to the LASSO analysis. The results of survival analyses suggested that patients in the 4-gene combination low-risk group had better OS and RFS than those in the 4-gene combination high-risk group. The 4-gene mentioned was demonstrated to be independent prognostic factor of patients with LUAD in the training set (OS, HR=11.962, P<0.001; RFS, HR=9.281, P<0.001) and test set (OS, HR=5.377, P=0.003; RFS, HR=2.949, P=0.104). More importantly, its prognosis performance was well in the validation set (OS, HR=0.955, P=0.002; RFS, HR=1.042, P<0.001).

**Conclusions** We introduced a 4-gene combination which could predict the survival of LUAD patients and might be an independent prognostic factor in LUAD.

Key words: lung adenocarcinoma, prognostication, least absolute shrinkage and selection operator, survival analysis

## Introduction

Lung cancer represents the lion's share of cancer morbidity and mortality worldwide, with 2.1 million new lung cancer cases and 1.8 million deaths in 2018, or nearly 1 in 5 (18.4%) of the total cancer mortality[1] in other words. Lung cancer is mainly classified into small cell lung cancer and non-small cell lung cancer. Approximately, 85% of patients are non-small cell lung cancer, of which lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the

most common subtypes. According to the epidemiological surveys in recent years, the incidence of LUAD has exceeded that of LUSC in both smokers and non-smokers in many countries, accounting for almost a half of all lung cancers; Even worse, in some countries, the incidence of LUAD in women is multiplying, and it also becomes the most common pathological type of lung cancer in young people [2, 3]. Worst of all, LUAD is a kind of non-small cell lung

cancer with extremely high lethality in its advanced stage but no obvious symptoms in its early stages, and thus most patients with LUAD may have delayed detection due to the late onset of symptoms and lack of specificity[4, 5]. Thanks to the continuous improvements of novel treatments and diagnosis techniques[6], the management strategies for LUAD have been improved considerably[7, 8]. However, the prognosis of LUAD patients is generally remaining to be unsatisfactory (the 5-year survival remains about 15%). Therefore, early diagnosis and treatment are crucial to the improvement of the prognosis of LUAD patients, which stresses the demands to develop biomarkers which can help identifying at-risk patients in that they can benefit from early interventions[9]. Biomarkers can be used for screening, diagnosis, and prognosis of tumors, but the individual biomarker is far from enough to meet clinical requirements[10, 11]. In consequence, combined detection of multiple biomarkers is advocated in clinical practice to improve sensitivity and specificity of diagnosis. Nowadays, biomarkers for lung cancer detection mainly include carbohydrate antigens, tumor embryo antigens, differentiation antigens, and proliferating antigens[12, 13]. Nevertheless, the prognostic performance of established biomarkers remains controversial and limited. Thus, in this study, we identified a 4-gene combination to predict the prognosis of patients with LUAD.

## Materials and methods

### Discovery datasets

The cancer genomic atlas (TCGA) program, hosted by the National Cancer Institute's (NCI), molecularly studied more than 20,000 primary cancers and corresponding normal samples across 33 cancer types and the genomic data of these 33 cancer types was uniformly reposit in the NCI's Genomic Data Commons (GDC), which facilitates precision medicine. The expression profile of TCGA-LUAD cohort was measured by using the Illumina HiSeq 2000 RNA Sequencing platform created by the University of North Carolina TCGA genome characterization center. Apart from that, the gene expression profile is generated using GDC mRNA quantification analysis pipeline and obtained from UCSC Xena (<https://xenabrowser.net/datapages/>). The matched clinical data was also obtained from GDC Xena Hub (<https://gdc.xenahubs.net>). Patients in the TCGA-LUAD including 503 LUAD samples were included in the presents according to the following inclusion criteria: patients with primary lung adenocarcinoma; patients whose survival information was documented; patients whose tumor

tissues were subjected to RNA sequencing; patients who were not previously treated; Meanwhile, patients meeting the exclusion criteria (patients with history of other malignancies, patients whose tumor tissues were not undertaken to RNA sequencing analysis, patients with metastasis disease; patients whose survival information was not available; repeated mRNAs sequencing samples) were excluded. Then, the remaining patients in the TCGA-LUAD were randomly assigned to a training set and test set in a 1:1 ratio for subsequent analysis using the R package "caret".

### Validation dataset

To perform an independently external validation, we used GSE31210 as a validation dataset. The gene expression profile and clinical data were downloaded from the Gene Expression Omnibus (GEO)(<https://www.ncbi.nlm.nih.gov/geo/>). GSE31210 was measured by Affymetrix Human Genome U133 Plus 2.0 Array including 226 LUAD samples[11, 14]. The inclusion and exclusion criteria were the same with those used in the TCGA-LUAD.

### Statistical analysis

Firstly, we screened the prognosis-related genes in TCGA-LUAD by using univariate Cox proportional hazards regression (CPH) model, genes with Bonferroni corrected  $P < 0.05$  were obtained. Then, R package "caret" was implied to perform stratified random sampling based on survival status, which was randomly divided into training set and test set according the ratio of 1:1. The R package "glmnet" was used to perform a least absolute shrinkage and selection operator (LASSO) regression to penalized CPH model for minimizing the overfitting of the Cox regression on the training set. After that, the risk score of each patient was calculated utilizing the predict function based on this penalized CPH model in the training set, test set and validation set. Since time-dependent survival receiver operating characteristic (ROC) is an important method to evaluate the predictive ability of prognostic models, R packages "survivalROC" was utilized to perform time-dependent ROC for identifying optimal cutoff value corresponding to the highest true-positive rate and the lowest false-positive rate, which was then used to divide the LUAD patients into low-risk groups and high-risk groups. Subsequently, prognostic significance of the multigene combination was estimated by evaluating the differences of Kaplan-Meier survival curve analysis between the two groups in terms of overall survival (OS) and relapse-free survival (RFS). Furthermore, univariate and multivariable CPH models were utilized to

identify independent prognostic factors associated with survival, and we considered risk score, age, gender, and pathologic stage as covariate. The above survival analysis was conducted using the R packages "survminer" and "survival".

After obtaining the above results, GSE31210 was used as an external validation dataset. Therefore, the risk score of each patient was calculated based on the regression coefficients of genes in the multigene combination, and the cut off value was also found as introduced above. According to the optimal cutoff value, the LUAD patients in the external validation set were separated into high-risk group and low-risk group. In addition, Kaplan-Meier survival curve analysis, univariate and multivariable CPH model were performed to validate the prognostic performance of discovery dataset. To further confirm the clinical relevance of the multigene combination, we compared the prognostication performance of the multigene combination with several well-established biomarkers[15-26] in terms of C-index. All statistical analyses were performed using the R 3.5.2 software, and  $P < 0.05$  was considered statistically significant.

### Enrichment analysis

In order to explore the potential molecular mechanisms of the prognosis-related genes affecting the survival of LUAD patients, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were conducted by using the DAVID online tool (<https://david.ncifcrf.gov/>)[27, 28]. We considered  $P < 0.05$  and false discovery rate (FDR)  $P < 0.05$  as significantly enrichment, and used R package "ggplot2" to visualize the significantly enriched GO and KEGG terms.

### Clinical application of the 4-gene combination

Nomogram is a commonly used prediction tool in medicine. It integrates diverse prognostic and determinant variables to generate individual and numerical probabilities of clinical events[29]. To clarify the clinical application ability of the 4-gene combination, we built a 4-gene combination based nomogram estimating the 3-year, and 5-year OS of LUAD patients, which included the age, gender, stage, and the risk score of each BC patient into a multivariate survival model. After using 1000 resampled bootstraps to internally verify the predicted values of the nomogram, we applied the R package "rms" to draw the nomogram. Meantime, we performed decision curve analysis (DCA) to verify the clinical role of the nomogram for the 4-gene combination[30].

## Results

### Characteristics of LUAD patients in the discovery dataset and validation dataset

A total of 729 LUAD patients with survival information were included in the present study. There were 502 patients in the discovery dataset, and the characteristics of patients in the training set and test set are balanced (median age in years was 66 [range 38-87], 138 females [52.98%], and 118 males [47.01%] in the training set; median age in years was 67 [range 40-88], 133 females [52.99%] in the test set). Meanwhile, there were 226 patients in the validation dataset (GSE31210) (median age in years was 61 [range 30-76], 121 females [53.54%], and 105 males [46.46%]), detailed characteristics of patients in the training set, test set, and GSE31210 were summarized in supplementary table 1.

### Development of the prognostic multigene combination and validation of its prognostic performance

A total of 99 prognosis-related genes were identified in the TCGA-LUAD after Bonferroni correction(Supplementary table 2). As a result of LASSO penalized CPH model in the training set, four genes (CIDEA [ubiquitin specific peptidase 4], ZFP3 [ZFP3 zinc finger protein], DKK1 [dickkopf WNT signaling pathway inhibitor 1], and USP4 [ubiquitin specific peptidase 4]) were finally identified. Thus, a 4-gene combination was built on the basis of the coefficients of these 4 genes in the LASSO penalized CPH model (supplementary table 3). According to the optimal cutoff value of 0.97, 0.978 and -3.829 (Figure 1), we classified the patients in training set, test set and validation set into the 4-gene combination high-risk group and 4-gene combination low-risk group, respectively. The characteristics of the 4-gene combination were shown in Figure 2.

The Kaplan-Meier curve of OS and RFS, and univariate CPH model of OS and RFS results suggested that patients in the 4-gene combination low-risk group was associated with better OS and RFS compared with those in the 4-gene combination high-risk group in the training set (OS, hazards ratio (HR)=11.962, 95% CI: 6.232-22.961,  $P < 0.001$ ; RFS, HR=9.281, 95% CI: 4.064-21.193,  $P < 0.001$ ; figure 3 and supplementary table 4), test set (OS, HR=5.377, 95% CI: 1.736-16.657,  $P = 0.003$ ; RFS, HR=2.949, 95% CI: 0.800-10.867,  $P = 0.104$ ; figure 3 and supplementary table 4) and validation set (OS, HR=1.057, 95% CI: 1.016-1.079,  $P = 0.002$ ; RFS, HR=1.042, 95% CI: 1.018-1.066,  $P < 0.001$ ; figure 4 and Supplementary table 4). Meanwhile, the multivariable CPH model of OS and RFS results exhibited that the 4-gene

combination might be an independent prognostic factor of LUAD patients in the training set, test set, and validation set (Supplementary table 4) after adjusting other clinical factors such as age, gender, smoking status, and stage. The results showed that the 4-gene combination was effective in predicting OS and RFS in LUAD patients.

### Validation of the prognostic performance of the 4-gene combination

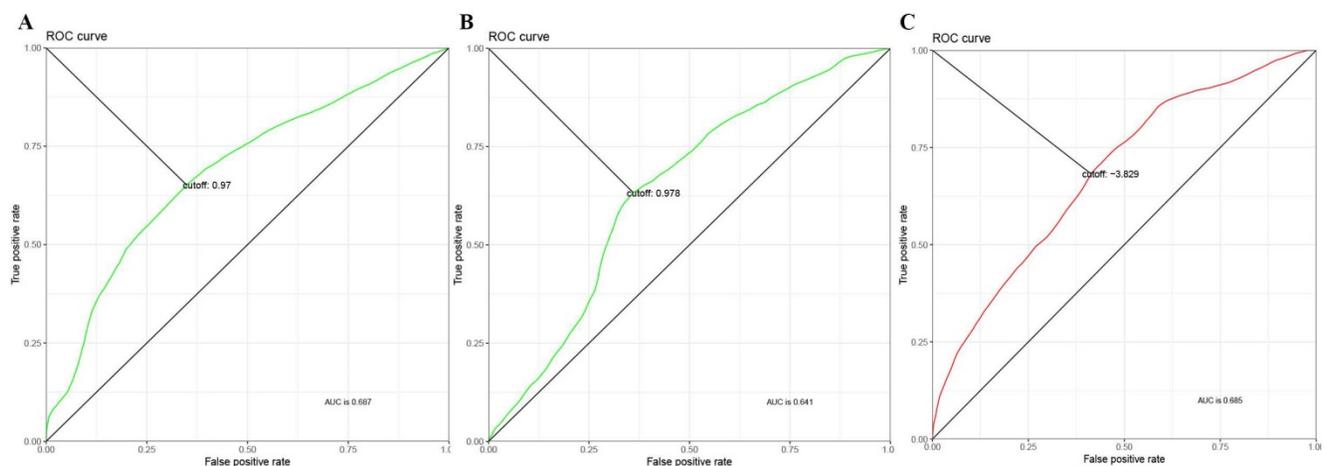
As shown in figure 3, the Kaplan-Meier curve of OS and RFS, and univariate CPH model of OS and RFS (OS, HR=0.955, 95% CI: 1.016-1.079, P=0.002; RFS, (HR)=1.042, 95% CI: 1.018-1.066, P<0.001; Figure 3 and Supplementary table 4) in the 4-gene combination low-risk group were also significantly better than those in the 4-gene combination high-risk group. In addition, although the multivariable CPH model of OS and RFS results were not statistically significant (OS, HR=1.029, 95% CI: 0.997-1.062, P=0.076; RFS, (HR)=1.024, 95% CI: 0.999-1.049, P=0.052; Supplementary table 4), it indicated that the 4-gene combination has a tendency of being an independent prognostic factor in the validation cohort. The above results suggested that the prognostic performance of the 4-gene combination in the discovery dataset and the validation dataset was significant.

Given the fact that several well-established prognostic signatures or biomarkers have been introduced, we tried to compare the prognostic performance of these biomarkers with our 4-gene signature. Zhang et al. introduced a 3-gene signature based on the main members of kinesin family member genes (KIF14, KIF18B, and KIF20A), and they demonstrated that it can significantly stratify patients into low-risk group and high-risk group[15]. Sun et al. suggested a 2-gene signature composed of BRCA1 and ERBB3 and applied it to the prognosis prediction in patients with lung adenocarcinoma[16].

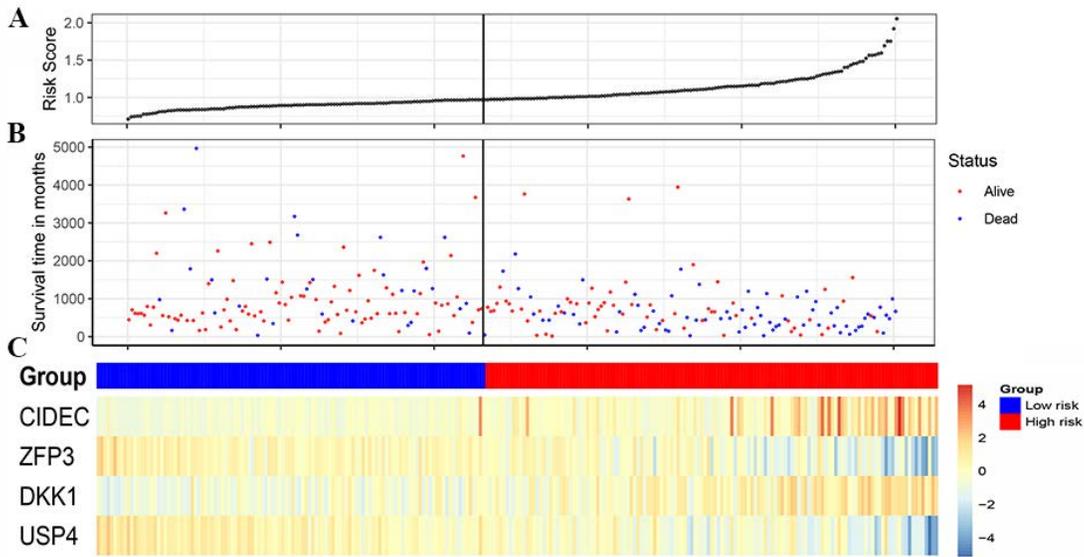
Meanwhile, there were several well-established biomarkers including (ALDH1A1[17], CD117[18], CELIAC1[19], CX3CL1[20], IFITM1[21], ITGA2B[22], LETM1[23], PHLPP2[24], RRBP1[25], and TMEM213[26]) in lung adenocarcinoma. We calculated the C-indexes of our 4-gene signature and the well-established signatures mentioned above. As shown in the following figure 1, the C-index of the 4-gene signature was obviously higher than that of the other biomarkers in the training set, test set, and the independent validation set GSE31210, indicating that our 4-gene signature has superior predictive performance compared to other models and is more reliable in clinical settings.

### The results of the enrichment analysis

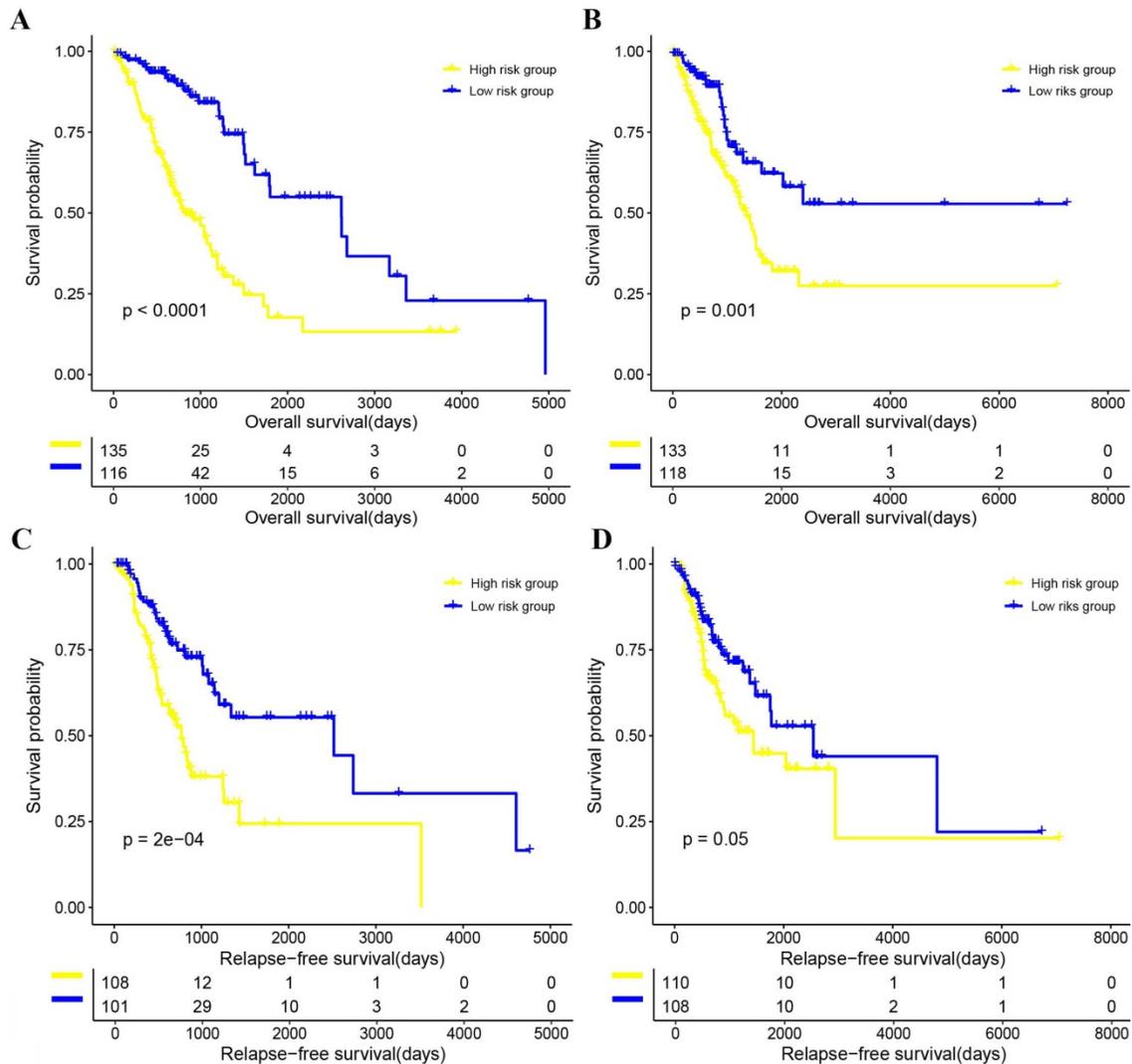
In order to have a rudimentary knowledge of the biological meaning of the prognosis-related genes, GO and KEGG enrichment analysis were conducted. As shown in Figure 5, the prognosis-related genes were mostly enriched in GO terms related to gene regulation ("negative regulation of peptidyl-serine phosphorylation", "negative regulation of insulin receptor signaling pathway", "negative regulation of protein binding", "positive regulation of inhibitory postsynaptic potential", "positive regulation of circadian rhythm", "negative regulation of platelet aggregation" and "negative regulation of microtubule polymerization"), cell proliferation ("intermediate filament cytoskeleton organization", "activation of protein kinase activity" and "L-fucose catabolic process") and signal transduction ("intracellular signal transduction" and "Wnt signaling pathway involved in somitogenesis") (Figure 5). Furthermore, the results of KEGG pathway enrichment analysis of prognosis-related genes showed that these genes were mainly enriched in metabolic pathways ("Metabolic pathways" and "GABAergic synapse") (Figure 5).



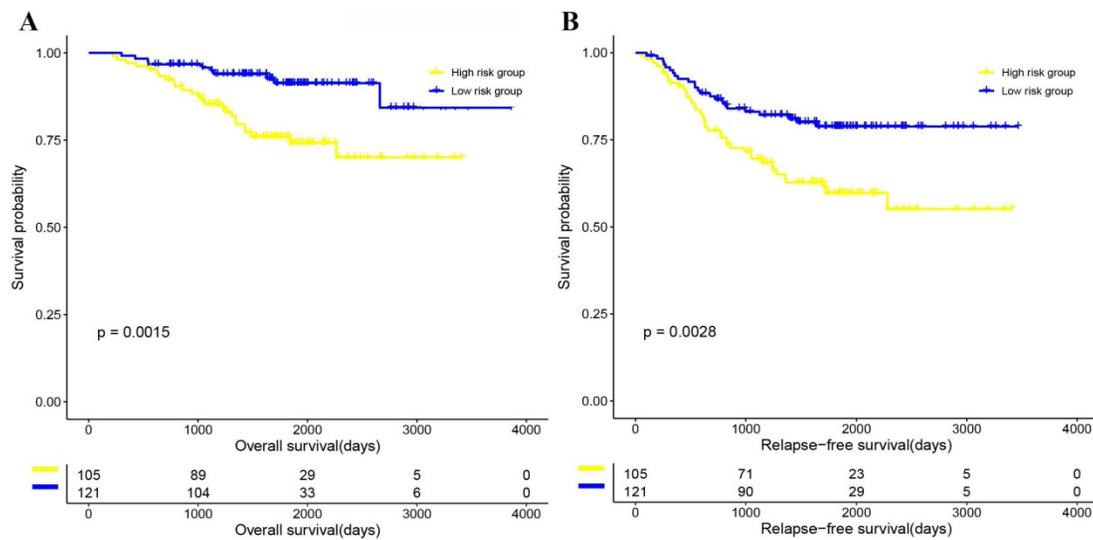
**Figure 1.** Time-Dependent ROC Curve of 4-gene combination. (A) ROC in the training set. (B) ROC in the test set. (C) ROC in the validation set.



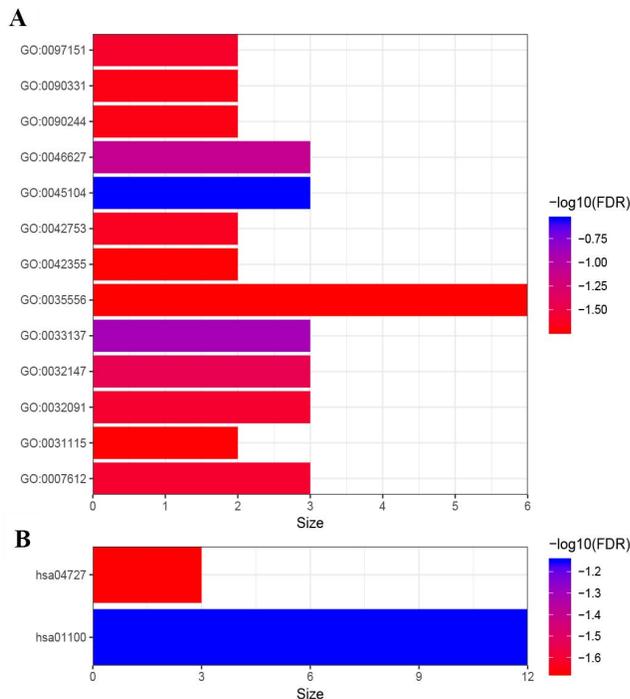
**Figure 2.** Characteristics of the 4-gene combination of the discovery cohort. (A) Risk score (On the left is the low-risk group and on the right is the high-risk group). (B) Survival time in days (Red dot indicates Alive, blue dot indicates death). (C) Gene expression heatmap (The blue color is the low-risk group and the red color is the high-risk group)



**Figure 3.** The correlations between the 4-gene combination and the overall survival (OS) and relapse-free survival (RFS) of patients with LUAD. (A) OS in the training set. (B) OS in the test set. (C) RFS in the training set. (D) RFS in the test set.



**Figure 4.** Kaplan-Meier curve survival analysis on validation set. (A) OS in the validation set. (B) RFS in the validation set.



**Figure 5.** Functional enrichment analysis of 99 genes. (A) GO enrichment analysis. (B) KEGG enrichment analysis. GO:0045104 intermediate filament cytoskeleton organization; GO:0033137~negative regulation of peptidyl-serine phosphorylation; GO:0046627~negative regulation of insulin receptor signaling pathway; GO:0032147~activation of protein kinase activity; GO:0032091~negative regulation of protein binding; GO:0007612~learning; GO:0097151~positive regulation of inhibitory postsynaptic potential; GO:0042753~positive regulation of circadian rhythm; GO:0090244~Wnt signaling pathway involved in somitogenesis; GO:0090331~negative regulation of platelet aggregation; GO:0031115~negative regulation of microtubule polymerization; GO:0042355~L-fucose catabolic process; GO:0035556~intracellular signal transduction. hsa01100~Metabolic pathways; hsa04727~GABAergic synapse.

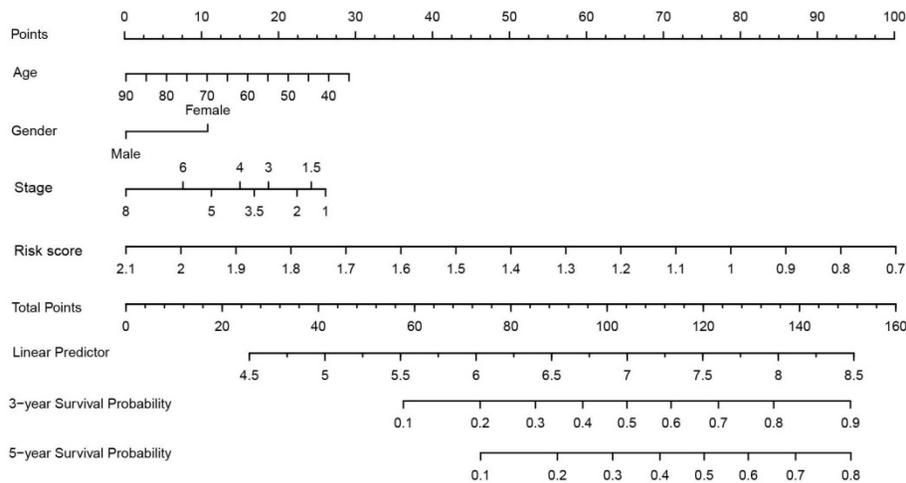
**Clinical application of the 4-gene combination**

As shown in figure 6, we constructed a prognostic nomogram which included age, gender,

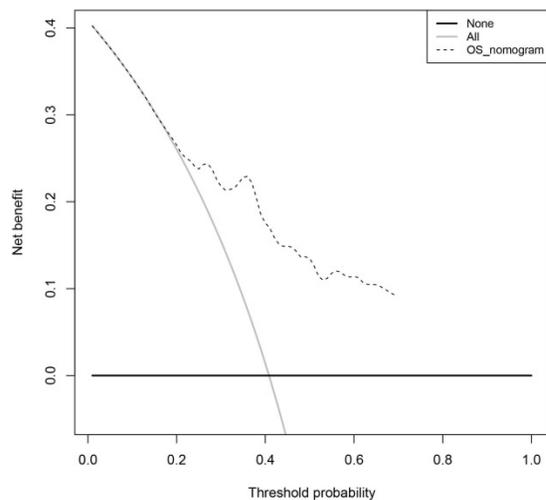
stage and the risk score to predict the 3- and 5-year OS of patients with LUAD. The internally and externally validated Harrell's c index were 0.726 and 0.654 , indicating that the 4 gene combination performed well in clinical application. DCA could determine a range of threshold probabilities for a prediction mode, as shown in Figure 7, the nomogram threshold probability based on 4-gene combinations was significantly better than the default strategies of treating all or none at a threshold probability more than 4%.

**Discussions**

Given the low rate of 5-year survival as well as high rate of recurrence of LUAD patients, and also early diagnosis along with intervention can improve the clinical outcomes of LUAD patients, identification and evaluation of novel biomarkers for patients with LUAD is of great importance. In the present study, univariate CPH analysis was performed to analyze the relations between the expression of genes and the OS of patients with LUAD, and the results exhibited that 99 genes were associated with the OS of LUAD patients. LASSO (34) was introduced in order to improve the prediction accuracy and interpretability of regression models by forcing certain coefficients to be set to zero, and effectively to choose a simpler model that did not include those coefficients. Based on this, the above 99 genes were included into a LASSO penalized CPH model, and as a result 4 genes with none-zero coefficients in this model were obtained. Therefore, the risk score of each LUAD patients was calculated, and the 4-gene combination was built for prediction of overall survival and recurrence of patients with LUAD.



**Figure 6.** Nomogram construction based on 4-gene combination.



**Figure 7.** The decision curve analysis of the 4-gene combination.

The OS and RFS of LUAD patients in the 4-gene combination low-risk group significantly superior compared with those in the 4-gene combination high-risk group, and the 4-gene combination was shown to be an independently prognostic combination in LUAD. Actually, studies have reported that these four genes in the combination are associated with cancer progression or suppression. DKK1 was a member of the dickkopf family and members of this family were secreted proteins characterized by two cysteine-rich domains that mediate protein-protein interactions[31]. This gene played a role in embryonic development and might be important in bone formation in adults[32]. Elevated expression of this gene had been observed in numerous human cancers and its corresponding protein promoted proliferation, invasion and growth in several kinds of cancer cell lines[33, 34]. Aufderklamm S et al. demonstrated that DKK-1 could inhibit osteoblast activity by blocking the Wnt

pathway, which led to progression of metastatic prostate cancer[35]. CIDEc, a member of the cell death-inducing DNA fragmentation factor-like effector family, could promote lipid droplet formation in adipocytes and mediate adipocyte apoptosis[36]. It was reported that the CIDE family regulated lipid metabolism and played an important role in the development of metabolic disorders such as obesity[37], insulin resistance[38] and hepatic steatosis[39]. Ming Yu et al. found that CIDE family was highly expressed in clear cell renal cell carcinoma[40], USP4, the protein encoded by this gene is a protease that deubiquitinates target proteins, shuttles between the nucleus and cytoplasm and is involved in maintaining operational fidelity in the endoplasmic reticulum[41]. Hou X et al. demonstrated that USP4 was significantly upregulated in head and neck squamous cell carcinoma. USP4 might negatively regulate RIP1-mediated NF- $\kappa$ B activation and promotes TNF- $\alpha$ -induced apoptosis in FaDu cells, as well as directly interacted with receptor-interacting protein 1 (RIP1) and deubiquitinated K63-linked ubiquitination from RIP1[42]. Yao R et al. suggested that esophageal cancer patients with high USP4 expression had relative longer survival time[43]. Some researchers also concluded that USP4 expression might be a favorable biomarker for OS and RFS in LUAD patients, while it was worth noting that USP4 expression has no prognostic value for OS and RFS in lung squamous cell carcinoma[44]. In addition, Ye et al. suggested that USP4 had anti-cancer effect in breast cancer, and found that USP4 could inhibit the growth of tumors in a mouse tumor xenograft model[45]. ZFP3 contains a conserved structural motif that mediates its binding to protein, DNA and RNA, in humans ZFP3 has been localized to chromosome 17p12-17pter[46]. Ming et al. reported that the ZFP3 had protective effect in clear cell renal cell

carcinoma[47], which was consistent with our research.

Meanwhile, the GSE31210 dataset was used to validate the prognostication ability of the 4-gene combination. It has been proved that 4-gene combination can be used as an independent prognostic factor for LUAD patients in the validation cohort. Results of functional enrichment analysis of the prognosis-related genes suggested that the prognosis-related genes were mainly enriched in gene regulation, cell proliferation and signal transduction related GO terms and pathways, this was in accordance with the prognostic value of the 4-gene combination. In addition, we obtained good results by detecting the clinical potency of the 4 gene combination by nomogram and DCA. Collectively, we suggested that our 4-gene combination was a good supplement for the prognosis prediction of patients with LUAD.

Although the four gene combinations showed excellent performance in the training set, test set and validation set, it had the following defects. First, this study was an integration and reanalysis of existing published LUAD gene expression studies. Although it showed good performance in prediction of the survival and recurrence of LUAD patients, it had not been verified by large scale prospective trials. Second, the associated mechanisms had not been validated in LUAD cells. Based on this, our follow-up researches will focus on verifying the conclusions of this study in terms of clinical application and molecular mechanisms.

In conclusions, we introduced a 4-gene combination which could predict the survival of LUAD patients and might be an independent prognostic factor in LUAD.

## Supplementary Material

Supplementary figures and tables.

<http://www.jcancer.org/v11p1940s1.pdf>

## Competing Interests

The authors have declared that no competing interest exists.

## References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018; 68: 394-424.
- Yang B, Li J, Li F, Zhou H, Shi W, Shi H, et al. Comprehensive analysis of age-related somatic mutation profiles in Chinese young lung adenocarcinoma patients. *Cancer Med*. 2019; 8: 1350-8
- Mao Y, Yang D, He J, Krasna MJ. Epidemiology of Lung Cancer. *Surg Oncol Clin N Am*. 2016; 25: 439-45.
- Bunn PA, Jr. Worldwide overview of the current status of lung cancer diagnosis and treatment. *Arch Pathol Lab Med*. 2012; 136: 1478-81.
- Ridge CA, McErean AM, Ginsberg MS. Epidemiology of lung cancer. *Semin Intervent Radiol*. 2013; 30: 93-8.
- Eggert JA, Palavanzadeh M, Blanton A. Screening and Early Detection of Lung Cancer. *Semin Oncol Nurs*. 2017; 33: 129-40.
- Vargas AJ, Harris CC. Biomarker development in the precision medicine era: lung cancer as a case study. *Nat Rev Cancer*. 2016; 16: 525-37.
- Yang F, Li H, Wang J. [The precision surgical treatment strategies of lung adenocarcinoma under the guidance of the WHO new classification]. *Zhonghua Wai Ke Za Zhi*. 2017; 55: 49-53.
- Villalobos P, Wistuba II. Lung Cancer Biomarkers. *Hematol Oncol Clin North Am*. 2017; 31: 13-29.
- Li S, Gao M, Li Z, Song L, Gao X, Han J, et al. Role of microRNAs in metastasis of non-small cell lung cancer. *Front Biosci (Landmark Ed)*. 2016; 21: 998-1005.
- Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res*. 2012; 72: 100-11.
- I H, Cho JY. Lung Cancer Biomarkers. *Adv Clin Chem*. 2015; 72: 107-70.
- Scott A, Salgia R. Biomarkers in lung cancer: from early detection to novel therapeutics and decision making. *Biomark Med*. 2008; 2: 577-86.
- Yamauchi M, Yamaguchi R, Nakata A, Kohno T, Nagasaki M, Shimamura T, et al. Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma. *PLoS One*. 2012; 7: e43923.
- Zhang L, Zhu G, Wang X, Liao X, Huang R, Huang C, et al. Genomewide investigation of the clinical significance and prospective molecular mechanisms of kinesin family member genes in patients with lung adenocarcinoma. *Oncol Rep*. 2019; 42: 1017-34.
- Sun Y, Hou L, Yang Y, Xie H, Yang Y, Li Z, et al. Two-gene signature improves the discriminatory power of IASLC/ATS/ERS classification to predict the survival of patients with early-stage lung adenocarcinoma. *Onco Targets Ther*. 2016; 9: 4583-91.
- Koh YW, Han JH, Haam S, Jung J. ALDH1 expression correlates with an epithelial-like phenotype and favorable prognosis in lung adenocarcinoma: a study based on immunohistochemistry and mRNA expression data. *J Cancer Res Clin Oncol*. 2019; 145: 1427-36.
- Guo YN, Liang L, Ren S, Wu M, Shi DC, Mo WJ, et al. CD117 expression is correlated with poor survival of patients and progression of lung carcinoma: a meta-analysis with a panel of 2645 patients. *Pol J Pathol*. 2019; 70: 63-78.
- Zhang L, Li M, Deng B, Dai N, Feng Y, Shan J, et al. HLA-DQB1 expression on tumor cells is a novel favorable prognostic factor for relapse in early-stage lung adenocarcinoma. *Cancer Manag Res*. 2019; 11: 2605-16.
- Liu J, Li Y, Zhu X, Li Q, Liang X, Xie J, et al. Increased CX3CL1 mRNA expression level is a positive prognostic factor in patients with lung adenocarcinoma. *Oncol Lett*. 2019; 17: 4877-90.
- Koh YW, Han JH, Jeong D, Kim CJ. Prognostic significance of IFITM1 expression and correlation with microvessel density and epithelial-mesenchymal transition signature in lung adenocarcinoma. *Pathol Res Pract*. 2019; 215: 152444.
- Xing S, Zeng T, Xue N, He Y, Lai YZ, Li HL, et al. Development and Validation of Tumor-educated Blood Platelets Integrin Alpha 2b (ITGA2B) RNA for Diagnosis and Prognosis of Non-small-cell Lung Cancer through RNA-seq. *Int J Biol Sci*. 2019; 15: 1977-92.
- Piao L, Yang Z, Feng Y, Zhang C, Cui C, Xuan Y. LETM1 is a potential biomarker of prognosis in lung non-small cell carcinoma. *BMC Cancer*. 2019; 19: 898.
- Huang H, Pan X, Jin H, Li Y, Zhang L, Yang C, et al. PHLPP2 Downregulation Contributes to Lung Carcinogenesis Following B[a]P/B[a]PDE Exposure. *Clin Cancer Res*. 2015; 21: 3783-93.
- Liu S, Lin M, Ji H, Ding J, Zhu J, Ma R, et al. RRBP1 overexpression is associated with progression and prognosis in endometrial endometrioid adenocarcinoma. *Diagn Pathol*. 2019; 14: 7.
- Zou J, Li Z, Deng H, Hao J, Ding R, Zhao M. TMEM213 as a novel prognostic and predictive biomarker for patients with lung adenocarcinoma after curative resection: a study based on bioinformatics analysis. *J Thorac Dis*. 2019; 11: 3399-410.
- Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4: 44-57.
- Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009; 37: 1-13.
- Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol*. 2015; 16: e173-80.
- Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006; 26: 565-74.
- Daoussis D, Papachristou DJ, Dimitroulas T, Sidiropoulos T, Antonopoulos I, Andonopoulos AP, et al. Dickkopf-1 is downregulated early and universally in the skin of patients with systemic sclerosis despite normal circulating levels. *Clin Exp Rheumatol*. 2018; 36 Suppl 113: 45-9.
- Rojas A, Mardones R, Pritzker K, van Wijnen AJ, Galindo MA, Las Heras F. Dickkopf-1 reduces hypertrophic changes in human chondrocytes derived from bone marrow stem cells. *Gene*. 2019; 687: 228-37.
- Hong SA, Yoo SH, Lee HH, Sun S, Won HS, Kim O, et al. Prognostic value of Dickkopf-1 and ss-catenin expression in advanced gastric cancer. *BMC Cancer*. 2018; 18: 506.
- Gao H, Li L, Xiao M, Guo Y, Shen Y, Cheng L, et al. Elevated DKK1 expression is an independent unfavorable prognostic indicator of survival in head and neck squamous cell carcinoma. *Cancer Manag Res*. 2018; 10: 5083-9.

35. Aufderklamm S, Hennenlotter J, Leidenberger P, Rausch S, Hohneder A, Kuhs U, et al. Systemic Alterations of Wnt Inhibitors in Patients with Prostate Cancer and Bone Metastases. *Dis Markers*. 2018; 2018: 1874598.
36. Xu Y, Gu Y, Liu G, Zhang F, Li J, Liu F, et al. Cidec promotes the differentiation of human adipocytes by degradation of AMPKalpha through ubiquitin-proteasome pathway. *Biochim Biophys Acta*. 2015; 1850: 2552-62.
37. Zhou L, Yu M, Arshad M, Wang W, Lu Y, Gong J, et al. Coordination Among Lipid Droplets, Peroxisomes, and Mitochondria Regulates Energy Expenditure Through the CIDE-ATGL-PPARalpha Pathway in Adipocytes. *Diabetes*. 2018; 67: 1935-48.
38. Wang H, Ti Y, Zhang JB, Peng J, Zhou HM, Zhong M, et al. Single nucleotide polymorphisms in CIDE gene are associated with metabolic syndrome components risks and antihypertensive drug efficacy. *Oncotarget*. 2017; 8: 27481-8.
39. Andersen E, Ingerslev LR, Fabre O, Donkin I, Altintas A, Versteyhe S, et al. Preadipocytes from obese humans with type 2 diabetes are epigenetically reprogrammed at genes controlling adipose tissue function. *Int J Obes (Lond)*. 2019; 43: 306-18.
40. Yu M, Wang H, Zhao J, Yuan Y, Wang C, Li J, et al. Expression of CIDE proteins in clear cell renal cell carcinoma and their prognostic significance. *Mol Cell Biochem*. 2013; 378: 145-51.
41. Yun SI, Kim HH, Yoon JH, Park WS, Hahn MJ, Kim HC, et al. Ubiquitin specific protease 4 positively regulates the WNT/beta-catenin signaling in colorectal cancer. *Mol Oncol*. 2015; 9: 1834-51.
42. Hou X, Wang L, Zhang L, Pan X, Zhao W. Ubiquitin-specific protease 4 promotes TNF-alpha-induced apoptosis by deubiquitination of RIP1 in head and neck squamous cell carcinoma. *FEBS Lett*. 2013; 587: 311-6.
43. Yao R, Pu J, Fan R, Zhu W, Ding X, Shen X, et al. Ubiquitin-specific protease 4 improves the prognosis of the patients in esophageal cancer. *Cancer Biomark*. 2017; 20: 317-23.
44. Zhong M, Jiang Q, Jin R. USP4 expression independently predicts favorable survival in lung adenocarcinoma. *IUBMB Life*. 2018; 70: 670-7.
45. Li Y, Jiang D, Zhang Q, Liu X, Cai Z. Ubiquitin-specific protease 4 inhibits breast cancer cell growth through the upregulation of PDCD4. *Int J Mol Med*. 2016; 38: 803-11.
46. Ashworth A, Williams BP, Buchberg AM, Goodfellow PN, Solomon E, Potter J, et al. Chromosomal localization of zinc finger protein genes in man and mouse. *Genomics*. 1989; 4: 323-7.
47. Yang J, Yang J, Gao Y, Zhao L, Liu L, Qin Y, et al. Identification of potential serum proteomic biomarkers for clear cell renal cell carcinoma. *PLoS One*. 2014; 9: e111364.