Identification and Validation of an RNA Binding Proteins-Associated Prognostic Signature for Non-Small Cell Lung Cancer

Ti-wei Miao

sichuan university <https://orcid.org/0000-0003-3705-1568>

Fang-ying Chen

Department of Tuberculosis, the Third People's Hospital of Tibet Autonomous Region, Lhasa, China

Wei Xiao

Respiratory Group, Department of Integrated Traditional Chinese and Western Medicine, West China Hospital, Sichuan University, Chengdu, China

Long-yi Du

Respiratory Group, Department of Integrated Traditional Chinese and Western Medicine, West China Hospital, Sichuan University, Chengdu, China

Bing Mao

Respiratory Group, Department of Integrated Traditional Chinese and Western Medicine, West China Hospital, Sichuan University, Chengdu, China

Yan Wang

Research Core Facility, West China Hospital, Sichuan University, Chengdu, China

Juan-juan Fu (✉ fu.juanjuan@scu.edu.cn)

Respiratory Group, Department of Integrated Traditional Chinese and Western Medicine, West China Hospital, Sichuan University, Chengdu, China <https://orcid.org/0000-0002-7957-1409>

Research

Keywords: RBPs, NSCLC, prognosis, meta-analysis, bioinformatics

Posted Date: November 4th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-100245/v1>

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

[Read Full License](#)

Abstract

Background: Non-small cell lung cancer (NSCLC) is a malignancy with relatively high incidence and poor prognosis. RNA-binding proteins (RBPs) were reported to be dysregulated in multiple cancers and were closely associated with tumor initiation and progression. However, the functions of RBPs in NSCLC remain unclear.

Method: The RNA sequencing data and corresponding clinical information of NSCLC was downloaded from The Cancer Genome Atlas (TCGA) database. We identified aberrantly expressed RBPs between tumor and control tissue, and systemically investigated the expression and prognostic value of these RBPs by a series of bioinformatics analysis.

Results: A total of 459 aberrantly expressed RBPs (291 up-regulated and 168 down-regulated RBPs) were identified. Functional enrichment analysis indicated that the differentially expressed RBPs were mainly associated with RNA splicing, ncRNA metabolic process, regulation of translation, mRNA surveillance pathway, RNA degradation, and RNA transport. Thirteen RBPs (*ZC3H12C*, *ZC3H12D*, *BOP1*, *CASC3*, *DDX24*, *IGF2BP1*, *KHDC1*, *FASTKD3*, *TARBP1*, *INTS7*, *NOL12*, *SNRPB*, *PABPC1L*) were identified as prognostic RBPs by multivariate Cox regression analysis, and were used to construct a prognostic signature. Further analysis demonstrated that high-risk group were significantly related to poor overall survival in training and testing cohort. The area under receiver operator characteristic curve of the prognostic signature was 0.703 in training cohort and 0.636 in testing cohort. In addition, the prognostic signature was further validated in differently clinical subgroup (≥ 65 , < 65 , female, male, stage I-II, III-IV, T1-2, T3-4, N0, N1-3, M0 and M1). The risk score was an independent prognostic factor of NSCLC. A nomogram based on thirteen RBPs was constructed to predict the survival of patients.

Conclusion: Our results provide novel insights into the pathogenesis of NSCLC. The RBPs-associated prognostic signature showed predictive value for NSCLC prognosis, with potential applications in clinical decision-making and individualized treatment.

Introduction

Non-small cell lung cancer (NSCLC) is the leading cause of cancer death worldwide, with an incidence rate of 1.3 million cases per year [1]. NSCLC remains asymptomatic during early stages, and only 25% of patients with NSCLC are diagnosed at this stage [2]. However, about 80% of patients are already in metastatic stage and their 5-year survival rate are below 15% [3]. Despite recent progress in the development of novel medications and immunotherapies in treating NSCLC, the therapeutic efficacy still is unsatisfactory [4]. Therefore, it is an urgent need for effective prediction marker to accurately predict the prognosis of patients with NSCLC.

RNA binding proteins (RBPs) are a group of proteins which could function with the RNA-binding domain to distinguish and bind to target RNAs, including coding RNA and non-coding RNAs (rRNAs, ncRNAs, snRNAs, miRNAs, tRNAs, and snoRNAs) [5]. Up to now, more than 1,500 RBPs have been identified

through high-throughput screening in human genome validated as members of a unique database [6, 7]. RBPs could modulate the fate of binding RNAs by regulating transcription, editing, splicing, polyadenylation, translocation, turnover [8]. In addition, RBPs could also regulate DNA repair [9]. In recent years, studies in genetics and proteomics have shown that most RBPs exhibit functional abnormalities in several cancer. Upregulating RNA-binding protein Quaking could inhibit gastric cancer cell proliferation and metastasis [10]. A cohort of mRNA stability involved in the cell cycle and cell proliferation could be regulated by IGF2BP3 in colorectal cancer[11]. HuR could regulate epithelial to mesenchymal transition (EMT) in pancreatic cancer [12]. CPEB3 could inhibit the epithelial-mesenchymal transition and metastasis of hepatocellular cancer cells by suppressing the translation of MTDH mRNA[13].

Some reports demonstrated that RBPs associated with initiation, progression and prognosis of NSCLC. QKI-5, RALY and KHSRP could promote cancer cell migration, proliferation, and invasion and was associated with metastasis and poorer overall survival in NSCLC patients [14–16]. MSI1 promoted NSCLC malignancy and chemoresistance via activating the Akt signaling pathway [17]. However, RBM47 could inhibit NSCLC metastasis through modulation of AXIN1 mRNA stability and Wnt/ β -catenin signaling[18]. In addition, NONO is highly expressed in lung cancer tissues, and its expression has been correlated with the prognosis of lung cancer patients [19]. However, the role of single RBP for exploring molecular mechanism and predicting prognosis of NSCLC was limited. Thus, a systematic analysis of RBPs may present novel and comprehensive insights into the underlying mechanism during cancer progression and could construct a prognostic signature. For example, in lung squamous cell carcinoma, a nine-RBP gene signature showed predictive value for prognosis, with potential applications in clinical decision-making [20]. In addition. RBPs-associated prognostic signature was also constructed in lung adenocarcinoma, and the high-risk group had poor overall survival (OS) [21, 22]. However, the RBPs associated prognostic signature involved with lung cancer only presents a part of lung cancer patients. Thus, a prognostic signature that can predict most lung cancer patients is lacking.

In the present study, RNA-sequencing data and relevant clinical information of NSCLC was obtained from The Cancer Genome Atlas (TCGA) database (<https://tcga-data.nci.nih.gov/tcga/>) and screened differently expressed RBPs associated with prognosis and explored molecular mechanism in patients with NSCLC. In addition, a prognostic RBPs signature was constructed to predict survival and guide clinical decision-making.

Materials And Methods

Data preprocessing and identification of differentially expressed RBPs

RNA sequencing data of 1037 NSCLC samples and 108 control samples with corresponding clinical information were downloaded from TCGA. A total of 1542 RBPs were included in the current study. Differentially expressed RBPs were identified by R language (Version 4.0.2) based on |log fold change

(FC)| ≥ 0.5 and adjusted P value < 0.05 when two groups were compared. The “Ggplot2 package” and “Gplots package” in R software were used to perform the volcano plots and the heatmap.

GO and KEGG functional enrichment analysis

Enrichment analysis of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway was performed using 459 differentially expressed RBPs by R language. GO mainly depicts molecular function (BP), biological process (CC) and cellular component (MF). $P < 0.05$ was regarded as an accepted threshold criterion.

Identification of prognostic RBP and construction of prognostic RBPs signature

The patients were randomly divided into training cohort and testing cohort with 1:1 ratio. The univariate Cox regression analysis was performed to identify prognostic RBPs which were significantly correlated with survival with $P < 0.05$. Subsequently, the prognostic RBPs signature was constructed by the multivariate Cox regression analysis in training cohort according to a linear combination of RBP expression values multiplied by a regression coefficient (β). The formula was as follows: risk score = expression of RBP 1 $\times \beta_1$ RBP 1 + expression of RBP 2 $\times \beta_2$ RBP 2 +... expression of RBP n $\times \beta_n$ RBP n. The patients were divided into low- or high-risk group according to the median risk score and compared by a log-rank test in training cohort. Receiver operating characteristic (ROC) curve analysis by R package was performed to assess the predictive accuracy of prognostic RBPs signature in training cohort. Furthermore, the univariate Cox regression analysis and multivariable Cox regression analysis were performed for checking whether the risk score was a prognostic factor of NSCLC. In addition, the prognostic signature was validated by the testing cohort and differently clinical subgroup (≥ 65 , < 65 , female, male, stage I-II, III-IV, T1-2, T3-4, N0, N1-3, M0 and M1). A nomogram plot was applied to forecast the likelihood of OS using the rms R package in training cohort.

The level of differentially expressed RBPs in Human Protein Atlas database

The Human Protein Atlas (HPA) online database (<http://www.proteinatlas.org/>) was used to detect the expression of thirteen RBPs at a protein level.

Results

Data acquisition and identification of differently expressed RBPs

The flowchart was displayed in Figure 1. RNA sequencing data of 1145 samples (108 control and 1037 NSCLC) and clinical information of 1027 NSCLC patients including the gender, age, tumor stage, survival time and survival status were downloaded from TCGA. The NSCLC patients with incomplete follow-up information were deleted. Finally, 996 patients were included for further analysis and the baseline characteristics of 996 NSCLC patients were listed in Table 1. A total of 459 differentially expressed RBPs

(291 up-regulated and 168 down-regulated RBPs) were identified when NSCLC compared to control (Figure 2).

Functional enrichment analysis

Functional enrichment analysis was performed with 459 differentially expressed RBPs. The biological process was primarily associated with ncRNA metabolic process, ribonucleoprotein complex biogenesis, RNA splicing, regulation of translation, RNA catabolic process (Figure 3). The cellular component mainly located in cytoplasmic ribonucleoprotein granule, ribonucleoprotein granule, spliceosomal complex (Figure 3). The molecular function was mainly related to catalytic activity, acting on RNA, translation regulator activity, mRNA 3'-UTR binding (Figure 3). The KEGG pathway was mainly enriched in RNA transport, Ribosome biogenesis in eukaryotes, mRNA surveillance pathway (Figure 4).

Identification of prognostic RBP and construction of prognostic RBPs signature

38 prognostic RBPs were identified from 459 differentially expressed RBPs by univariate Cox regression analysis (Figure 5). Subsequently, 13 RBPs (*ZC3H12C*, *ZC3H12D*, *BOP1*, *CASC3*, *DDX24*, *IGF2BP1*, *KHDC1*, *FASTKD3*, *TARBP1*, *INTS7*, *NOL12*, *SNRPB*, *PABPC1L*) were verified as independent indicator of prognosis by multivariable Cox regression analysis in training cohort (Figure 6), and a prognostic risk score formula was established by multivariate Cox regression analysis: Risk score = $(0.166 \times \text{expression of } ZC3H12C) + (-0.295 \times \text{expression of } ZC3H12D) + (0.015 \times \text{expression of } BOP1) + (0.037 \times \text{expression of } CASC3) + (0.037 \times \text{expression of } DDX24) + (0.045 \times \text{expression of } IGF2BP1) + (-0.423 \times \text{expression of } KHDC1) + (-0.089 \times \text{expression of } FASTKD3) + (-0.064 \times \text{expression of } TARBP1) + (0.045 \times \text{expression of } INTS7) + (-0.130 \times \text{expression of } NOL12) + (0.001 \times \text{expression of } SNRPB) + (0.041 \times \text{expression of } PABPC1L)$. The Kaplan–Meier survival analysis was constructed according to risk score and survival time and status in training cohort. The high-risk group has a shorter survival time compared to the low-risk group ($P < 0.001$; Figure 7A) and area under ROC curve was 0.703 (Figure 7B). All patients were divided into low- or high-risk group according to the median risk score (Figure 8A). The survival time of each patient was shown in Figure 8B and the patients with a high-risk score had a higher mortality rate than those with a low-risk score. A heatmap was shown to present the different expression profile of the thirteen RBPs between two groups (Figure 8C).

Risk score was an independent prognostic factor of NSCLC

The clinical features included risk score, age, sex, TNM stage were considered as covariables to screen independent prognostic factor. The risk score and tumor stage in training cohort were associated with patient prognosis in the univariate analysis (hazard ratio [HR]: 1.741; 95% confidence interval [CI]: 1.522–1.991; $P < 0.001$. HR: 1.521; 95% CI: 1.305–1.773; $P < 0.001$) and the multivariate analysis (HR: 1.779; 95% CI: 1.548–2.045; $P < 0.001$. HR: 1.578; 95% CI: 1.343–1.854; $P < 0.001$) (Figure 9), which indicates the risk score could independently predict the prognosis of patients with NSCLC.

Validation of prognostic RBPs signature in testing cohort

To validate the accuracy of the prognostic signature, it was further assessed in testing cohort. Compared to the low-risk group, the high-risk group has a poor survival time ($P < 0.001$; Figure 10A) and area under ROC curve was 0.636 (Figure 10B). The risk score, survival status and RBPs expression data of NSCLC patients in the testing cohort were illustrated in Figure 11. In addition, the risk score and tumor stage in testing cohort were also independently associated with prognosis in the univariate analysis (HR: 1.377; 95% CI: 1.179–1.607; $P < 0.001$. HR: 1.458; 95%CI: 1.245–1.708; $P < 0.001$) and the multivariate analysis (HR: 1.332; 95% CI: 1.134–1.566; $P < 0.001$; HR: 1.444; 95% CI: 1.223–1.705; $P < 0.001$) (Figure 12). The results in testing cohort consisted with the results in training cohort. In order to further validate the accuracy of prognostic RBPs signature, it was applied to different age (≥ 65 and < 65), gender (female and male), TNM stage (stage I-II and III-IV), T stage (T1-2 and T3-4), N stage (N0 and N1-3), M stage (M0 and M1) by the Kaplan–Meier curves analysis in the training cohort. The high-risk group in all subgroup above were significantly associated with poor survival time compared to low-risk group (Figure 13). Thus, a RBPs-associated prognostic signature could predict prognosis of NSCLC patients with superior accuracy.

Nomogram predicting model

In order to construct a quantitative model for NSCLC prognosis, thirteen prognostic RBPs were combined to build a nomogram plot (Figure 14). This allowed us to calculate the estimated survival probabilities of NSCLC patients at 1, 3 and 5 years by plotting a vertical line between the total point axis and each prognosis axis.

The level of differentially expressed RBPs in Human Protein Atlas database

To determine the expression of 13 differentially expressed RBPs between NSCLC and control, the immunohistochemistry results of 10 RBPs (ZC3H12C, ZC3H12D, BOP1, CASC3, DDX24, IGF2BP1, FASTKD3, TARBP1, NOL12, SNRPB,) were downloaded from the Human Protein Atlas database due to KHDC1, INTS7 and PABPC1L without immunohistochemistry results in this database (Figure 15).

Discussion

In recent years, a multiple of studies have suggested the crucial roles of RBPs in the development and progression of various human cancers [23–25]. However, a comprehensive analysis of the potential function of RBPs on NSCLC tumorigenesis and prognosis has not been performed. In present study, A total of 459 differentially expressed RBPs were identified between cancer and control tissue. Functional enrichment analysis was performed by 459 differentially expressed RBPs. Thirteen RBPs were independently identified as prognostic RBPs by multivariate Cox regression analysis and were used to construct the prognostic RBPs signature. The high-risk group in this prognostic RBPs signature was associated with a shorter survival time. Furthermore, the risk score was an independent prognostic factor of NSCLC. In addition, the signature was validated by the testing cohort and differently clinical subgroup. Therefore, A RBPs-associated prognostic signature was successfully constructed with highly diagnostic value, which could apply to predict prognosis of patients with NSCLC and guide clinical decision-making.

The biological process in NSCLC was primarily associated with RNA splicing, ncRNA metabolic process, ribonucleoprotein complex biogenesis, regulation of translation, RNA catabolic process. Abnormality in RNA splicing have been associated with tumor initiation and progression [26, 27]. The ribonucleoprotein complex involved in initiation, progression and prognosis of cancer [28, 29]. In addition, in cancer cells, lncRNA were aberrantly expressed as classical oncogenes or tumor suppressors and correlate with the altered metabolism [30, 31]. Given the direct control of translation by tRNA modifications, tRNA modifying enzymes may function as regulators of cancer progression [32]. KEGG pathways were mainly enriched in mRNA surveillance pathway, RNA degradation, RNA transport, Ribosome biogenesis in eukaryotes. The previous research has reported that mRNA surveillance pathway has involved with initiation and progression of cancer [33–35]. RBPs were associated with tumorigenicity and metastasis involved in mRNA stability [11, 18]. In addition, the RNA degradation pathway was involved in drug anti-oral tumorigenesis [36]. Alterations of crucial pathways underlying mRNA nuclear export could result in genome instability and cancer initiation [37]. Given the important function of ribosome and ribosomal proteins (RPs) in cells, ribosomal dysfunction would result in a variety of diseases [38–41] and RPL15, a large ribosomal subunit protein, was remarkably upregulated in human cancer tissues and cultured cell lines and was closely correlated with clinicopathological characteristics [42]. However, in NSCLC, those molecular mechanisms and pathways need further explore in the future.

Several RBPs in our prognostic signature had been studied in human disease, especially in cancer. IGF2BP1 silencing could inhibit the occurrence of tumor event in NSCLC [43]. In addition, the CASC3 overexpressed in cancer tissue and promoted cancer cell proliferation, migration, invasion, and glycolysis [44, 45]. ZC3H12C could inhibit vascular inflammation and proinflammatory gene expression in endothelial cells [46] and regulate proinflammatory activation of macrophages [47, 48]. ZC3H12D contributed to inflammation-associated signaling pathways involved in the pathogenesis of Leukoaraisosis [49]. BOP1 could prompt cell migration and invasion in colorectal cancer [50] and loss of BOP1 conferred resistance to BRAF kinase inhibitors in melanoma [51]. DDX24 negatively regulated cytosolic RNA-mediated innate immune signaling [52]. FASTKD3 was an essential component of mitochondrial respiration that may modulate energy balance in cells exposed to adverse conditions by functionally coupling mitochondrial protein synthesis to respiration [53]. TARBP1 was overexpressed in NSCLC and hepatocellular cancer, and was associated with pathological grade, clinical stage and pathological type [54, 55]. Increased levels of INTS7 in prostate cancer was associated with an aggressive phenotype [56]. NOL12 was a multifunctional RNA binding protein at the nexus of RNA and DNA metabolism [57]. NOL12 repression induced nucleolar stress-driven cellular senescence and was associated with normative aging [58]. PABPC1L might promote colon tumorigenesis by regulating mRNA splicing [59]. In addition, PABPC1L depletion inhibited proliferation and migration via blockage of AKT pathway in colorectal cancer [60]. In current bioinformatic analysis, the level of ZC3H12D, BOP1, IGF2BP1, KHDC1, FASTKD3, TARBP1, INTS7, NOL12, SNRPB, PABPC1L mRNA and protein in lung cancer tissue were higher than that in control tissue and were associated with survival time. However, the level of ZC3H12C, CASC3, DDX24 mRNA and protein in lung cancer tissue were lower compared with control tissue and were also associated with survival time. Therefore, the thirteen RBPs involved in tumorigenicity

and prognosis of NSCLC, which deserved further explore the molecular mechanism in vitro and in vivo experiments in the future.

The previous research has demonstrated that prognostic RBPs signature could predict the prognosis of several of cancer [61, 62]. However, in lung cancer, the previous prognostic RBPs signature was constructed based on lung squamous cell carcinoma or lung adenocarcinoma [20–22], which only predict a part of patients. A prognostic RBPs signature was constructed by multivariable Cox regression analysis concluding 13 differentially expressed RBPs (*ZC3H12C*, *ZC3H12D*, *BOP1*, *CASC3*, *DDX24*, *IGF2BP1*, *KHDC1*, *FASTKD3*, *TARBP1*, *INTS7*, *NOL12*, *SNRPB*, *PABPC1L*) in the current study, which could predict most lung cancer patients. The high-risk group had a shorter survival time compared to the low-risk group in training and testing cohort (both $P < 0.001$). Area under ROC curve was 0.703 in training cohort and 0.636 in testing cohort with moderate accuracy and specificity. In addition, risk score was an independent prognostic factor of NSCLC by the univariate Cox analysis and the multivariate Cox analysis. Furthermore, the prognostic RBPs signature was applied to differently clinical subgroup (≥ 65 , < 65 , female, male, stage I-II, III-IV, T1-2, T3-4, N0, N1-3, M0 and M1), and high-risk group in all subgroup has a shorter survival time compared with the low-risk group, which further validated the prognostic RBPs signature. Therefore, the RBPs-associated prognostic signature was successfully constructed. In addition, A nomogram was established to predict 1-year, 3-year and 5-year survival of patients with NSCLC.

There were some highlights in current study. Our study provided innovative insights into the roles of RBPs during NSCLC development and progression and established a promising prognostic signature to evaluate patient prognosis. In addition, the prognostic RBPs signature was validated by the testing cohort and differently clinical subgroup, and had a good predict value. Nevertheless, this study had several limitations which could not be ignored. First, our results were based on the bioinformatics analysis of the TCGA data, which need to be validated in clinical patient cohort and multicenter and prospective study. Second, the prognostic RBPs was not validated by qPCR. Finally, further studies were required to clarify the molecular mechanism for clinical practice in the future.

Conclusion

A total of 459 differentially expressed RBPs were identified between cancer and control tissue. Functional enrichment analysis was mainly associated with RNA splicing, regulation of translation, RNA catabolic process, mRNA surveillance pathway, RNA degradation, RNA transport. Thirteen RBPs were independently identified as prognostic RBPs by multivariate Cox regression analysis and were used to construct the prognostic RBPs signature. The high-risk group in this prognostic RBPs signature was associated with a shorter survival time. Furthermore, the risk score was an independent prognostic factor of NSCLC. In addition, the signature was validated by the testing cohort and differently clinical subgroup (≥ 65 , < 65 , female, male, stage I-II, III-IV, T1-2, T3-4, N0, N1-3, M0 and M1). Therefore, a RBPs-associated prognostic signature was successfully constructed with highly diagnostic value, which could apply to predict prognosis of patients with NSCLC and guide clinical decision-making.

Abbreviations

NSCLC: Non-small cell lung cancer; RBPs:RNA-binding proteins; TCGA:The Cancer Genome Atlas; EMT:epithelial to mesenchymal transition; FC:fold change; GO:Gene Ontology; KEGG:Kyoto Encyclopedia of Genes and Genomes; BP:molecular function; CC:biological process ; MF:cellular component; ROC:Receiver operating characteristic; HPA:Human Protein Atlas; HR:hazard ratio; CI:confidence interval.

Declarations

Ethics approval and consent to participate

Current research did not involve in ethical issues due to all the data from TCGA public database

Consent for publication

Not applicable.

Availability of data and materials

The datasets involved in our study were extracted from TCGA (<https://portal.gdc.cancer.gov/>), and HPA (<https://www.proteinatlas.org/>). All the data we used were available in our study.

Competing interests

No potential conflicts of interest are to be disclosed by authors.

Funding

This study was funded the Thousand Talents Program, Sichuan province, China [0082204151045] and the National Natural Science Foundation of China [No.81870014].

Acknowledgments

The data of this study was download from TCGA and HPA, we gratefully acknowledge the patients and operations.

Availability of data and materials

The datasets involved in our study were extracted from TCGA (<https://portal.gdc.cancer.gov/>), and HPA (<https://www.proteinatlas.org/>). All the data we used were available in our study.

Authors' contributions

Ti-wei Miao and Fang-ying Chen conceived and designed the study and analyzed the data and wrote the manuscript. Wei Xiao and Long-yi Du prepared the figures and tables and designed the study. Yan Wang

revised the figure. Bing Mao and Juan-juan Fu designed the study and revised the figures, tables and manuscript.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. 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. doi: 10.3322/caac.21492.
2. Ni M, Liu X, Wu J, et al. Identification of Candidate Biomarkers Correlated With the Pathogenesis and Prognosis of Non-small Cell Lung Cancer via Integrated Bioinformatics Analysis. *Front Genet.* 2018; 9: 469. doi: 10.3389/fgene.2018.00469.
3. Goldstraw P, Chansky K, Crowley J, et al. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol.* 2016; 11: 39-51. doi: 10.1016/j.jtho.2015.09.009.
4. Facchinetti F, Marabelle A, Rossi G, et al. Moving Immune Checkpoint Blockade in Thoracic Tumors beyond NSCLC. *J Thorac Oncol.* 2016; 11: 1819-36. doi: 10.1016/j.jtho.2016.05.027.
5. Hong S. RNA Binding Protein as an Emerging Therapeutic Target for Cancer Prevention and Treatment. *J Cancer Prev.* 2017; 22: 203-10. doi: 10.15430/jcp.2017.22.4.203.
6. Cook KB, Kazan H, Zuberi K, et al. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.* 2011; 39: D301-8. doi: 10.1093/nar/gkq1069.
7. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. *Nat Rev Genet.* 2014; 15: 829-45. doi: 10.1038/nrg3813.
8. Müller-McNicoll M, Neugebauer KM. How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nat Rev Genet.* 2013; 14: 275-87. doi: 10.1038/nrg3434.
9. Vågbø CB, Slupphaug G. RNA in DNA repair. *DNA Repair (Amst).* 2020; 95: 102927. doi: 10.1016/j.dnarep.2020.102927.
10. He Z, Duan Z, Chen L, et al. Long non-coding RNA Loc490 inhibits gastric cancer cell proliferation and metastasis by upregulating RNA-binding protein Quaking. *Aging (Albany NY).* 2020; 12. doi: 10.18632/aging.103876.
11. Li K, Huang F, Li Y, et al. Stabilization of oncogenic transcripts by the IGF2BP3/ELAVL1 complex promotes tumorigenicity in colorectal cancer. *Am J Cancer Res.* 2020; 10: 2480-94. doi: 10.1158/1535-7163.mct-19-0822.
12. Dong R, Chen P, Polireddy K, et al. An RNA-binding protein, Hu-antigen R, in pancreatic cancer epithelial to mesenchymal transition, metastasis, and cancer stem cells. *Mol Cancer Ther.* 2020. doi: 10.1158/1535-7163.mct-19-0822.
13. Zhang H, Zou C, Qiu Z, et al. CPEB3-mediated MTDH mRNA translational suppression restrains hepatocellular carcinoma progression. *Cell Death Dis.* 2020; 11: 792. doi: 10.1038/s41419-020-02984-y.

14. Liang G, Meng W, Huang X, et al. miR-196b-5p-mediated downregulation of TSPAN12 and GATA6 promotes tumor progression in non-small cell lung cancer. *Proc Natl Acad Sci U S A*. 2020; 117: 4347-57. doi: 10.1073/pnas.1917531117.
15. Song G, Guo G, Du T, et al. RALY may cause an aggressive biological behavior and a dismal prognosis in non-small-cell lung cancer. *Exp Cell Res*. 2020; 389: 111884. doi: 10.1016/j.yexcr.2020.111884.
16. Yan M, Sun L, Li J, et al. RNA-binding protein KHSRP promotes tumor growth and metastasis in non-small cell lung cancer. *J Exp Clin Cancer Res*. 2019; 38: 478. doi: 10.1186/s13046-019-1479-2.
17. Lang Y, Kong X, He C, et al. Musashi1 Promotes Non-Small Cell Lung Carcinoma Malignancy and Chemoresistance via Activating the Akt Signaling Pathway. *Cell Physiol Biochem*. 2017; 44: 455-66. doi: 10.1159/000485012.
18. Shen DJ, Jiang YH, Li JQ, et al. The RNA-binding protein RBM47 inhibits non-small cell lung carcinoma metastasis through modulation of AXIN1 mRNA stability and Wnt/ β -catentin signaling. *Surg Oncol*. 2020; 34: 31-9. doi: 10.1016/j.suronc.2020.02.011.
19. Kim SJ, Ju JS, Park SS, et al. An RNA-binding-protein, NONO governs energy metabolism by regulating NAMPT in lung cancer. *Biochem Biophys Res Commun*. 2020; 528: 376-82. doi: 10.1016/j.bbrc.2020.01.011.
20. 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. doi: 10.3389/fgene.2020.00185.
21. Li W, Gao LN, Song PP, et al. Development and validation of a RNA binding protein-associated prognostic model for lung adenocarcinoma. *Aging (Albany NY)*. 2020; 12: 3558-73. doi: 10.18632/aging.102828.
22. Li W, Li N, Gao L, et al. Integrated analysis of the roles and prognostic value of RNA binding proteins in lung adenocarcinoma. *PeerJ*. 2020; 8: e8509. doi: 10.7717/peerj.8509.
23. Kudinov AE, Karanicolas J, Golemis EA, et al. Musashi RNA-Binding Proteins as Cancer Drivers and Novel Therapeutic Targets. *Clin Cancer Res*. 2017; 23: 2143-53. doi: 10.1158/1078-0432.ccr-16-2728.
24. Pereira B, Billaud M, Almeida R. RNA-Binding Proteins in Cancer: Old Players and New Actors. *Trends Cancer*. 2017; 3: 506-28. doi: 10.1016/j.trecan.2017.05.003.
25. Dong W, Dai ZH, Liu FC, et al. The RNA-binding protein RBM3 promotes cell proliferation in hepatocellular carcinoma by regulating circular RNA SCD-circRNA 2 production. *EBioMedicine*. 2019; 45: 155-67. doi: 10.1016/j.ebiom.2019.06.030.
26. Adamia S, Haibe-Kains B, Pilarski PM, et al. A genome-wide aberrant RNA splicing in patients with acute myeloid leukemia identifies novel potential disease markers and therapeutic targets. *Clin Cancer Res*. 2014; 20: 1135-45. doi: 10.1158/1078-0432.ccr-13-0956.
27. Sun L, Goodman PA, Wood CM, et al. Expression of aberrantly spliced oncogenic ikaros isoforms in childhood acute lymphoblastic leukemia. *J Clin Oncol*. 1999; 17: 3753-66. doi: 10.1200/jco.1999.17.12.3753.

28. Ke RS, Zhang K, Lv LZ, et al. Prognostic value and oncogene function of heterogeneous nuclear ribonucleoprotein A1 overexpression in HBV-related hepatocellular carcinoma. *Int J Biol Macromol*. 2019; 129: 140-51. doi: 10.1016/j.ijbiomac.2019.02.012.
29. Li KK, Mao CY, Zhang JG, et al. Overexpression of U three protein 14a (UTP14a) is associated with poor prognosis of esophageal squamous cell carcinoma. *Thorac Cancer*. 2019; 10: 2071-80. doi: 10.1111/1759-7714.13176.
30. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011; 144: 646-74. doi: 10.1016/j.cell.2011.02.013.
31. Liu H, Luo J, Luan S, et al. Long non-coding RNAs involved in cancer metabolic reprogramming. *Cell Mol Life Sci*. 2019; 76: 495-504. doi: 10.1007/s00018-018-2946-1.
32. Begley U, Sosa MS, Avivar-Valderas A, et al. A human tRNA methyltransferase 9-like protein prevents tumour growth by regulating LIN9 and HIF1- α . *EMBO Mol Med*. 2013; 5: 366-83. doi: 10.1002/emmm.201201161.
33. Zhang L, Lu D, Liu M, et al. Identification and interaction analysis of key miRNAs in medullary thyroid carcinoma by bioinformatics analysis. *Mol Med Rep*. 2019; 20: 2316-24. doi: 10.3892/mmr.2019.10463.
34. Long NP, Jung KH, Yoon SJ, et al. Systematic assessment of cervical cancer initiation and progression uncovers genetic panels for deep learning-based early diagnosis and proposes novel diagnostic and prognostic biomarkers. *Oncotarget*. 2017; 8: 109436-56. doi: 10.18632/oncotarget.22689.
35. Popp MW, Maquat LE. Nonsense-mediated mRNA Decay and Cancer. *Curr Opin Genet Dev*. 2018; 48: 44-50. doi: 10.1016/j.gde.2017.10.007.
36. Chang NW, Huang YP. The RNA degradation pathway is involved in PPAR α -modulated anti-oral tumorigenesis. *Biomedicine (Taipei)*. 2019; 9: 27. doi: 10.1051/bmdcn/2019090427.
37. Hautbergue GM. RNA Nuclear Export: From Neurological Disorders to Cancer. *Adv Exp Med Biol*. 2017; 1007: 89-109. doi: 10.1007/978-3-319-60733-7_6.
38. McCann KL, Baserga SJ. Genetics. Mysterious ribosomopathies. *Science*. 2013; 341: 849-50. doi: 10.1126/science.1244156.
39. Narla A, Ebert BL. Ribosomopathies: human disorders of ribosome dysfunction. *Blood*. 2010; 115: 3196-205. doi: 10.1182/blood-2009-10-178129.
40. Kenmochi N. [Ribosomopathies—defective ribosome biogenesis and diseases]. *Seikagaku*. 2013; 85: 909-15. doi:
41. Freed EF, Bleichert F, Dutca LM, et al. When ribosomes go bad: diseases of ribosome biogenesis. *Mol Biosyst*. 2010; 6: 481-93. doi: 10.1039/b919670f.
42. Dong Z, Jiang H, Liang S, et al. Ribosomal Protein L15 is involved in Colon Carcinogenesis. *Int J Med Sci*. 2019; 16: 1132-41. doi: 10.7150/ijms.34386.

43. Zhang J, Luo W, Chi X, et al. IGF2BP1 silencing inhibits proliferation and induces apoptosis of high glucose-induced non-small cell lung cancer cells by regulating Netrin-1. *Arch Biochem Biophys*. 2020; 108581. doi: 10.1016/j.abb.2020.108581.
44. Degot S, Régnier CH, Wendling C, et al. Metastatic Lymph Node 51, a novel nucleo-cytoplasmic protein overexpressed in breast cancer. *Oncogene*. 2002; 21: 4422-34. doi: 10.1038/sj.onc.1205611.
45. Liu W, Yin C, Liu Y. Circular RNA circ_0091579 Promotes Hepatocellular Carcinoma Proliferation, Migration, Invasion, and Glycolysis Through miR-490-5p/CASC3 Axis. *Cancer Biother Radiopharm*. 2020. doi: 10.1089/cbr.2019.3472.
46. Liu L, Zhou Z, Huang S, et al. Zc3h12c inhibits vascular inflammation by repressing NF- κ B activation and pro-inflammatory gene expression in endothelial cells. *Biochem J*. 2013; 451: 55-60. doi: 10.1042/bj20130019.
47. Liang J, Wang J, Azfer A, et al. A novel CCCH-zinc finger protein family regulates proinflammatory activation of macrophages. *J Biol Chem*. 2008; 283: 6337-46. doi: 10.1074/jbc.M707861200.
48. Huang S, Qi D, Liang J, et al. The putative tumor suppressor Zc3h12d modulates toll-like receptor signaling in macrophages. *Cell Signal*. 2012; 24: 569-76. doi: 10.1016/j.cellsig.2011.10.011.
49. Huang WQ, Yi KH, Li Z, et al. DNA Methylation Profiling Reveals the Change of Inflammation-Associated ZC3H12D in Leukoaraisosis. *Front Aging Neurosci*. 2018; 10: 143. doi: 10.3389/fnagi.2018.00143.
50. Chen X, Zhao Y. Block of proliferation 1 promotes cell migration and invasion in human colorectal cancer cells via the JNK pathway. *J Clin Lab Anal*. 2020; 34: e23283. doi: 10.1002/jcla.23283.
51. Gupta R, Bugide S, Wang B, et al. 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. doi: 10.1073/pnas.1821889116.
52. Ma Z, Moore R, Xu X, et al. DDX24 negatively regulates cytosolic RNA-mediated innate immune signaling. *PLoS Pathog*. 2013; 9: e1003721. doi: 10.1371/journal.ppat.1003721.
53. Simarro M, Gimenez-Cassina A, Kedersha N, et al. Fast kinase domain-containing protein 3 is a mitochondrial protein essential for cellular respiration. *Biochem Biophys Res Commun*. 2010; 401: 440-6. doi: 10.1016/j.bbrc.2010.09.075.
54. Ye J, Wang J, Zhang N, et al. Expression of TARBP1 protein in human non-small-cell lung cancer and its prognostic significance. *Oncol Lett*. 2018; 15: 7182-90. doi: 10.3892/ol.2018.8202.
55. Ye J, Wang J, Tan L, et al. Expression of protein TARBP1 in human hepatocellular carcinoma and its prognostic significance. *Int J Clin Exp Pathol*. 2015; 8: 9089-96. doi: 10.1002/ijcp.2015.08.009.
56. Kwon OK, Ha YS, Lee JN, et al. Comparative Proteome Profiling and Mutant Protein Identification in Metastatic Prostate Cancer Cells by Quantitative Mass Spectrometry-based Proteogenomics. *Cancer Genomics Proteomics*. 2019; 16: 273-86. doi: 10.21873/cgp.20132.
57. Scott DD, Trahan C, Zindy PJ, et al. Nol12 is a multifunctional RNA binding protein at the nexus of RNA and DNA metabolism. *Nucleic Acids Res*. 2017; 45: 12509-28. doi: 10.1093/nar/gkx963.

58. Pinho M, Macedo JC, Logarinho E, et al. NOL12 Repression Induces Nucleolar Stress-Driven Cellular Senescence and Is Associated with Normative Aging. *Mol Cell Biol*. 2019; 39. doi: 10.1128/mcb.00099-19.

59. Zhang B, Babu KR, Lim CY, et al. A comprehensive expression landscape of RNA-binding proteins (RBPs) across 16 human cancer types. *RNA Biol*. 2020; 17: 211-26. doi: 10.1080/15476286.2019.1673657.

60. Wu YQ, Ju CL, Wang BJ, et al. PABPC1L depletion inhibits proliferation and migration via blockage of AKT pathway in human colorectal cancer cells. *Oncol Lett*. 2019; 17: 3439-45. doi: 10.3892/ol.2019.9999.

61. Jin Y, Qin X. Comprehensive Analysis of the Roles and Prognostic Value of RNA-Binding Proteins in Head and Neck Squamous Cell Carcinoma. *DNA Cell Biol*. 2020. doi: 10.1089/dna.2020.5641.

62. Gao L, Meng J, Zhang Y, et al. Development and validation of a six-RNA binding proteins prognostic signature and candidate drugs for prostate cancer. *Genomics*. 2020. doi: 10.1016/j.ygeno.2020.08.034.

Tables

Due to technical limitations, table 1 is only available as a download in the Supplemental Files section.

Figures

Figure 1

flowchart

Figure 2

A total of 459 differentially expressed RBPs (291 up-regulated and 168 down-regulated RBPs) were identified when NSCLC compared to control

Figure 3

The biological process was primarily associated with ncRNA metabolic process, ribonucleoprotein complex biogenesis, RNA splicing, regulation of translation, RNA catabolic process

Figure 4

The KEGG pathway was mainly enriched in RNA transport, Ribosome biogenesis in eukaryotes, mRNA surveillance pathway

	pvalue	Hazard ratio
ZC3H12C	0.008	1.135(1.034-1.245)
ZC3H12D	0.040	0.769(0.598-0.988)
RPP25	0.047	1.019(1.000-1.038)
BOP1	0.009	1.012(1.003-1.021)
EIF5A	0.049	1.001(1.000-1.003)
SPATS2	0.010	1.062(1.015-1.112)
CASC3	0.010	0.972(0.952-0.993)
MRPL15	0.005	1.006(1.002-1.011)
DDX24	0.002	0.968(0.949-0.988)
SMAD9	0.020	0.896(0.817-0.983)
CLK2	0.019	0.978(0.960-0.996)
IGF2BP1	0.003	1.047(1.015-1.080)
GAPDH	0.032	1.000(1.000-1.000)
CIRBP	0.004	0.989(0.982-0.996)
KHDC1	0.037	0.870(0.763-0.992)
DDX52	0.004	1.089(1.027-1.155)
THOC1	0.045	0.916(0.840-0.998)
ZRSR2	0.024	0.954(0.916-0.994)
FASTKD3	0.045	0.952(0.908-0.999)
DHX37	0.007	1.090(1.024-1.160)
RNASE7	<0.001	1.014(1.007-1.020)
TARBP1	0.015	0.964(0.935-0.993)
RAN	0.009	1.006(1.001-1.010)
CWC25	0.021	0.935(0.883-0.990)
RAVER2	0.024	0.926(0.865-0.990)
INTS7	0.039	1.040(1.002-1.079)
TRMT1	0.043	0.977(0.956-0.999)
PRIM1	0.045	1.037(1.001-1.074)
NOP56	0.034	1.010(1.001-1.019)
NOL12	0.033	0.887(0.794-0.991)
PCF11	0.005	0.902(0.838-0.970)
SNRPB	0.014	1.001(1.000-1.003)
SNRPN	0.040	0.989(0.978-0.999)
GTPBP4	0.025	1.028(1.003-1.052)
PABPC1L	0.033	0.981(0.964-0.998)
PABPC1	0.018	1.001(1.000-1.001)
TERT	0.035	1.141(1.009-1.290)
YARS2	0.015	1.024(1.005-1.044)

Figure 5

38 prognostic RBPs were identified from 459 differentially expressed RBPs by univariate Cox regression analysis

	pvalue	Hazard ratio
ZC3H12C	0.023	1.180(1.023–1.361)
ZC3H12D	0.129	0.745(0.509–1.089)
BOP1	0.025	1.015(1.002–1.029)
CASC3	0.015	0.963(0.935–0.993)
DDX24	0.025	0.964(0.933–0.995)
IGF2BP1	0.057	1.046(0.999–1.095)
KHDC1	<0.001	0.655(0.525–0.817)
FASTKD3	0.016	0.915(0.851–0.983)
TARBP1	0.028	0.938(0.886–0.993)
INTS7	0.128	1.046(0.987–1.109)
NOL12	0.178	0.878(0.727–1.061)
SNRPB	0.130	1.001(1.000–1.003)
PABPC1L	0.022	1.042(1.006–1.079)

Figure 6

Subsequently, 13 RBPs (ZC3H12C, ZC3H12D, BOP1, CASC3, DDX24, IGF2BP1, KHDC1, FASTKD3, TARBP1, INTS7, NOL12, SNRPB, PABPC1L) were verified as independent indicator of prognosis by multivariable Cox regression analysis in training cohort

Figure 7

Figure 7A) and area under ROC curve was 0.703 (Figure 7B).

Figure 8

All patients were divided into low- or high-risk group according to the median risk score (Figure 8A). The survival time of each patient was shown in Figure 8B and the patients with a high-risk score had a higher mortality rate than those with a low-risk score. A heatmap was shown to present the different expression profile of the thirteen RBPs between two groups (Figure 8C).

Figure 9

]: 1.522–1.991; $P < 0.001$. HR: 1.521; 95% CI: 1.305–1.773; $P < 0.001$) and the multivariate analysis (HR: 1.779; 95% CI: 1.548–2.045; $P < 0.001$. HR: 1.578; 95% CI: 1.343–1.854; $P < 0.001$)

Figure 10

Compared to the low-risk group, the high-risk group has a poor survival time ($P < 0.001$; Figure 10A) and area under ROC curve was 0.636 (Figure 10B).

Figure 11

The risk score, survival status and RBPs expression data of NSCLC patients in the testing cohort were illustrated in Figure 11.

Figure 12

In addition, the risk score and tumor stage in testing cohort were also independently associated with prognosis in the univariate analysis (HR: 1.377; 95% CI: 1.179–1.607; $P < 0.001$. HR: 1.458; 95%CI: 1.245–1.708; $P < 0.001$) and the multivariate analysis (HR: 1.332; 95% CI: 1.134–1.566; $P < 0.001$; HR: 1.444; 95% CI: 1.223–1.705; $P < 0.001$) (Figure 12)

Figure 13

The high-risk group in all subgroup above were significantly associated with poor survival time compared to low-risk group (Figure 13).

Figure 14

In order to construct a quantitative model for NSCLC prognosis, thirteen prognostic RBPs were combined to build a nomogram plot (Figure 14).

Figure 15

CASC3, DDX24, IGF2BP1, FASTKD3, TARBP1, NOL12, SNRPB,) were downloaded from the Human Protein Atlas database due to KHDC1, INTS7 and PABPC1L without immunohistochemistry results in this database (Figure 15).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table1.docx](#)