Development and validation of a prediction model for lung adenocarcinoma based on RNA-binding protein

Longjun Yang[1,2], Rusi Zhang[1,2], Guangran Guo[1,2], Gongming Wang[1,2], Yingsheng Wen[1,2], Yongbin Lin[1,2], Xuewen Zhang[1,3], Xiangyang Yu[4], Zirui Huang[1,2], Dechang Zhao[1,2], Lanjun Zhang[1,2]; written on behalf of the AME Thoracic Surgery Collaborative Group

[1]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China; [2]Department of Thoracic Surgery, Sun Yat-sen University Cancer Center, Guangzhou, China; [3]Department of Anesthesiology, Sun Yat-sen University Cancer Center, Guangzhou, China; [4]Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Contributions: (I) Conception and design: L Yang; (II) Administrative support: Y Lin, Y Wen, and L Zhang; (III) Provision of study materials or patients: L Yang, R Zhang, G Guo; (IV) Collection and assembly of data: R Zhang, G Guo, G Wang, Y Wen, Z Huang, X Zhang, D Zhao; (V) Data analysis and interpretation: L Yang, R Zhang, G Guo, X Yu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Lanjun Zhang. Department of Thoracic Surgery, Sun Yat-sen University Cancer Center, 651 Dongfeng Road East, Guangzhou 510060, China. Email: zhanglj@sysucc.org.cn.

Background: RNA-binding proteins (RBPs) have been found to participate in the development and progression of cancer. This present study aimed to construct a RBP-based prognostic prediction model for lung adenocarcinoma (LUAD).

Methods: RNA sequencing data and corresponding clinical information were acquired from The Cancer Genome Atlas (TCGA) and served as a training set. The prediction model was validated using the dataset in Gene Expression Omnibus (GEO) databases. Univariate and multivariate Cox regression analyses were conducted to identify the RBPs associated with survival. R software (http://www.r-project.org) was used for analysis in this study.

Results: Nine hub prognostic RBPs (*CIRBP*, *DARS2*, *DDX24*, *GAPDH*, *LARP6*, *SNRPE*, *WDR3*, *ZC3H12C*, *ZC3H12D*) were identified by univariate Cox regression analysis and multivariate Cox regression analysis. Using a risk score based on the nine-hub RBP model, we separated the LUAD patients into a low-risk group and a high-risk group. The outcomes revealed that patients in the high-risk group had poorer survival than those in the low-risk group. This signature was validated in the GEO database. Further study revealed that the risk score can be an independent prognostic biomarker for LUAD. A nomogram based on the nine hub RBPs was built to quantitatively predict the prognosis of LUAD patients.

Conclusions: Our nine-gene signature model could be used as a marker to predict the prognosis of LUAD and has potential for use in treatment individualization.

Keywords: Lung adenocarcinoma (LUAD); RNA-binding protein (RBPs); overall survival (OS); prognostic model

Introduction

Lung cancer has the highest mortality among cancers and the second highest incidence in both men and women (1). There were 228,820 new cases of lung cancer diagnosed in the United States and more than 135,720 related deaths in 2020 (1). An estimated 85% of lung cancers are non-small cell lung cancer (NSCLC), which mainly consists of lung squamous cell carcinoma and lung adenocarcinoma (LUAD) (2). Of these, LUAD is the most prevalent

Page 2 of 14

Yang et al. Prediction model for LUAD based on RNA-binding protein

NSCLC subtype, constituting about 40% of all lung cancer cases (3). Despite significant development in the diagnosis and treatment of lung cancer, the prognosis of lung cancer remains unfavorable mostly due to a lack of early diagnoses (4). Therefore, in order to better improve the prognosis of lung cancer patients, new prognostic indicators for lung cancer urgently need to be identified. RNA-binding proteins (RBPs) are proteins that can bind and impact a variety of RNAs (5). RBP dysregulation can lead to alteration in RNA metabolism, resulting in a change in the transcriptome and proteome of the cells, further affecting cell growth, proliferation, invasion, and death (6). RBPs can bind and influence RNA to improve the stability of target messenger RNA (mRNAs) and thus promote the expression of genes. Through this mechanism, RBPs perform a crucial role in the development of various diseases (7,8). Thus far, nearly 1,500 RBPs have been reported in the human genome (9). Accumulating evidence indicates that RBPs play an essential role in processes such as preserving the physiological balance of cells, cell development, and stress response (10). Mechanisms of RBPs regulation have been explored in cancer cells, including stability, translation, alternative splicing, subcellular localization, and polyadenylation (11). It has been demonstrated that RBP dysregulation contributes to cancer progression by altering the expression of oncogenes or tumor suppressor genes (12). LIN-28 homolog B (*LIN28B*) promotes versatility through interaction with the let-7 family microRNA, and is pivotal in the pathogenesis of colorectal cancer (13,14). Meanwhile, polypyrimidine tract-binding protein 3 (*PTBP3*) mediates the variable shear of caveolin 1 (*CAV1*), thus affecting the migration and invasion of gastric cancer (15). In lung cancer, the downregulation of cell proliferation splicing factor Quaking (*QKI*) is associated with poor survival rate (16). RBPs play a regulatory role in many steps of gene expression, Li *et al.* have established a predictive model in lung squamous cell carcinoma (LUSC) (17), but few RBPs have been systematically investigated in LUAD thus far.

The purpose of this study was therefore to construct an RBP-based prognosis prediction model for LUAD. We downloaded RNA sequencing data and clinical information for LUAD from The Cancer Genome Atlas, and differentially expressed RNA-binding protein (DERBP) genes were screened. For the purpose of recognizing key RBPs in LUAD that could serve as prognostic markers, we performed a functional analysis of the downloaded data.

We present the following article in accordance with the TRIPOD reporting checklist (available at http://dx.doi.org/10.21037/atm-21-452).

Methods

Screening of differentially expressed RBPs in LUAD and data processing

We obtained the RNA sequencing data of 497 LUAD tumor tissue samples and 54 normal lung tissue samples, along with matching clinical information from TCGA. The P values of differentially expressed genes (DEGs) between the LUAD sample of TCGA was analyzed with Wilcoxon test using the "limma" R package (18). To identify RBPs that were differentially expressed in LUAD, the cutoff threshold in TCGA was set as |log2 fold change (FC)| ≥1.0, and the false discovery rate (FDR) was <0.05. All samples were used as the training cohort. In addition, the external validation cohort consisted of the expression data and comparative clinical data that were acquired from the Gene Expression Omnibus (GEO) database (GSE13213, n=117). This research was conducted in line with the Declaration of Helsinki (as revised in 2013).

Kyoto Encyclopedia of Genes and Genomes pathway and Gene Ontology enrichment analyses

In order to identify the biological function of differentially expressed RBPs, R software was used for Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) analyses. The GO analysis terms included biological process (BP), cellular component (CC), and molecular function (MF). The FDR (q value) <0.05 was regarded as the statistical criterion.

Protein-protein interaction network structuring and key module screening

We submitted 349 RBPs to the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (http://www.string-db.org/). A protein-protein interaction (PPI) network was framed with Cytoscape v3.7.1 software. and its subnetworks were visualized. The MCODE tool in Cytoscape was used to manage the PPI networks, recognize the key networks, and construct a network of 349 nodes and 3,782 edges.

Building a prognostic model

The "survival" package from R software was used to conduct univariate Cox regression analysis for all hub RBPs in the key modules of the training data set. We further screened for significant candidate genes using the log-rank test. Subsequently, based on the initially screened significant candidate genes noted above, a multivariate Cox regression model was built to calculate the risk score to evaluate the prognosis of LUAD patients. The risk score of each LUAD specimen was determined by the following formula:

$$\text{Risk score} = \text{Exp1} \times \text{coe1} + \text{Exp2} \times \text{coe2} + \ldots\ldots + \text{Expi} \times \text{coei} \quad [1]$$

LUAD patients were separated into high-risk and low-risk groups according to the median value of the risk score. Subsequently, a log-rank test was performed to compare the difference of OS between the two subgroups. Furthermore, we used the "survivalROC" package to assess the performance of the prognostic model and drew a receiver operating characteristic (ROC) curve (19). Then, to confirm the predictive power of this prognostic model, 117 LUAD patient samples with credible prognostic data were downloaded from the GSE13213 dataset as a validation cohort.

Nomogram construction and validation of expression level

We used R software to construct a nomogram consisting of nine hub RBPs with the aim of providing a basis for clinicians to evaluate the prognosis of lung cancer patients. The expression level of 9 key RBPs was verified in the Human Protein Atlas (HPA) database (http://www.proteinatlas.org/) (20).

Statistical analysis

Survival curves were constructed by Kaplan-Meier analysis, and the significance of differences was evaluated by log-rank test. Wilcoxon signed-rank test were used to probe into quantitative variables. Significance was defined as $P < 0.05$. ALL statistical analyses were performed using the R software v4.0.3.

Results

Selection and identification of differentially express RBPs in LUAD

The study process is diagrammatized in *Figure 1*. We acquired LUAD data from TCGA database, which included 54 normal lung tissue and 497 tumor tissue samples. The data were processed to find the differently expressed RBP (DERBP) genes by using R software. The RNA sequences of 1,542 RBPs (9) were ultimately included in the analysis; a total of 368 RBPs including 239 upregulated and 129 downregulated genes were deemed to be differentially expressed between normal tissues and those with tumor ($P < 0.05$, |log2 fold change| > 0.5, Table S1). A heat map and volcano plot of the results of the analysis are shown in *Figure 2*.

KEGG pathway and GO enrichment analysis

The enrichment of the identified RBP-encoding genes was evaluated by R software. The KEGG enrichment analysis indicated that the upregulated DERBP genes expression was dramatically enriched in three biological processes: ribosome, RNA transport, and ribosome biogenesis in eukaryotes (*Figure 3A*). Meanwhile, the downregulated DERBP genes expression was significantly enriched in mRNA surveillance pathway, RNA transport, and spliceosome (*Figure 3B*). GO analysis showed that of the upregulated DERBP genes enriched in BP, the majority included non-coding RNA (ncRNA) metabolic processing, ncRNA processing, ribonucleoprotein complex biogenesis, transfer RNA (tRNA) metabolic process, and ribosome biogenesis. Meanwhile, the downregulated DERBP genes were mainly involved in the regulation of translation, regulation of cellular amide metabolic process, regulation of mRNA metabolic process, RNA splicing, and RNA catabolic process. Under CC, we detected that upregulated DERBP genes were observably enriched in preribosome, ribosome, mitochondrial matrix, and ribosomal subunit. Meanwhile, the downregulated DERBP genes were significantly enriched in five biological processes: cytoplasmic ribonucleoprotein granule, ribonucleoprotein granule, nuclear speck spliceosomal complex, and P-body. In the MF analysis, the upregulated DERBP genes were notably enriched in catalytic activity, acting on RNA, catalytic activity, acting on a tRNA, ribonuclease activity, structural constituent of ribosome and nuclease activity, whereas the upregulated DERBP genes were significantly enriched in mRNA 3'-untranslated region (3'UTR) binding, translation regulator activity, translation regulator activity, nucleic acid binding, and single-stranded RNA binding (*Figure 3C,D*).
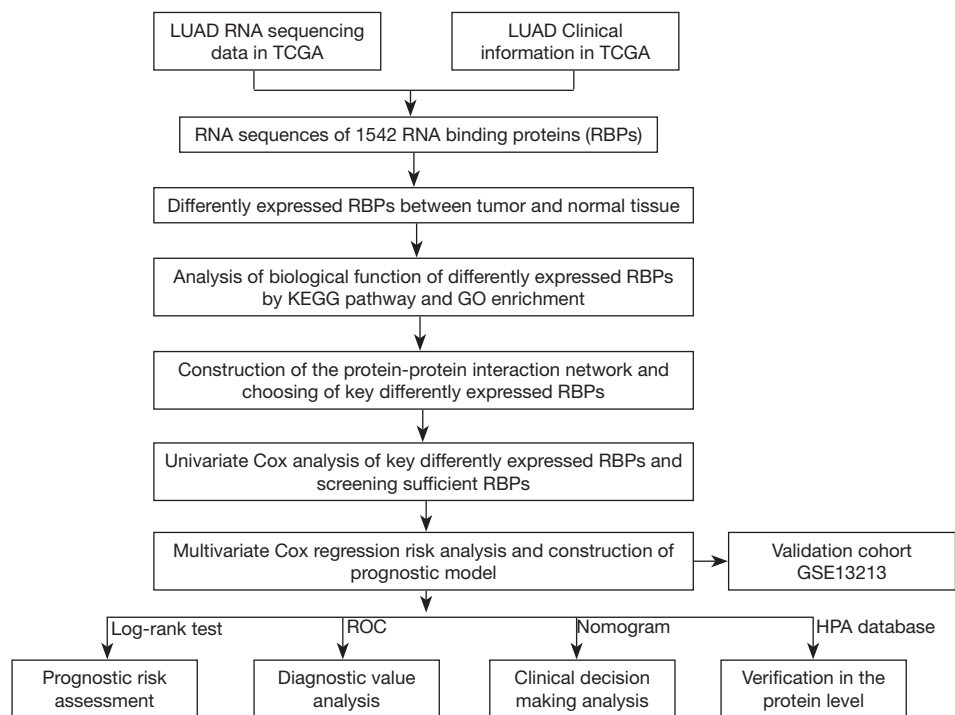
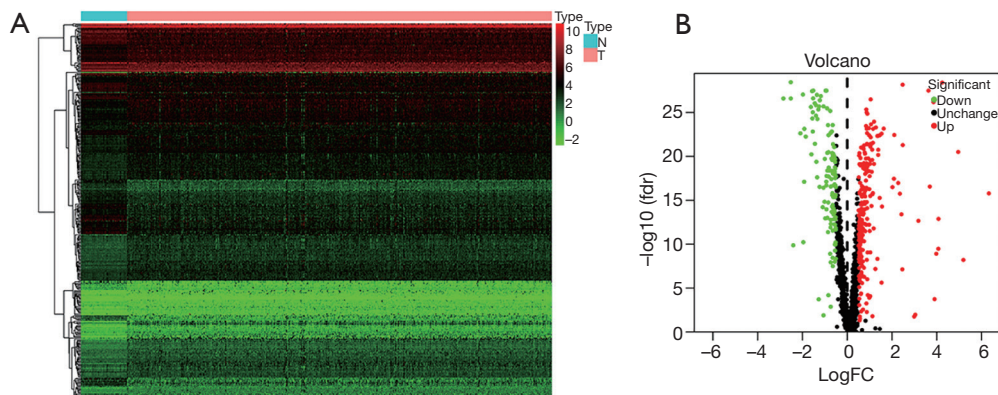Figure 1 Flowchart of lung adenocarcinoma (LUAD) RNA binding proteins (RBPs) analysis.

Figure 2 Differentially expressed RBPs (DERBPs) in LUAD. (A) Heat map of DERBPs between tumor and normal tissues; (B) Volcano plot of DERBPs: upregulated DERBPs are indicated by red dots, and downregulated DERBPs are indicated by green dots. LUAD, lung adenocarcinoma.

PPI network structure and selection of key modules

To further investigate the mutual effect of DERBPs and identify the key RBPs in LUAD, we constructed a PPI network using Cytoscape software. This network, which is based on the data came from the STRING database, contained 349 nodes and 3,782 edges (*Figure 4A*). Next, the MODE tool in Cytoscape was used to manage the PPI

networks and detect potential critical modules. From this, four modules were selected, the most important of which was composed of 44 nodes and 861 edges (*Figure 4B*).

Selection of prognosis-related RBPs

A total of 349 DERBPs were recognized from the PPI network. In order to further explore the prognostic value of these RBP-
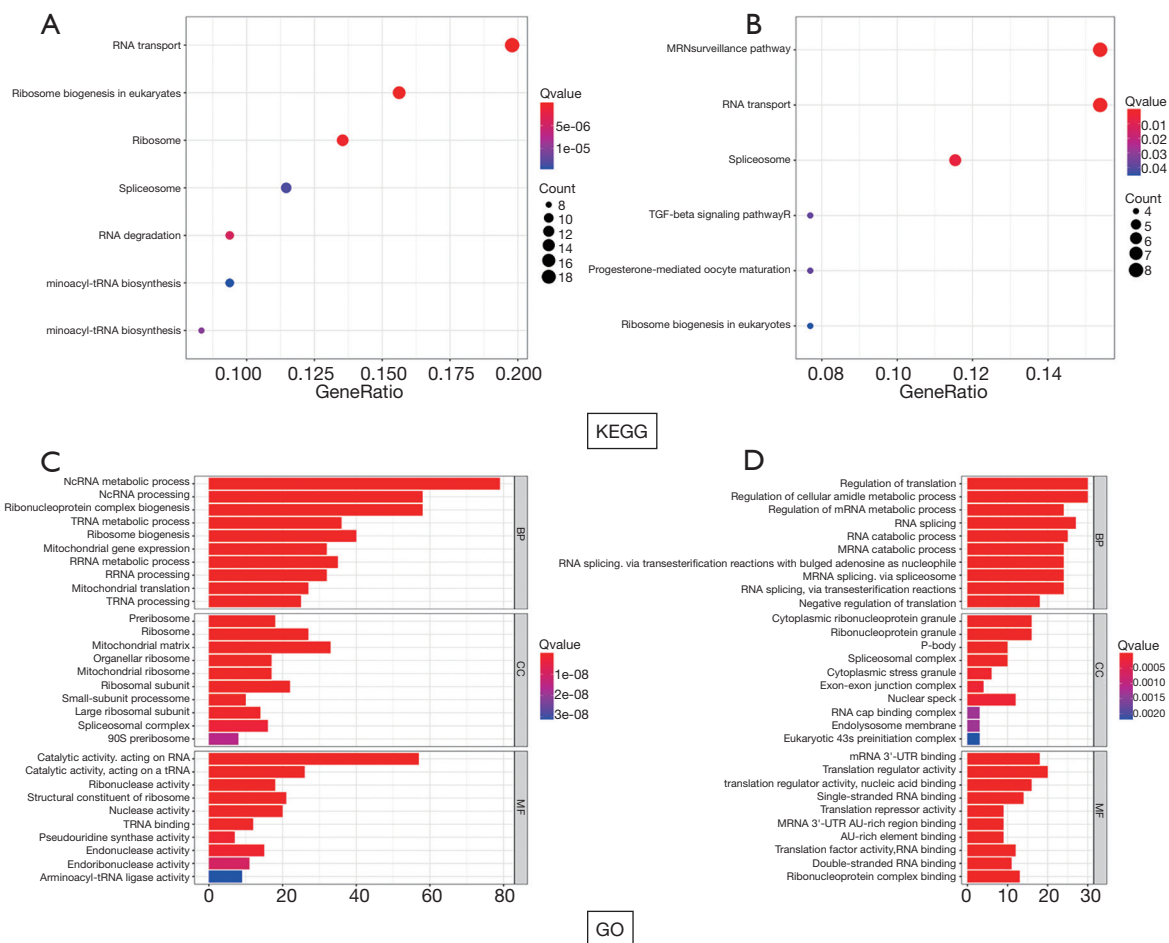
Figure 3 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis and Gene Ontology (GO) enrichment analysis of DERBPs. In KEGG analysis, the abundance of RBP is expressed in terms of dot size. Low q value is represented in red, and high q value is represented in blue. A darker red color displaying a smaller q value, in the same way, darker blue color indicating a higher q value. And the q value of GO terms are illustrated by color bar. BP, biological process; CC, cellular component; MF, molecular function. (A) KEGG pathway analysis of upregulated DERBPs. (B) KEGG pathway analysis of downregulated DERBPs. (C) GO terms of upregulated DERBPs. (D) GO terms of downregulated DERBPs.

related genes, we conducted a univariate Cox regression analysis and acquired 17 candidate hub DERBP genes associated with prognosis (*Figure 5*). Subsequently, a multiple stepwise Cox regression analysis based on these 17 prognostic-associated candidate hub RBPs was performed to determine their influence on the overall survival (OS) of LUAD patients. This yielded nine hub DERBP genes that were independent predictors in LUAD patients (*Figure 6, Table 1*).

Construction and analysis of an RBP-based prognosis

A prediction model was established based on nine key

DERBP genes recognized by multivariate Cox regression analysis. The LUAD patient risk score was calculated by the following formula:

$$\text{Risk score} = (0.23938\times \text{Exp}[\text{CIRBP}]) + (0.24583\times \text{Exp}[\text{DARS2}]) + (-0.38377\times \text{Exp}[\text{DDX24}]) + (0.30115\times \text{Exp}[\text{GAPDH}]) + (0.21273\times \text{Exp}[\text{LARP6}]) + (0.37201\times \text{Exp}[\text{SNRPE}]) + (0.31985\times \text{Exp}[\text{WDR3}]) + (0.41663\times \text{Exp}[\text{ZC3H12C}]) + (-0.41679\times \text{Exp}[\text{ZC3H12D}]).$$ [2]

Next, we performed a survival analysis to evaluate the predictive power of the model. To account for the risk scores, the 497 LUAD patients were separated into high-
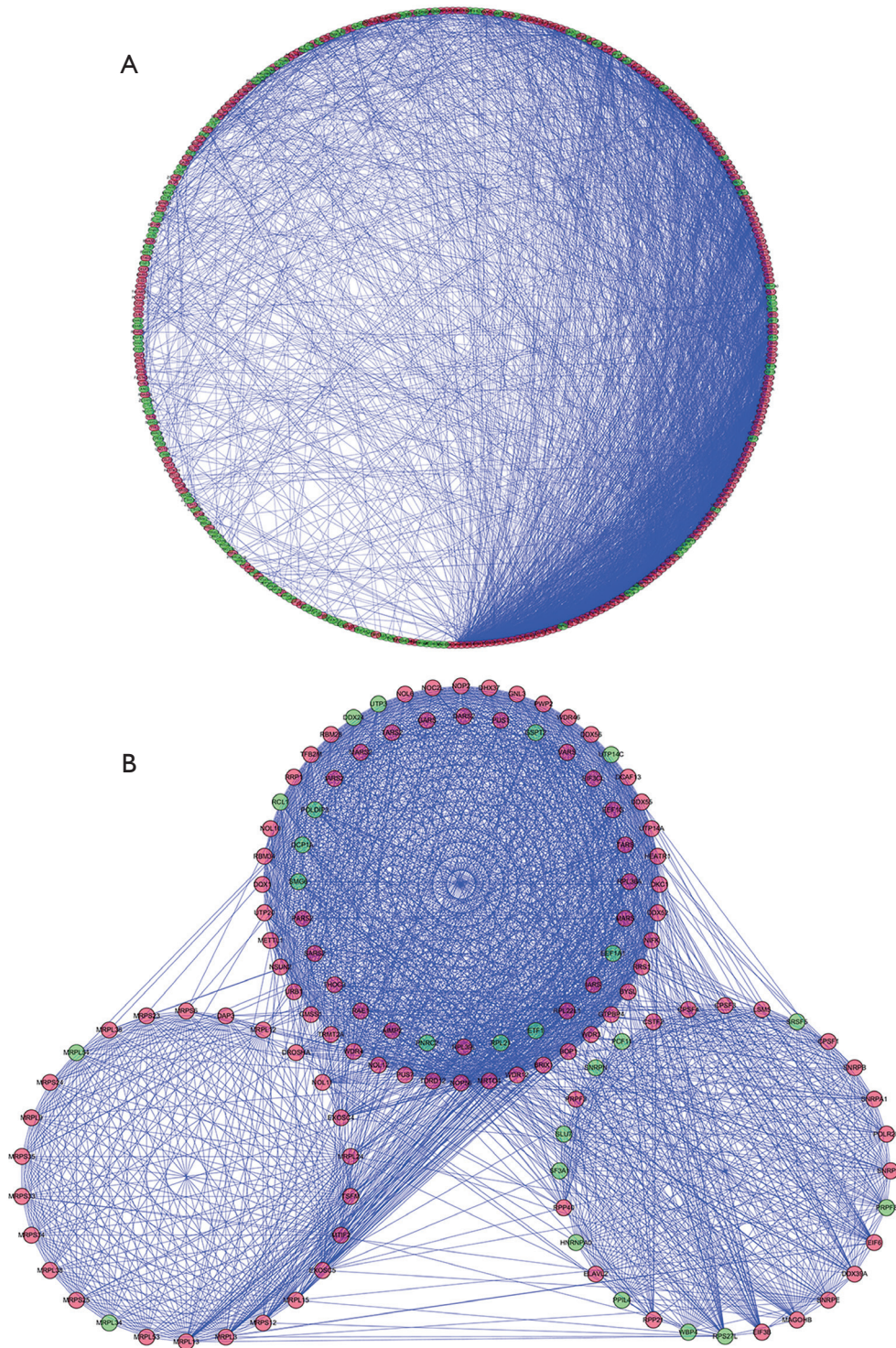
Figure 4 Protein-protein interaction (PPI) network and modules analysis. (A) PPI network of DERBPs; (B) four critical modules from PPI network. Green circles: downregulation. Red circles: upregulation.
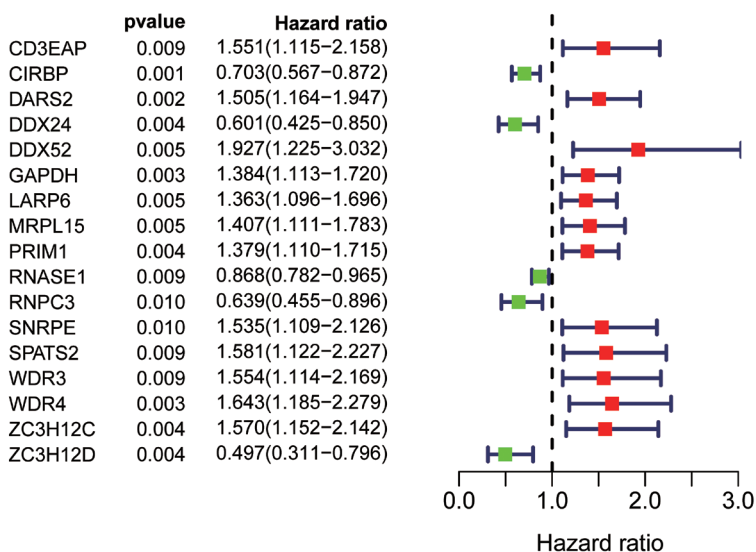
	pvalue	Hazard ratio
CD3EAP	0.009	1.551(1.115–2.158)
CIRBP	0.001	0.703(0.567–0.872)
DARS2	0.002	1.505(1.164–1.947)
DDX24	0.004	0.601(0.425–0.850)
DDX52	0.005	1.927(1.225–3.032)
GAPDH	0.003	1.384(1.113–1.720)
LARP6	0.005	1.363(1.096–1.696)
MRPL15	0.005	1.407(1.111–1.783)
PRIM1	0.004	1.379(1.110–1.715)
RNASE1	0.009	0.868(0.782–0.965)
RNPC3	0.010	0.639(0.455–0.896)
SNRPE	0.010	1.535(1.109–2.126)
SPATS2	0.009	1.581(1.122–2.227)
WDR3	0.009	1.554(1.114–2.169)
WDR4	0.003	1.643(1.185–2.279)
ZC3H12C	0.004	1.570(1.152–2.142)
ZC3H12D	0.004	0.497(0.311–0.796)

Hazard ratio

Figure 5 Forrest plot of univariate Cox regression analyses to identify hub RBPs in LUAD. RBPs, RNA binding proteins; LUAD, lung adenocarcinoma.

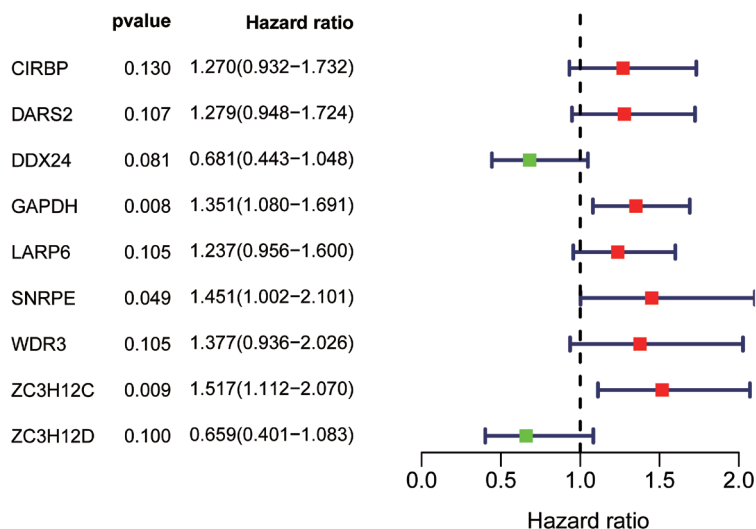|	pvalue	Hazard ratio
CIRBP	0.130	1.270(0.932–1.732)
DARS2	0.107	1.279(0.948–1.724)
DDX24	0.081	0.681(0.443–1.048)
GAPDH	0.008	1.351(1.080–1.691)
LARP6	0.105	1.237(0.956–1.600)
SNRPE	0.049	1.451(1.002–2.101)
WDR3	0.105	1.377(0.936–2.026)
ZC3H12C	0.009	1.517(1.112–2.070)
ZC3H12D	0.100	0.659(0.401–1.083)

Hazard ratio

Figure 6 Forrest plot of multivariate Cox regression analysis for identification of prognosis related hub RBPs.

risk and low-risk subgroups. The results confirmed that the survival curve of the high-risk subgroup was lower than that of the low-risk subgroup (*Figure 7A*). We further performed received operating curve (ROC) analysis to assess the predictive power of the OS-related prognostic models based on the nine RBPs and drew an ROC curve. The results showed that the area under the ROC curve (AUC) of this OS-related prognostic model was 0.707 for the training cohorts (*Figure 7B*), which demonstrated that it had certain

predictive ability. The heat map of RBP expression, the status of patient survival, and the distribution of risk scores are summarized in *Figure 7C,D,E*. Next, we used the same formula to verify the predictive value of the OS-related model in the GSE13213 dataset and evaluated whether the model had similar predictive power in other datasets. We discovered that, in the GSE13213 cohorts, the test group and training cohorts had similar results *Figure 8A,B,C,D,E*. This indicated that the nine DERBP genes-based model

Table 1 Nine prognosis-related hub DERBPs identified by multivariate Cox regression analysis

RBP	Coef	HR	Lower 95% CI	Upper 95% CI	P value
CIRBP	0.239382	1.270463	0.931909	1.732012	0.130036
DARS2	0.245832	1.278685	0.948215	1.724329	0.10709
DDX24	−0.38377	0.681291	0.442956	1.047863	0.080617
GAPDH	0.301149	1.35141	1.080187	1.690734	0.008418
LARP6	0.212729	1.23705	0.956383	1.600082	0.105171
SNRPE	0.372009	1.450647	1.001597	2.101021	0.049021
WDR3	0.319847	1.376916	0.935578	2.026446	0.104755
ZC3H12C	0.416629	1.51684	1.111659	2.069702	0.0086
ZC3H12D	-0.41679	0.659161	0.401293	1.082733	0.099756

(DE)RBPs, differentially expressed RNA-binding proteins; Coef, coefficient; HR, hazard ratio; CI, confidence interval

had good specificity and sensitivity in predicting the prognosis of patients with LUAD.

Building a nine hub DERBP genes-based nomogram

To quantitatively evaluate the effect of the nine hub DERBP genes-based gene signature in the prognostic model, we constructed a nomogram to the predict survival time of LUAD patients (*Figure 9*). From the results of multivariate Cox analysis, we built a nomogram that would be capable of predicting the 1-, 3-, and 5-year OS of LUAD patients using the R package "rms".

We used Cox regression analysis to evaluate the prognostic consequence of the different clinical characteristics of LUAD patients. The results of univariate Cox regression analysis revealed that in both the training cohort and test cohort, the risk score was a significant risk factor for OS and that tumor stage was correlated with the OS of LUAD patients (P<0.01). In contrast, sex and age were irrelevant to OS (*Figure 10*). In short, multivariate Cox regression analysis revealed that risk score and tumor stage were independent risk factors for OS in LUAD patients (P<0.05; *Figure 11*).

Validation the expression of key RBPs

To further investigate the expression of the nine risk-related genes in LUAD patients, we surveyed the protein expression of these DERBPs in the HPA database: *DDX24*, *GAPDH*, *LARP6*, and *ZC3H12D* were overexpressed in LUAD tumor tissue compared to normal tissue. By contrast, the

expression of *WDR3* in LUAD tumor tissue was relatively low (*Figure 12*). Furthermore, the immunoreactivity of *DARS2* and *ZC3H12C* was not significantly different between the LUAD tumor tissue and normal tissue. Information on *CIRBP* and *SNRPE* expression levels is not available in the HPA.

Discussion

An increasing amount of research has confirmed that the dysregulation of RBPs plays a critical role in cancer emergence, development, and progression (6). Nevertheless, merely a fraction of RBPs that contribute to tumorigenesis and progression have been extensively studied and identified (21). In the present study, we identified 368 DERBPs between normal and LUAD tissue of LUAD based on data from TCGA. Through KEGG and GO enrichment analysis, relevant biological pathways were systematically analyzed, and co-expression networks and PPI networks of DERBPs were constructed. In addition, univariate Cox regression and multivariate Cox regression were performed to build a risk prediction model. Survival analysis was implemented to further investigate the model's clinical significance. We established a nomogram based on the nine hub DERBP genes to predict LUAD prognosis. These discoveries may help develop new biomarkers for the diagnosis and prognosis of LUAD patients.

The GO and KEGG pathway analysis demonstrated that the DERBPs were notably enriched in RNA transport, ribosome, ncRNA metabolic process, preribosome, and translation regulator activity. Furthermore, a DERBP
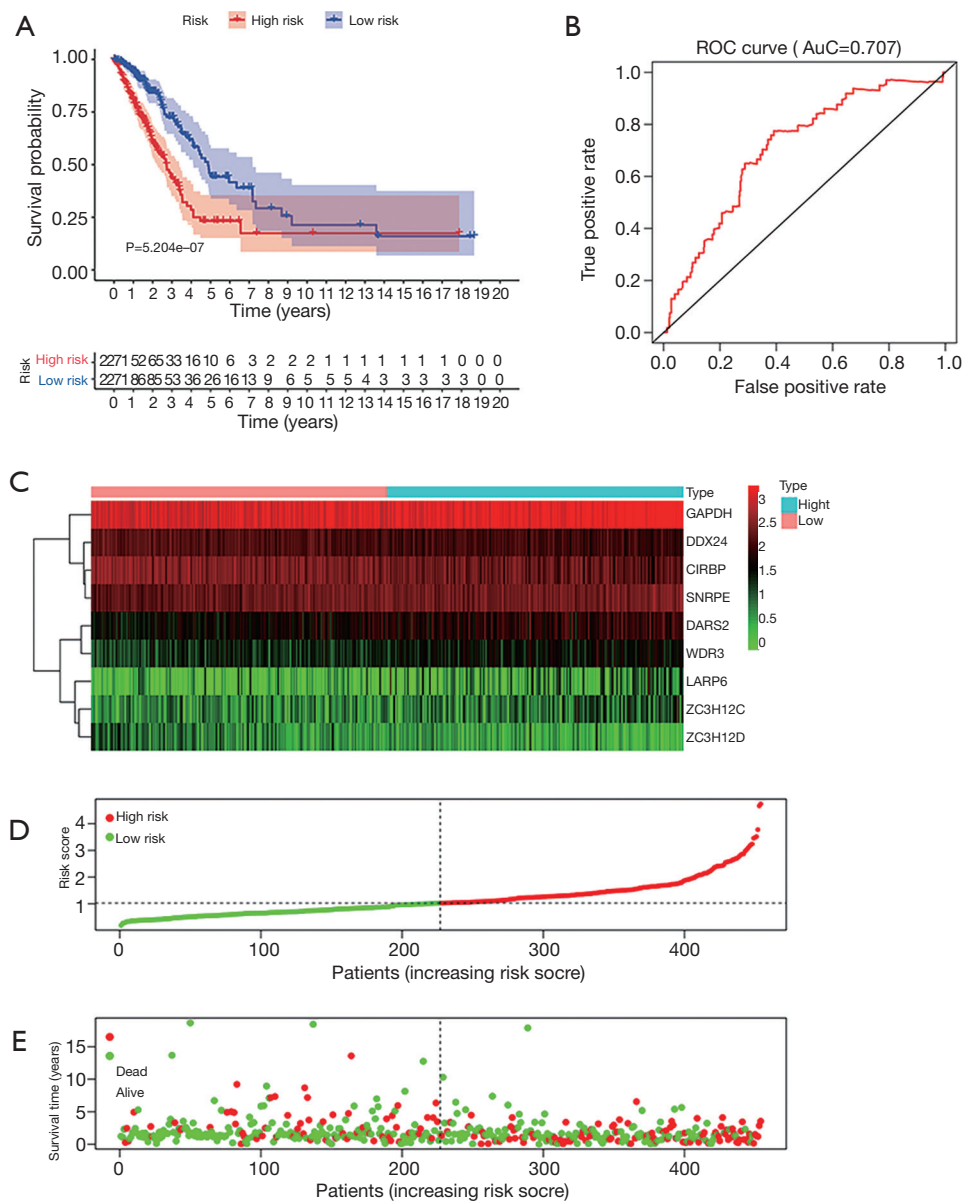
Figure 7 The nine RBP-related gene prediction model associated with overall survival (OS) of LUAD in TCGA dataset. (A) Survival curve for low-risk and high-risk groups in the prediction model; (B) receiver operating characteristic (ROC) analysis of the prediction model; (C) Heat map of 9 RBP gene expression in TCGA dataset; (D) risk score distribution; (E) survival status. The dotted line indicates the median risk score; most of the patients on the right died, showing an increased risk of death as the risk score increased. LUAD, lung adenocarcinoma.

protein interaction network was created, and we obtained modules containing 349 key DERBPs. Some of these hub DERBP genes have been revealed to serve a critical role in the progression of tumors (22,23). For instance, eukaryotic translation initiation factors (EIFs) have been used as a potential therapeutic target for patients with a variety of

tumors. In previous studies, lung cancer patients with high *EIF6* expression were predicted to have poor prognosis (24,25). Moreover, the absence of *BOP1* was discovered to lead to the activation of MAPK pathway, leading to BRAF inhibitor resistance in melanoma (26). *GNL3*, as an oncogene, promotes the progression of osteosarcoma by
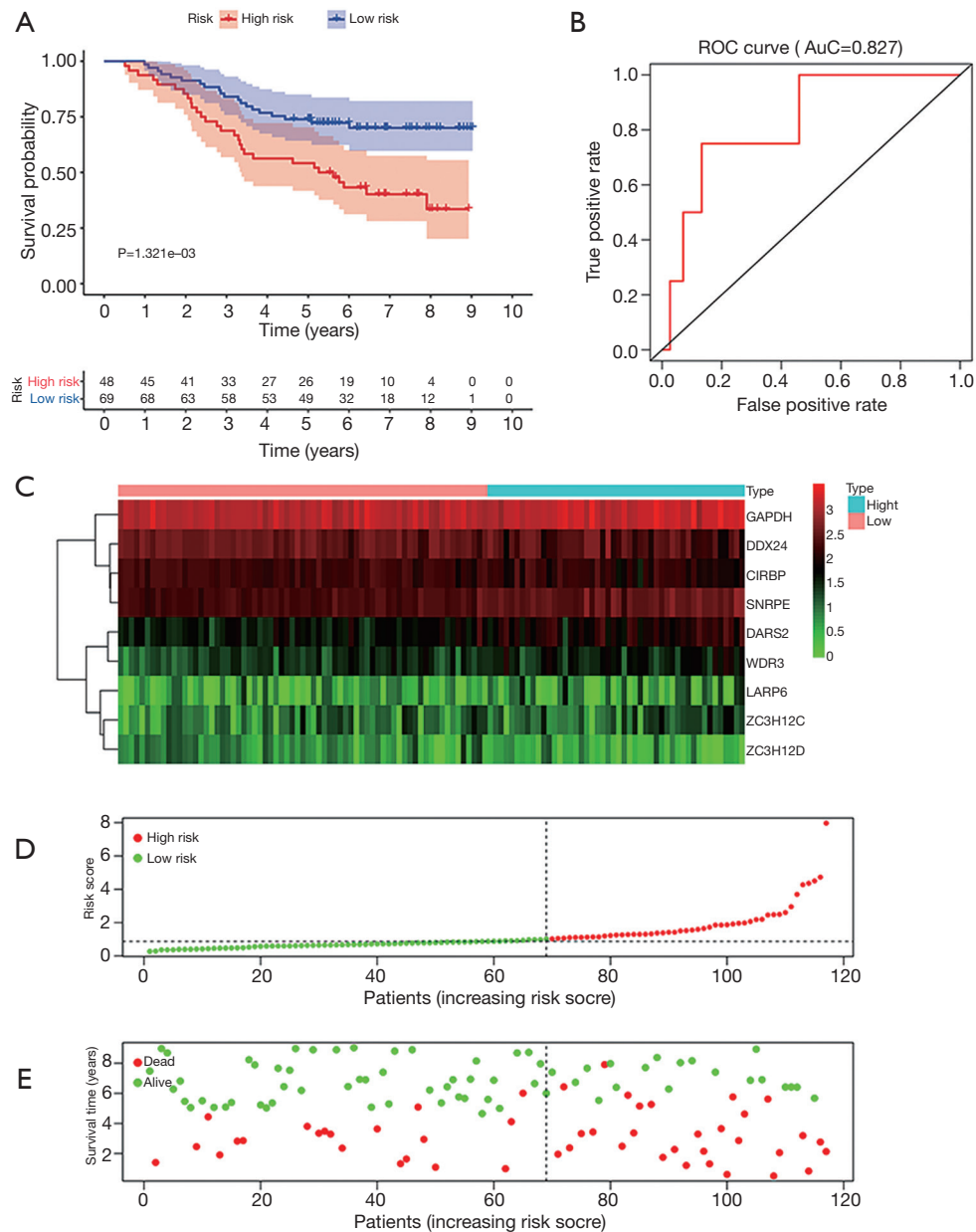
Figure 8 The nine RBP-related gene prediction model associated with OS of LUAD in GSE13213 dataset. (A) Survival curve for low-risk and high-risk groups in the prediction model; (B) ROC analysis of the prediction model; (C) Heat map of 9 RBP gene expression in GSE13213 dataset; (D) risk score distribution; (E) survival status. The dotted line indicates the median risk score; most of the patients on the right died, showing an increased risk of death as the risk score increased. OS, overall survival; LUAD, lung adenocarcinoma.

regulating epithelial-mesenchymal transition (EMT) (27). *BYSL* was found to enhance glioblastoma cell invasion, cell migration, and EMT through the GSK-3β/β-catenin signaling pathway (28). Furthermore, Larp6-mediated mRNA localization was shown to be a key regulator of ribosomal biogenesis during cell migration, while EMT was demonstrated to upregulate the expression of *LARP6* and promote the invasion and metastasis of tumor cells (29). The module analysis of PPI network in our study indicated that ribosome biogenesis in eukaryotes, ncRNA metabolic process, catalytic activity, acting on RNA, and preribosome were associated with LUAD.
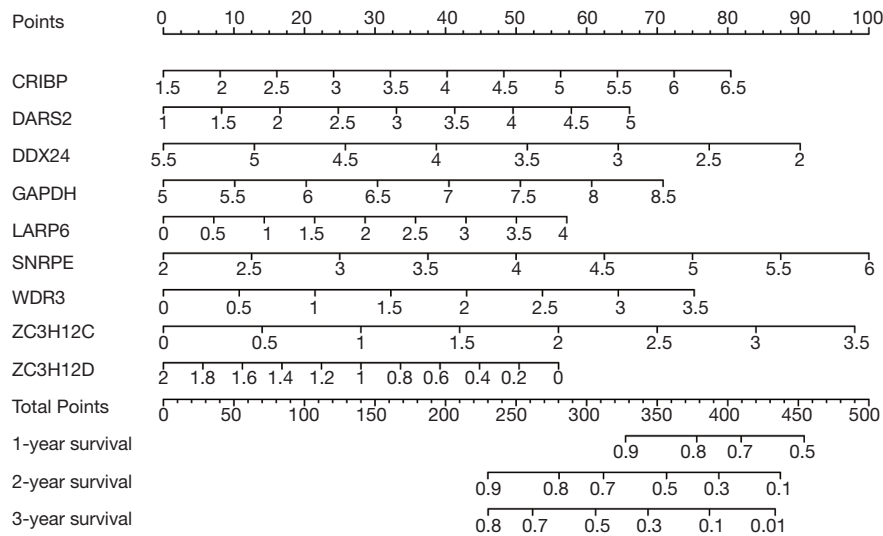
Figure 9 A nomogram was built on the basis of the nine RBP genes for predicting OS of LUAD patients. RBPs, RNA binding proteins; OS, overall survival; LUAD, lung adenocarcinoma.
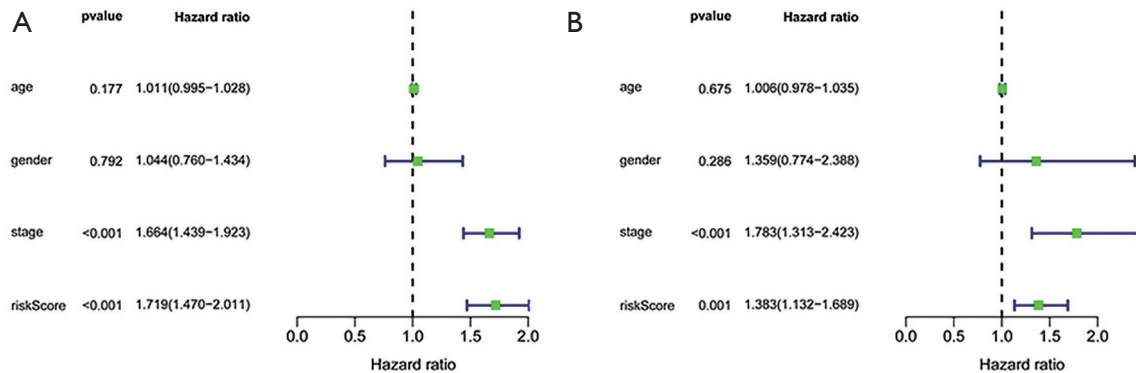

Figure 10 Univariate analysis was performed to assess the clinicopathological prognostic value of the prediction model in LUAD. (A) Training cohort; (B) test cohort. LUAD, lung adenocarcinoma.
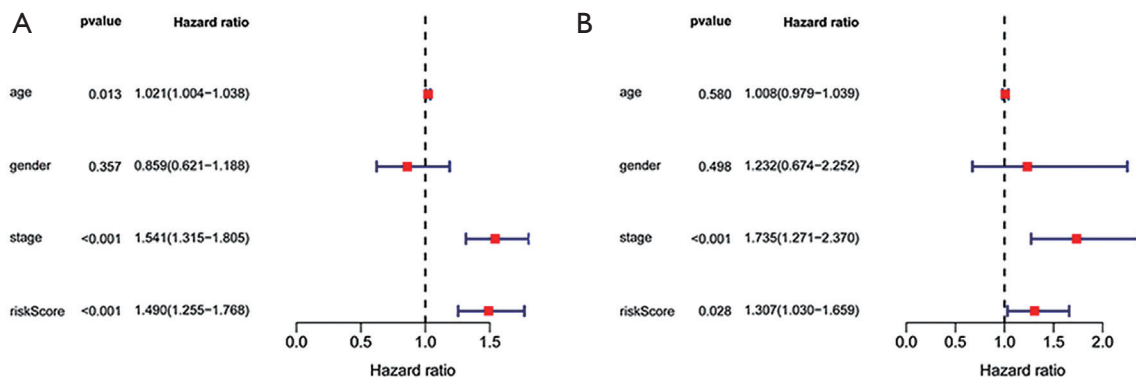

Figure 11 Multivariate analysis was performed to assess the clinicopathological prognostic value of the prediction model in LUAD. (A) Training cohort; (B) test cohort. LUAD, lung adenocarcinoma.
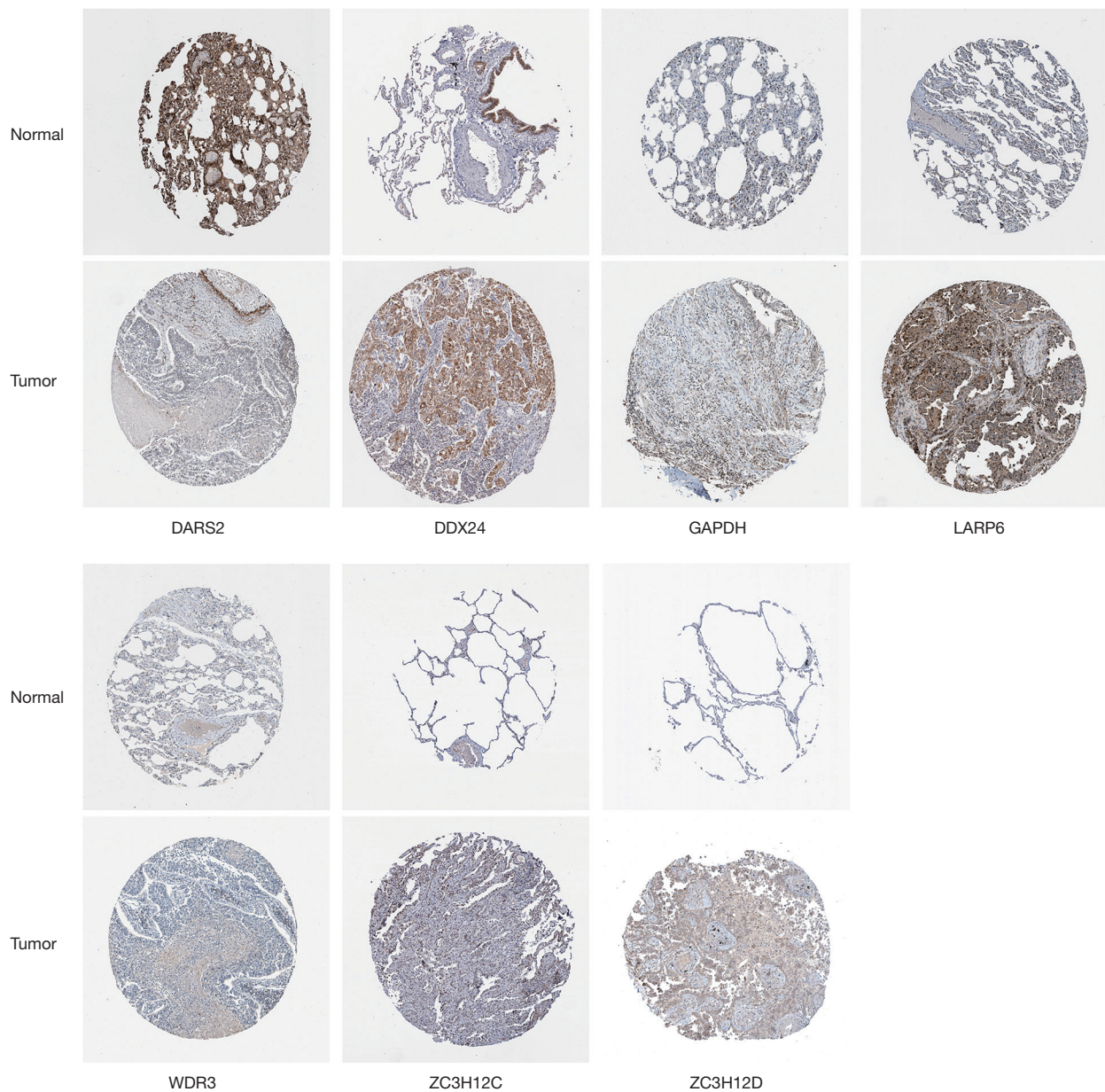
Page 12 of 14

Yang et al. Prediction model for LUAD based on RNA-binding protein

Figure 12 Validation of representative RBP expression of LUAD tumor and normal tissue in the Human Protein Atlas (HPA) database. RBP, RNA binding protein; LUAD, lung adenocarcinoma.

Univariate and multivariate Cox regression analysis identified nine DERBP genes which showed prognostic value in LUAD, including *CIRBP*, *DARS2*, *DDX24*, *GAPDH*, *LARP6*, *SNRPE*, *WDR3*, *ZC3H12C*, and *ZC3H12D*. Previous studies have reported that the expression of *SNRPE* (30,31) and *GAPDH* (32) were relevant to tumor occurrence and development of lung cancer, which is consistent with our finding. Multivariate

Cox regression analysis was then used to establish a risk model based on nine hub DERBPs to predict LUAD prognosis. This model used a TCGA cohort as a training set and a GEO cohort as a testing set. Subsequently, the ROC curves showed that the nine-gene signature had good diagnostic ability and could be used as a biomarker for screening LUAD patients with poor prognosis. However, the mechanisms by which these nine DERBP genes interact

with tumorigenesis have not been clarified, and thus further studies are needed to fully elucidate the underlying mechanisms. Subsequently, we further built nomograms to estimate 1-, 3-, and 5-year OS. Our findings suggested that the nine DERBP genes-based gene model may be applied clinically for the diagnosis, prognostic management, and treatment of patients with LUAD.

However, there were some limitations in this study. First, the generation of prediction models depended on retrospective data from TCGA, which were not verified using clinical samples or prospective clinical study. Moreover, the clinical information acquired from the TCGA was insufficient, and may produce discrepancies in the results. Even so, our nine DERBP genes-related gene model shows immense potential for predicting prognosis in patients with LUAD. It may provide a useful reference for clinicians in determining customized clinical treatment options and providing personalized prognoses.

Conclusions

Bioinformatics analysis was performed to identify the potential prognostic key RPBs correlated with LUAD. The prognostic model we obtained was found to be an independent prognostic factor for LUAD. These nine DERBPs might be involved in LUAD tumor formation, development, and prognosis. Therefore, the nine-gene signature can be used as a marker to predict the postoperative prognosis and tumor progression of LUAD and may serve as novel a strategy for targeted intervention in LUAD.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at http://dx.doi.org/10.21037/atm-21-452

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/atm-21-452). The authors have no conflicts of interest to declare.

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

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020;70:7-30.

2. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature 2018;553:446-54.

3. Duma N, Santana-Davila R, Molina JR. Non-Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. Mayo Clin Proc 2019;94:1623-40.

4. Miller KD, Nogueira L, Mariotto AB, et al. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin 2019;69:363-85.

5. Corley M, Burns MC, Yeo GW. How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. Mol Cell 2020;78:9-29.

6. Mohibi S, Chen X, Zhang J. Cancer the'RBP'eutics-RNA-binding proteins as therapeutic targets for cancer. Pharmacol Ther 2019;203:107390.

7. Siang DTC, Lim YC, Kyaw AMM, et al. The RNA-binding protein HuR is a negative regulator in adipogenesis. Nat Commun 2020;11:213.

8. Li W, Gao L-N, Song P-P, et al. Development and validation of a RNA binding protein-associated prognostic model for lung adenocarcinoma. Aging (Albany NY) 2020;12:3558-73.

9. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. Nat Rev Genet 2014;15:829-45.

Page 14 of 14

Yang et al. Prediction model for LUAD based on RNA-binding protein

10. Masuda K, Kuwano Y. Diverse roles of RNA-binding proteins in cancer traits and their implications in gastrointestinal cancers. Wiley Interdiscip Rev RNA 2019;10:e1520.

11. Garzia A, Morozov P, Sajek M, et al. PAR-CLIP for Discovering Target Sites of RNA-Binding Proteins. Methods Mol Biol 2018;1720:55-75.

12. Pereira B, Billaud M, Almeida R. RNA-Binding Proteins in Cancer: Old Players and New Actors. Trends Cancer 2017;3:506-28.

13. Balzeau J, Menezes MR, Cao S, et al. The LIN28/let-7 Pathway in Cancer. 2017.

14. King CE, Cuatrecasas M, Castells A, et al. LIN28B promotes colon cancer progression and metastasis. Cancer Res 2011;71:4260-8.

15. Liang X, Chen W, Shi H, et al. PTBP3 contributes to the metastasis of gastric cancer by mediating CAV1 alternative splicing. Cell Death Dis 2018;9:569.

16. Zong FY, Fu X, Wei WJ, et al. The RNA-binding protein QKI suppresses cancer-associated aberrant splicing. PLoS Genet 2014;10:e1004289.

17. Li W, Li X, Gao LN, et al. Integrated Analysis of the Functions and Prognostic Values of RNA Binding Proteins in Lung Squamous Cell Carcinoma. Front Genet 2020;11:185.

18. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.

19. Carter JV, Pan J, Rai SN, et al. ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. Surgery 2016;159:1638-45.

20. Thul PJ, Åkesson L, Wiking M, et al. A subcellular map of the human proteome. Science 2017;356:eaal3321.

21. Chen H, Liu J, Wang H, et al. Inhibition of RNA-Binding Protein Musashi-1 Suppresses Malignant Properties and Reverses Paclitaxel Resistance in Ovarian Carcinoma. J Cancer 2019;10:1580-92.

22. Jain A, Brown SZ, Thomsett HL, et al. Evaluation of Post-transcriptional Gene Regulation in Pancreatic Cancer Cells: Studying RNA Binding Proteins and Their mRNA Targets. Methods Mol Biol 2019;1882:239-52.

23. Kim TH, Tsang B, Vernon RM. Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. PHASE SEPARATION 2019.

24. Brina D, Miluzio A, Ricciardi S, et al. eIF6 coordinates insulin sensitivity and lipid metabolism by coupling translation to transcription. Nat Commun 2015;6:8261.

25. Gantenbein N, Bernhart E, Anders I, et al. Influence of eukaryotic translation initiation factor 6 on non-small cell lung cancer development and progression. Eur J Cancer 2018;101:165-80.

26. Gupta R, Bugide S, Wanga B. Loss of BOP1 confers resistance to BRAF kinase inhibitors in melanoma by activating MAP kinase pathway. Proc Natl Acad Sci U S A 2019;116:4583-91.

27. Li T, Li L, Wu X, et al. The oncogenic role of GNL3 in the progression and metastasis of osteosarcoma. Cancer Manag Res 2019;11:2179-88.

28. Sha Z, Zhou J, Wu Y, et al. BYSL Promotes Glioblastoma Cell Migration, Invasion, and Mesenchymal Transition Through the GSK-3beta/beta-Catenin Signaling Pathway. Front Oncol 2020;10:565225.

29. Dermit M, Dodel M, Lee F, et al. Subcellular mRNA Localization Regulates Ribosome Biogenesis in Migrating Cells. Dev Cell 2020;55:298-313.e10.

30. Valles I, Pajares MJ, Segura V, et al. Identification of novel deregulated RNA metabolism-related genes in non-small cell lung cancer. PLoS One 2012;7:e42086.

31. Quidville V, Alsafadi S, Goubar A, et al. Targeting the deregulated spliceosome core machinery in cancer cells triggers mTOR blockade and autophagy. Cancer Res 2013;73:2247-58.

32. Guo C, Liu S, Sun MZ. Novel insight into the role of GAPDH playing in tumor. Clin Transl Oncol 2013;15:167-72.