Comprehensive Analysis of the Functions, Prognostic and Diagnostic Values of RNA Binding Proteins in Head and Neck Squamous Cell Carcinoma

Xueping Ke

The Affiliated Stomatological Hospital of Nanjing Medical University, Nanjing Medical University

Zhen Fu

Nanjing University of Chinese Medicine, Jiangsu Province Academy of Traditional Chinese Medicine

Jingjing Yang

The Affiliated Stomatological Hospital of Nanjing Medical University, Nanjing Medical University

Shijin Yu

The Affiliated Stomatological Hospital of Nanjing Medical University, Nanjing Medical University

Tingyuan Yan

The Affiliated Stomatological Hospital of Nanjing Medical University, Nanjing Medical University

Xin Li

The Affiliated Stomatological Hospital of Nanjing Medical University, Nanjing Medical University

Zhongwu Li

Affiliated Stomatological Hospital, Nanjing Medical University

Xiang Wu (✉ Wuxiang15851859210@163.com)

The Affiliated Stomatological Hospital of Nanjing Medical University, Nanjing Medical University

Research Article

Keywords: Head and neck squamous cell carcinoma (HNSCC), RNA-binding proteins (RBP), prognostic signature, Diagnostic signature

Posted Date: December 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1192565/v1>

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

[Read Full License](#)

Abstract

Background: Increasing evidence has suggested that RNA binding protein (RBP) dysregulation plays an important part in tumorigenesis. Here, we sought to explore the potential molecular functions and clinical significance of RBP and develop diagnostic and prognostic signatures based on RBP in patients with head and neck squamous cell carcinoma (HNSCC).

Methods: The Limma package was applied to identify the differently expressed RBPs between HNSCC and normal samples with $|\log_2 \text{fold change (FC)}| \geq 1$ and false discovery rate (FDR) < 0.05 . The immunohistochemistry images from the Human Protein Atlas database The diagnostic signature based on RBP was built by LASSO-logistic regression and random forest and the prognostic signature based on RBP was constructed by LASSO and stepwise Cox regression analysis in training cohort and validated in validation cohort. All these analyses were performed using the R software.

Results: A total of 84 aberrantly expressed RBPs were obtained, comprising 41 up-regulated and 43 down-regulated RBPs. Seven RBP genes (CPEB3, PDCD4, ENDOU, PARP12, DNMT3B, IGF2BP1, EXO1) were identified as diagnostic related hub gene and were used to establish a diagnostic RBP signature risk score (DRBPS) model by the coefficients in LASSO-logistic regression analysis and shown high specificity and sensitivity in the training (area under the receiver operating characteristic curve [AUC] = 0.998), and in all validation cohorts (AUC > 0.95 for all). Similarly, seven RBP genes (MKRN3, ZC3H12D, EIF5A2, AFF3, SIDT1, RBM24 and NR0B1) were identified as prognosis associated hub genes by least absolute shrinkage and selection operator (LASSO) and stepwise multiple Cox regression analyses and were used to construct the prognostic model named as PRBPS. The area under the curve of the time-dependent receiver operator characteristic curve of the prognostic model were 0.664 at 3 years and 0.635 at 5 years in training cohort and 0.720, 0.777 in the validation cohort, showing a favorable predictive efficacy for prognosis in HNSCC.

Conclusions: Our results demonstrate the values of consideration of RBP in the diagnosis and prognosis for HNSCC and provide a novel insights to understand potential role of dysregulated RBP in HNSCC.

Introduction

Head and neck cancer is one of the malignant diseases which is still one of the main causes of cancer deaths worldwide. It is mainly composed of tumors of the larynx, pharynx-throat and oral cavity, and its main histology is squamous cell carcinoma [1]. More than 600,000 patients with head and neck squamous cell carcinoma are diagnosed each year, making it the sixth most common malignant tumor in the world [2, 3]. Despite there has been great improvements in the methods of treatment including ablative surgery, chemotherapy and radiotherapy over the past few decades, the prognosis of patients with this highly malignant disease remains unsatisfactory as proved by unfavorable 5-year survival rates still remains around 50% [4, 5]. Various clinicopathological parameters such as tumor size, cervical lymph node involvement and clinical stage (TNM) have been identified as the pivotal prognostic indicators [6].

However, due to the specificity, sensitivity and performance of prognostic prediction, these prognostic indicators are far from optimal [7, 8]. Therefore, identify novel biomarkers with sufficient performance and clinical feasibility for early diagnosis and prognostic prediction will be crucial for improving the therapeutic effect and quality of life in patients with HNSCC.

RNA binding protein (RBP) is a type of protein that binds to various types of RNA through one or more spherical RNA binding domains (RBD) and is widely expressed in cells [9]. To date, more than 1,54 proteins have been determined as RBPs based on the presence of characteristic RNA-binding domains or their residence in established ribonucleoprotein complexes by genome-wide screening in human genome, approximately 7.5% of all protein coding genes [10]. These RBP have a far-going of biological functions, including post-transcriptional regulation of mRNA stability, mRNA splicing, mRNA editing and translation, mRNA polyadenylation and localization, which will ultimately affect the expression of genes in the cell [9-12]. Given the importance of post-transcriptional regulation in life processes, it is not surprising that abnormal expression of RBP is closely associated with the occurrence and development of various human cancers. Undoubtedly, the potential application of RBP as either a therapeutic target or a diagnostic marker in cancer represents a fast-growing research field.

In recent years, the fast-paced development in bioinformatics as well as the increasing availability of large-scale RNAsequencing (RNA-seq) transcriptome data of multiple cancers combined with clinical follow-up information makes exploring biomarkers for diagnosis and prognosis much more efficient and accurate, providing better approaches for the classification of patients and the implementation of personalized therapy, as well as uncovering the underlying mechanism of tumorigenesis [13]. Recently, genome-wide analysis has also shown that, compared to adjacent normal tissues, many RBPs are dysregulated in tumors, and their dysregulated expression were involved in carcinogenesis and related to the prognosis of patients with cancers [14-17]. For example, Wang.et.al have reported that the RNA-binding protein CELF1 enhances the malignant phenotypes of colorectal cancer cell such as cell invasion , migration and chemotherapy resistance by targeting ETS2 and Huan Feng.et al reported that RBM43 inhibits the development of hepatocellular carcinoma by regulating the expression of cyclin B1 [18, 19]. In summary, these studies show that RBP is closely related to the occurrence and development of human cancer.

In this study, we collected all available RNA sequencing (RNA-seq) data for HNSCC from Gene the Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases and performed comprehensive analysis to explore the potential molecular functions and clinical significance of RBPs in HNSCC. We distinguished diverse differentially expressed RBPs related to HNSCC, which provide new understanding of the pathogenesis of HNSCC. Then we also applied several machine learning methods to recognize pivotal RBP predictors and develop diagnostic and prognostic RBP signatures, which could provide new insights into predicting and evaluating the clinical survival of patients with HNSCC.

Materials And Methods

Data collection and preprocessing

The original TCGA RNA-seq data were downloaded from TCGA database. And all relevant epidemiological, clinicopathological as well as follow-up data were obtained and collated. Moreover, six independent HNSCC datasets (GSE41613, GSE42743, GSE23036, GSE30784, GSE25099 and GSE12452) from GEO database were found and utilized. Of which, the follow-up data of GSE41613 and GSE42743 were also downloaded from GEO database.

Identification of differentially expressed RBPs and gene function analysis

A total of 1542 RBPs were included in the analysis [10]. The Limma package was applied to identify the differentially expressed RBPs between HNSCC and normal samples with $|\log_2 \text{fold change (FC)}| \geq 1$ and false discovery rate (FDR) < 0.05 . Moreover, the biological functions of these differentially expressed RBPs were comprehensively detected by gene ontology (GO) enrichment analysis which including biological process (BP), cellular component (CC) and molecular function (MF). And the enrichment analyses were performed by clusterProfiler package in R with a P-value less than 0.05.

Protein–Protein Interaction (PPI) Network Construction, Module and Hub genes Screening

We obtained the protein-protein interaction information of differently expressed RBPs from the STRING database. Then we constructed the PPI network of RBPs by using the Cytoscape 3.6.1 and the key modules were screened by using the MCODE (Molecular Complex Detection) with the MCODE score and nodes numbers more than 5. The cytoHubba was used to select hub genes in Cytoscape. The two side $P < 0.05$ was considered statistically significant. Moreover, the immunohistochemical images from the Human Protein Atlas (HPA) database were used to detect the expression of the hub genes at the translational level.

The construction of diagnostic model for HNSCC

The TCGA HNSC dataset was defined as the training cohort to construct the diagnostic model. The random forest analysis was performed by using the “randomForest” package and the top ten important variables were screen based on the Gini values. Then LASSO-logistic analysis, as a statistical method, was carried out to identify the most important RBP genes, whose predictive accuracy could be improved significantly. The overlapping markers between these two methods were finally selected to build the diagnostic model which was built based on the coefficients of LASSO-logistic analysis. The GSE42743, GSE23036, GSE30784, GSE25099 and GSE12452 datasets were used to validate the result from TCGA dataset. In addition, the receiver operating characteristic (ROC) curve was plotted to evaluate the sensitivity and specificity by using the “ROCR” package in R software.

The construction of prognostic model for HNSCC patients

All HNSCC patients from TCGA-HNSC dataset were defined as the training cohort. The prognostic signature was constructed according to the following steps. Firstly, the univariate Cox regression was

used to identify survival-related RBPs from the differently expressed RBPs with the P -value <0.05 . Secondly, the LASSO regression was performed on these survival-related metabolic pathways and the non-zero coefficients' RBPs were filtered out for further analyses. Finally, the final prognostic model was constructed by stepwise multivariate Cox regression analysis. The LASSO analysis was performed by "glmnet" package. Moreover, Datasets in patients from GSE41613 and GSE42743 were pooled and defined as an independent, validation cohort to validate the result from TCGA dataset. And "Combat" in R package "sva" was used to remove batch effects. All these analyses were performed using the R software (version 3.6.3).

Statistical analyses

All HNSCC patients were divided into low-risk and high-risk groups based on the median value. The overall survival curve were plotted by Kaplan-Meier method and the difference were compared with the log-rank test. A time-dependent ROC curve with 1, 3, and 5 years as the defining point was performed to evaluate the predictive value of prognostic risk score by using the "survivalROC" package. Univariate and multivariate Cox regression analyses were employed to determine prognostic factors associated with survival. The evaluation of statistically significant differences of DRBPs between HNSCC and normal tissue in several datasets were performed by one-way ANOVA analyses. All these analyses were performed using the R software (version 3.6.3). And all statistical tests were two sided and p values less than 0.05 were considered statistically significant.

Results

Patient cohorts and RNA-seq datasets

502 HNSCC samples and 44 normal samples with original RNA-seq data and clinical follow-up information from TCGA were retrieved and defined as the training cohort. RNA-seq data and clinical information of 171 HNSCC samples (defined as validation cohort for prognostic model) were obtained from GSE41613 and GSE42743. Moreover, a total of 392 HNSCC samples and 111 normal samples were obtained from GSE42743, GSE23036, GSE30784, GSE25099 and GSE12452 (defined as validation cohorts for diagnostic model).

Identification of differently expressed RBP between HNSCC and normal samples from TCGA

The workflow of this study design was illustrated in Figure 1. The dataset of HNSCC were downloaded from TCGA contained 502 tumor samples and 44 normal tissue samples. The limma packages were applied to handle the data and discover the differently expressed RBPs. A total of 1542 RBPs were included in the analysis [10], and 84 RBPs met the screening standard of this study ($P<0.05$, $|\log_2FC|>1.0$), which consist of 41 upregulated and 43 downregulated RBPs were shown in **Supplementary Table 1**. The expression distribution of these differently expressed RBPs was displayed in **Figure 2**.

Functional Enrichment Analysis of the Differentially Expressed RBP

To explore the potential functional and molecular mechanisms of the identified RBPs, they were divided into two groups based on their expression level. Then we carried out GO analysis for these differentially expressed RBPs by R software. As showed in Figure 3, upregulated differentially expressed RBPs were significantly enriched in biological processes associated with defense response to virus, response to virus, RNA catabolic process and nucleic acid phosphodiester bond hydrolysis (**Figure 3A**). The downregulated differentially expressed RBPs were notably enriched in the regulation of cellular amide metabolic process, regulation of translation, RNA splicing and regulation of mRNA metabolic process (**Figure 3B**). The molecular function analysis showed that, among the differentially expressed RBPs, the upregulated RBPs were significantly enriched in RNA binding catalytic activity, acting on RNA, double-stranded RNA binding and RNA helicase activity (**Figure 3A**), whereas the downregulated RBPs were significantly enriched in mRNA 3'-UTR binding, AU-rich element binding and translation factor activity, RNA binding (**Figure 3B**). In regard to the cellular component, the upregulated RBPs and downregulated RBPs were mainly enriched in the cytoplasmic ribonucleoprotein granule and ribonucleoprotein granule (**Figure 3**).

Protein-protein interaction (PPI) network construction and key modules selecting

In order to further study the potential molecular functions of abnormal expressed RNA binding proteins in HNSCC, we used Cytoscape software to create a PPI network, which merged 84 nodes and 158 edges based on the data from the STRING database (**Figure 4A**). Then we used cytoHubba in Cytoscape to screen the hub genes by calculation and obtained the top 10 hub genes: IFIH1, DDX60, DDX58, EIF2AK2, OASL, OAS2, PARP12, IFIT3, IFIT1 and IFIT2. These hub genes were all unregulated in HNSCC samples.

In order to further determine these hub genes expression in HNSCC, we utilized immunohistochemistry results from the Human Protein Atlas database (HPA) to indicate that IFIH1, DDX60, DDX58, EIF2AK2, OASL, OAS2, PARP12, IFIT3, IFIT1 and IFIT2 were uncommonly increased in HNSCC compared with normal tissues. (**Figure 5**). Subsequently, we further analyzed the co-expression network by using the plug-in MODE in Cytoscape to detect potential key modules, and identified the two most important modules. Module 1 consisted of 12 nodes and 64 edges (**Figure 4B**), Module 2 contained 6 nodes and 14 edges (**Figure 4C**). The GO analysis indicated that the genes from module 1 were predominantly enriched in defense response to virus and double-stranded RNA binding, whereas the genes in module 2 were significantly enriched in piRNA metabolic process, macromolecule methylation and cytoplasmic ribonucleoprotein granule (**Supplementary Figure S1**).

Diagnostic signature building

The TCGA HNSCC cohort was selected as the training cohort to construct the diagnostic signature and validated in several HNSCC cohorts from GEO databases. Twenty-four potential predictors in the TCGA HNSCC cohort were identified and were features with nonzero coefficients in the LASSO-logistic regression model. Then, the random-forest analysis (**Figure 6A-B**) and LASSO analysis (**Figure 6C-D**) revealed seven overlapping markers such as CPEB3, PDPCD4, ENDOU, PARP12, DNMT3B, IGF2BP1, EXO1 between the two methods. We established a diagnostic RBP signature risk score (DRBPS) model based

on the seven overlapping markers by the coefficients in LASSO-logistic regression analysis. The DRBPS were calculated as the following: $DRBPS = (-0.16021961) \times \text{expression (CPEB3)} + (-0.41296683) \times \text{expression (PDCD4)} + (-0.27292101) \times \text{expression (ENDOU)} + (0.14022947) \times \text{expression (PARP12)} + (0.49962314) \times \text{expression (DNMT3B)} + (0.13455980) \times \text{expression (IGF2BP1)} + (0.57217627) \times \text{expression (EXO1)}$. The violin plot showed that the DRBPS value was markedly upregulated in HNSCC compared to normal tissues in all cohorts from TCGA and GEO databases (**Figure 6E**). The receiver operating characteristic curve revealed the DRBPS has a high accuracy in distinguishing HNSCC patients from normal controls with the AUC was 0.998 in the training cohort (**Figure 6F**). Furthermore, And the ROC curves also showed the excellent ability of DRBPS in differentiating between tumor and normal tissues in all validation cohorts with the AUC values were 0.958 in GSE42743, 0.994 in GSE23036, 0.993 in GSE30784, 0.997 in GSE25099 and 0.974 in GSE12452 cohorts (**Figure 6 G-K**).

Prognostic signature building

The TCGA HNSCC cohort was selected as the training cohort to construct the prognostic signature and validated in the validation cohort which combined with GSE41613 and GSE42743 datasets from GEO database. We identified 84 dysregulated RBP genes which differs between HNSCC and normal samples with $P < 0.05$ and then found 14 survival-related RBP genes by univariate Cox regression analysis ($P < 0.05$). Furthermore, we performed LASSO, stepwise multiple Cox regression analyses and developed a prognostic signature consisting of 7 RBP genes: MKRN3, ZC3H12D, EIF5A2, AFF3, SIDT1, RBM24 and NR0B1. The prognostic RBP signature was developed based on RBP genes was named as PRBPS. The formula was as following: $PRBPS = (0.15117) \times \text{expression (MKRN3)} + (-0.33896) \times \text{expression (ZC3H12D)} + (0.30465) \times \text{expression (EIF5A2)} + (-0.13525) \times \text{expression (AFF3)} + (0.13594) \times \text{expression (SIDT1)} + (0.11666) \times \text{expression (RBM24)} + (0.08505) \times \text{expression (NR0B1)}$. All patients in the training sets were divided into high risk group and low risk group according to median value. The Kaplan-Meier analyses indicated that patients in high-risk subgroup had a worse prognosis than those in low-risk subgroup in the training cohort and validation cohort (**Figure 7C, D**). Moreover, a time-dependent receiver operating characteristic (ROC) curve were performed to estimate the sensitivity and specificity of this PRBPS. Data listed in **Figure 7E, F** indicated that this signature to predict patient survival in 1, 3 and 5 years was satisfactory with AUC 0.667, 0.664, and 0.635 in the training cohort and 0.728, 0.720, 0.777 in the validation cohort. The expression heat map and survival status of patients with the seven prognostic RBP gene biomarkers in the low and high risk subgroups in training and validation cohorts are shown in **Figure 7H, I** and **Supplementary Figure S2**.

To further explore the prognostic value of seven prognostic RBPs in HNSCC, we divided patients into low or high expression subgroups based on the median values of gene expression. The Kaplan Meier-plotter was used to determine the relationship between seven RBP genes and OS. And the results demonstrated that most prognostic RBP genes were associated with OS in HNSCC patients ($P < 0.05$) except SIDT1 and NR0B1 (**Supplementary Figure S3**).

Univariate and multivariate Cox regression analyses for PRBPS

To further delineate the prognostic value of PRBPS in HNSCC, we performed univariate and multivariate Cox regression using the TCGA HNSCC cohort. Our data revealed that the PRBPS was associated with overall survival in HNSCC patients (HR, 1.961; 95% CI: 1.489-2.584; $P < 0.0001$). In addition, some well-established clinicopathological parameters such as margin status and tumor stage were also identified to be prognostic factors (**Supplementary Figure S4A**). Moreover, we performed multivariate Cox regression analyses to eliminate confounding factors and highlight the prognostic value of PRBPS. And PRBPS was identified as an independent predictive factor for HNSCC (HR, 2.019; 95% CI: 1.477-2.760; $P < 0.0001$) (**Supplementary Figure S4B**).

Discussion

The poor long-term survival ability of HNSCC and the lack of valuable prognostic prediction tools highlights the urgent need to identify new biomarkers with the best predictive performance to facilitate the diagnosis, treatment and prognosis monitoring of cancer. In recent years, many studies have shown that the expression of RBPs is dysregulated in a variety of human cancers [20-22]. However, the expression patterns and roles of RBPs in HNSCC are poorly understood. Microarray or high-throughput sequencing technology provides an effective detection tool for understanding the changes in key gene expression during cancer development, and also provides promising biomarkers for cancer treatment, diagnosis and prognosis [23-25]. In this study, we summarized all HNSCC-related TCGA RNA sequencing data, and identified 84 differentially expressed RBP genes between cancer and normal tissues. We systematically studied the relevant biological pathways and constructed the PPI network of these RBPs. Moreover, we established a diagnostic prediction signature (DRBPS) based on ten RBPs genes and we also constructed a prognostic RBP signature (PRBPS) to predict HNSCC prognosis in the light of seven prognostic-related RBP genes. These findings may be beneficial to the development of new biomarkers, thereby helping the diagnosis and prognosis prediction of HNSCC patients.

The enrichment analysis of functional pathways showed that differently expressed RBPs were greatly enriched in the adjustment of biological processes associated with mRNA metabolic process, RNA processing, RNA catabolic process, cellular amide metabolic process, regulation of translation, RNA splicing and regulation of mRNA metabolic process. Previous studies have proved that the regulatory processes of RNA processing, translation and RNA metabolism are associated with the occurrence and development of various human diseases[26]. The post-transcriptional regulation of RNA stability is an important biological process in the process of gene expression. RBPs can interact with RNA to form ribonucleoprotein complexes, thereby increasing the stability of target mRNA and promoting gene expression, which plays a key role in the development of various diseases [27]. For example, The RNA-binding protein NELFE is an oncogenic protein that may cause the imbalance of the HCC transcriptome by regulating MYC signaling [28]. RNA binding protein IGF2BP3 targets oncogenic transcripts and promotes proliferation of hematopoietic progenitor cells [15]. In addition, we created the protein-protein interaction network of these differently expressed RBPs and identified the two most important modules. Module 1 included 12 nodes and 64 edges, and module 2 consisted of 6 nodes and 14 edges. Then we screen the hub genes by computing degree, and obtained the top 10 hub genes including IFIH1,

DDX60, DDX58, EIF2AK2, OASL, OAS2, PARP12, IFIT3, IFIT1 and IFIT2. These hub genes were all upregulated at transcriptional level and translation level in HNSCC evidenced by the results from TCGA and HPA database. Among these Hub RBP genes, most of them have been shown to play an important role in the occurrence and development of cancer. For example, RBP-related proteins IFIT1 and IFIT3 contribute to the metastasis of oral squamous cell carcinoma and promote the anti-tumor effect of gefitinib by enhancing the p-EGFR recycling process [29]. EIF2AK2 impedes the DNA damage response, and is closely related to low survival rate of AML [30].

As we all know, prognostic and diagnostic features might be able to clarify the likelihood of a certain disease and its relative progress, recurrence and death, which could conduce to patient classification, treatment management, and monitoring of disease status in clinical practice [31]. Therefore, the role of abnormally expressed genes or other markers in cancer tissues on diagnosis and prognosis has aroused great interest. For example, Yang et al established a diagnostic and prognostic immune cell infiltration risk score for the diagnosis of digestive system cancer [32]. Zhou et al, established a diagnostic immune cell infiltration risk score for the diagnosis of colon cancer [33]. In this study, we established a diagnostic RBP signature (DRBPS) risk score model based on the seven RBP genes, including CPEB3, PDCD4, ENDOU, PARP12, DNMT3B, IGF2BP1, EXO1. The receiver operating characteristic curve revealed the DRBPS has a superior performance in distinguishing HNSCC from normal controls in the training cohort and it also showed the excellent ability in differentiating between tumor and normal tissues in all validation cohorts with the high AUC values. The significant stepwise increase in the DRBPS value from a normal tissue to tumor tissue, as well as the high AUC value not only demonstrated that the DRBPS model could effectively identify patients with HNSCC from individuals with healthy controls but also suggested that the RBPs participates in HNSCC carcinogenesis.

Recent studies have shown that the combination of several biomarkers has better predictive performance than a single biomarker. For example, Hess J et al build a 5-microRNA signature that can predict the survival and disease control of HPV-negative patients with head and neck cancer [34]. Liu et al obtained 5-lncRNA signatures using high-throughput lncRNAs data in the TCGA database to predict the survival time of patients with head and neck squamous cell carcinoma [35]. In addition, Wei Li et al found that a 9-RBP gene markers can predict the prognostic value of LUSC [36]. In this study, we analyzed the RBP related genes to find a more efficient and accurate features for prognosis-related RBPs, the prognostic RBP signature (PRBPS) were constructed by using LASSO regression analysis and stepwise multiple Cox regression analysis. We finally identified seven RBP-coding genes: MKRN3, ZC3H12D, EIF5A2, AFF3, SIDT1, RBM24 and NR0B1. High expression of ZC3H12D, SIDT1 and AFF3 was related to a good prognosis of HNSCC patients, whereas that of RBM24, EIF5A2, MKRN3 and NR0B1 were associated with poor prognosis. Previous research have confirmed that the dysregulated expression of SIDT1, AFF3, and NR0B1 were closely related to the occurrence and development of human tumors [37-39]. Consistently, abnormal overexpression of EIF5A2 was related to poor prognosis of patients with oral squamous cell carcinoma [40]. Our PRBPS robustly stratified patients into subgroups and served as an independent factor affecting the survival and prognosis. This predictive performance of RBPS was validated in an independent cohort, thus reinforce its translational potentials. In addition, the RBPs-related gene showed

key biological function, suggesting that they have a potential value for clinical adjuvant therapy. Together, these results revealed the prognostic value of PRBPS for HNSCC, and suggest that dysregulated RBP genes involved might facilitate the initiation and progression of HNSCC.

Nonetheless, we have to admit that our research has some unavoidable limitations. Firstly, the amount of data released in the publicly available data set is limited, so that the clinicopathological parameters analyzed in this study are not comprehensive, which might lead to potential error or bias. Secondly, our research was conducted on the basis of a retrospective analysis, and prospective research should be conducted to verify the results. Thirdly, the number of HNSCC sample is relatively small, this signature is still needed to be independently validated in more cohorts. However, data from multiple databases and our training validation queue design might rescue for these shortcomings.

In conclusion, our study demonstrates the values of consideration of RBPs in the diagnosis and prognosis for HNSCC. The proposed DRBPS and PRBPS models might provide much needed comprehensive clinical information for improving the personalized management of HNSCC patients.

Declarations

Acknowledgements

We would like to thank the team members for their contributions to this paper, and then we will continue to work hard to do relevant research.

Authors' contributions

Xueping Ke and Zhen Fu performed data collection and analysis. Jingjing Yang, Shijin Yu, Tingyuan Yan and Xin Li assisted out data collection and statistical analyses. The manuscript was drafted by Xueping Ke. Zhongwu Li and Xiang Wu conceived and supervised the whole project. All authors read and approved the final manuscript.

Funding

This work is financially supported, in whole or in part, by National Natural Science Foundation of China (81572669, 81602386, 81602378), Natural Science Foundation of Jiangsu Province (BK20161564, BK20161024, BK20151561).

Availability of data and materials

The datasets supporting the conclusion of this article are included within the article.

Competing interests

The authors declare that there are no conflict interests.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

References

1. Leemans, CR,PJF SnijdersRH Brakenhoff. The molecular landscape of head and neck cancer. *Nat Rev Cancer*. (2018) 18(5): 269-282. doi:10.1038/nrc.2018.11
2. Siegel, RL,KD MillerA Jemal. Cancer statistics, 2019. *CA Cancer J Clin*. (2019) 69(1): 7-34. doi:10.3322/caac.21551
3. Alsahafi, E, K Begg, I Amelio, N Raulf, P Lucarelli, T Sauter, et al. Clinical update on head and neck cancer: molecular biology and ongoing challenges. *Cell Death Dis*. (2019) 10(8): 540. doi:10.1038/s41419-019-1769-9
4. Cramer, JD,B Burtneess,QT LeRL Ferris. The changing therapeutic landscape of head and neck cancer. *Nat Rev Clin Oncol*. (2019) 16(11): 669-683. doi:10.1038/s41571-019-0227-z
5. Bratman, SV, JP Bruce, B O'Sullivan, TJ Pugh, W Xu, KW Yip, et al. Human Papillomavirus Genotype Association With Survival in Head and Neck Squamous Cell Carcinoma. *JAMA Oncol*. (2016) 2(6): 823-6. doi:10.1001/jamaoncol.2015.6587
6. Sacco, AGEE Cohen. Current Treatment Options for Recurrent or Metastatic Head and Neck Squamous Cell Carcinoma. *J Clin Oncol*. (2015) 33(29): 3305-13. doi:10.1200/JCO.2015.62.0963
7. Lydiatt, WM, SG Patel, B O'Sullivan, MS Brandwein, JA Ridge, JC Migliacci, et al. Head and Neck cancers-major changes in the American Joint Committee on cancer eighth edition cancer staging manual. *CA Cancer J Clin*. (2017) 67(2): 122-137. doi:10.3322/caac.21389
8. Denaro, N,EG RussiMC Merlano. Pros and Cons of the New Edition of TNM Classification of Head and Neck Squamous Cell Carcinoma. *Oncology*. (2018) 95(4): 202-210. doi:10.1159/000490415
9. Masuda, KY Kuwano. Diverse roles of RNA-binding proteins in cancer traits and their implications in gastrointestinal cancers. *Wiley Interdiscip Rev RNA*. (2019) 10(3): e1520. doi:10.1002/wrna.1520
10. Gerstberger, S,M HafnerT Tuschl. A census of human RNA-binding proteins. *Nat Rev Genet*. (2014) 15(12): 829-45. doi:10.1038/nrg3813
11. Nahalka, J. The role of the protein-RNA recognition code in neurodegeneration. *Cell Mol Life Sci*. (2019) 76(11): 2043-2058. doi:10.1007/s00018-019-03096-3
12. Campos-Melo, D,CA Droppelmann,K VolkeningMJ Strong. RNA-binding proteins as molecular links between cancer and neurodegeneration. *Biogerontology*. (2014) 15(6): 587-610. doi:10.1007/s10522-014-9531-2

13. Vogelstein, B,N Papadopoulos,VE Velculescu,S Zhou,LA Diaz, Jr.KW Kinzler. Cancer genome landscapes. *Science*. (2013) 339(6127): 1546-58. doi:10.1126/science.1235122
14. Pereira, B,M BillaudR Almeida. RNA-Binding Proteins in Cancer: Old Players and New Actors. *Trends Cancer*. (2017) 3(7): 506-528. doi:10.1016/j.trecan.2017.05.003
15. Palanichamy, JK, TM Tran, JM Howard, JR Contreras, TR Fernando, T Sterne-Weiler, et al. RNA-binding protein IGF2BP3 targeting of oncogenic transcripts promotes hematopoietic progenitor proliferation. *J Clin Invest*. (2016) 126(4): 1495-511. doi:10.1172/JCI80046
16. Cooke, A, T Schwarzl, I Huppertz, G Kramer, P Mantas, AM Alleaume, et al. The RNA-Binding Protein YBX3 Controls Amino Acid Levels by Regulating SLC mRNA Abundance. *Cell Rep*. (2019) 27(11): 3097-3106 e5. doi:10.1016/j.celrep.2019.05.039
17. Yun, Y, Y Zhang, C Zhang, L Huang, S Tan, P Wang, et al. Regulator of calcineurin 1 is a novel RNA-binding protein to regulate neuronal apoptosis. *Mol Psychiatry*. (2019). doi:10.1038/s41380-019-0487-0
18. Wang, H, R Huang, W Guo, X Qin, Z Yang, Z Yuan, et al. RNA binding protein CELF1 enhances cell migration, invasion, and chemoresistance by targeting ETS2 in colorectal cancer. *Clin Sci (Lond)*. (2020). doi:10.1042/CS20191174
19. Feng, H, J Liu, Y Qiu, Y Liu, H Saiyin, X Liang, et al. RNA-binding motif protein 43 (RBM43) suppresses hepatocellular carcinoma progression through modulation of cyclin B1 expression. *Oncogene*. (2020). doi:10.1038/s41388-020-1380-7
20. Hopkins, TG, M Mura, HA Al-Ashtal, RM Lahr, N Abd-Latip, K Sweeney, et al. The RNA-binding protein LARP1 is a post-transcriptional regulator of survival and tumorigenesis in ovarian cancer. *Nucleic Acids Res*. (2016) 44(3): 1227-46. doi:10.1093/nar/gkv1515
21. Vo, DT, D Subramaniam, M Remke, TL Burton, PJ Uren, JA Gelfond, et al. The RNA-binding protein Musashi1 affects medulloblastoma growth via a network of cancer-related genes and is an indicator of poor prognosis. *Am J Pathol*. (2012) 181(5): 1762-72. doi:10.1016/j.ajpath.2012.07.031
22. Forouzanfar, M,L Lachinani,K Dormiani,MH Nasr-Esfahani,AO GureK Ghaedi. Intracellular functions of RNA-binding protein, Musashi1, in stem and cancer cells. *Stem Cell Res Ther*. (2020) 11(1): 193. doi:10.1186/s13287-020-01703-w
23. Kulasingam, VEP Diamandis. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nat Clin Pract Oncol*. (2008) 5(10): 588-99. doi:10.1038/ncponc1187
24. Kourou, K,TP Exarchos,KP Exarchos,MV KaramouzisDI Fotiadis. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. (2015) 13: 8-17. doi:10.1016/j.csbj.2014.11.005
25. Ji, YAV Tulin. Post-transcriptional regulation by poly(ADP-ribosyl)ation of the RNA-binding proteins. *Int J Mol Sci*. (2013) 14(8): 16168-83. doi:10.3390/ijms140816168
26. Lin, CWO Miles. Beyond CLIP: advances and opportunities to measure RBP-RNA and RNA-RNA interactions. *Nucleic Acids Res*. (2019) 47(11): 5490-5501. doi:10.1093/nar/gkz295
27. . (!!! INVALID CITATION !!! [27, 28]).

28. Dang, H, A Takai, M Forgues, Y Pomyen, H Mou, W Xue, et al. Oncogenic Activation of the RNA Binding Protein NELFE and MYC Signaling in Hepatocellular Carcinoma. *Cancer Cell*. (2017) 32(1): 101-114 e8. doi:10.1016/j.ccell.2017.06.002
29. Pidugu, VK, MM Wu, AH Yen, HB Pidugu, KW Chang, CJ Liu, et al. IFIT1 and IFIT3 promote oral squamous cell carcinoma metastasis and contribute to the anti-tumor effect of gefitinib via enhancing p-EGFR recycling. *Oncogene*. (2019) 38(17): 3232-3247. doi:10.1038/s41388-018-0662-9
30. Cheng, X, M Byrne, KD Brown, MY Konopleva, SM Kornblau, RL Bennett, et al. PKR inhibits the DNA damage response, and is associated with poor survival in AML and accelerated leukemia in NHD13 mice. *Blood*. (2015) 126(13): 1585-94. doi:10.1182/blood-2015-03-635227
31. Nonaka, TDTW Wong. Liquid Biopsy in Head and Neck Cancer: Promises and Challenges. *J Dent Res*. (2018) 97(6): 701-708. doi:10.1177/0022034518762071
32. Yang, S,T Liu,Y Cheng,Y BaiG Liang. Immune cell infiltration as a biomarker for the diagnosis and prognosis of digestive system cancer. *Cancer Sci*. (2019) 110(12): 3639-3649. doi:10.1111/cas.14216
33. Zhou, R, J Zhang, D Zeng, H Sun, X Rong, M Shi, et al. Immune cell infiltration as a biomarker for the diagnosis and prognosis of stage I-III colon cancer. *Cancer Immunol Immunother*. (2019) 68(3): 433-442. doi:10.1007/s00262-018-2289-7
34. Hess, J, K Unger, C Maihoefer, L Schuttrumpf, L Wintergerst, T Heider, et al. A Five-MicroRNA Signature Predicts Survival and Disease Control of Patients with Head and Neck Cancer Negative for HPV Infection. *Clin Cancer Res*. (2019) 25(5): 1505-1516. doi:10.1158/1078-0432.CCR-18-0776
35. Liu, G, J Zheng, L Zhuang, Y Lv, G Zhu, L Pi, et al. A Prognostic 5-lncRNA Expression Signature for Head and Neck Squamous Cell Carcinoma. *Sci Rep*. (2018) 8(1): 15250. doi:10.1038/s41598-018-33642-1
36. Li, W,X Li, LN GaoCG You. 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
37. Wang, Y, H Li, J Ma, T Fang, X Li, J Liu, et al. Integrated Bioinformatics Data Analysis Reveals Prognostic Significance Of SIDT1 In Triple-Negative Breast Cancer. *Onco Targets Ther*. (2019) 12: 8401-8410. doi:10.2147/OTT.S215898
38. Lefevre, L, H Omeiri, L Drougat, C Hantel, M Giraud, P Val, et al. Combined transcriptome studies identify AFF3 as a mediator of the oncogenic effects of beta-catenin in adrenocortical carcinoma. *Oncogenesis*. (2015) 4: e161. doi:10.1038/oncsis.2015.20
39. Liu, XFX,XY Li,PS ZhengWT Yang. DAX1 promotes cervical cancer cell growth and tumorigenicity through activation of Wnt/beta-catenin pathway via GSK3beta. *Cell Death Dis*. (2018) 9(3): 339. doi:10.1038/s41419-018-0359-6
40. Lin, YM,ML Chen,CL Chen,CM YehWW Sung. Overexpression of EIF5A2 Predicts Poor Prognosis in Patients with Oral Squamous Cell Carcinoma. *Diagnostics (Basel)*. (2020) 10(7). doi:10.3390/diagnostics10070436

Figures

Figure 1

The procedures for analyzing RBPs in head neck squamous cell carcinoma (HNSCC).

Figure 2

The differentially expressed RBPs in HNSCC

(A) The heat map of differentially expressed RBPs between HNSCC and normal sample from TCGA database.

(B) The Volcano plot of differentially expression RBPs.

Figure 3

GO analysis results for differentially expressed RBPs.

(A) The up-regulated RBPs.

(B) The down-regulated RBPs.

A **B** **C** **Figure 4****PPI network and module analysis.**

(A) The PPI network for differentially expressed RBPs. The red blot represented up-regulated gene and the green blot represented down-regulated gene.

(B) The Key module 1 in PPI network

(C) The Key module 2 in PPI network.

Figure 5

Validation of protein expression of hub genes in HNSCC and normal tissue using the HPA database. (A) IFIH1, (B) DDX60, (C) DDX58, (D) EIF2AK2, (E) OASL, (F) OAS2, (G) PARP12, (H) IFIT3, (I) IFIT1, (J) IFIT2.

Figure 6

Construction and validation of diagnostic RBP score for HNSCC

(A) The top ten important RBPs in random forest analysis based on the Gini values.

(B) Multi-dimensional scaling plot of a proximity matrix generated from random forest analysis in the training cohort. The blue dots represent normal samples and the red dots indicate tumor samples. Dim dimension.

(C) The coefficient profile plot of 24 RBPs was produced against the log lambda sequence.

(D) Misclassification error for different numbers of variables revealed by the LASSO regression model. The red dots represent the value of misclassification error, the grey lines represent the standard error (SE), the two vertical dotted lines on the left and right, respectively, represent optimal values by the minimum criteria and 1-SE criteria.

(E) Distribution of DRBPS values between HNSCC and normal tissues in different datasets. The box plots inside the violin indicate the median value and interquartile range of DRBPS. The white points represent

mean RBPS values. RBPS diagnostic RBP risk score; N normal; T tumor

(F-K) The Receiver operating characteristic (ROC) curves of DRBPS in the training cohort **(F)** and validation cohorts: GSE42743 **(G)**, GSE23036 **(H)**, GSE30784 **(I)**, GSE25099 **(J)** and GSE12452 **(K)**.

Figure 7

Construction and validation of prognostic RBP score for HNSCC

(A) The coefficient profile plot of 14 survival-related RBPs was produced against the log lambda sequence.

(B) Tuning parameter (lambda) selection in the LASSO model used tenfold cross-validation via minimum criteria. Dotted vertical lines were drawn at the optimal values using the minimum criteria and the 1 standard error of the minimum criteria (the 1-SE criteria).

(C-D) The Kaplan-Meier analyses revealed significant associations between PRBPS and OS in patients from the training cohort and validation cohort.

(E-F) The time-dependent ROC curve analysis with 1, 3, 5 years as the defining point was performed to evaluate the predictive value of the PRBPS in training cohort and validation cohort.

(G-I) Patient subgroups with high and low-risk score were classified by the median value **(G)**. The survival status of patients with HNSCC in training cohort **(H)**. The heatmap showed the distribution of seven RBPs in high and low-risk group in the training cohort **(I)**.

Supplementary Files

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

- [SupplementaryFigureS1.pdf](#)
- [SupplementaryFigureS2.pdf](#)
- [SupplementaryFigureS3.pdf](#)
- [SupplementaryFigureS4.pdf](#)
- [SupplementaryTable1.csv](#)