PERTURBATION OF RNA BINDING PROTEIN REGULATION IN CANCER

EDITED BY: Pradeep Chaluvally Raghavan, Yongsheng Kevin Li, Li Guo and Juan Xu

PUBLISHED IN: Frontiers in Genetics

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-228-1
DOI 10.3389/978-2-88971-228-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

PERTURBATION OF RNA BINDING PROTEIN REGULATION IN CANCER

Topic Editors:
Pradeep Chaluvally Raghavan, Medical College of Wisconsin, United States
Yongsheng Kevin Li, Hainan Medical University, China
Li Guo, Nanjing University of Posts and Telecommunications, China
Juan Xu, Harbin Medical University, China

Citation: Raghavan, P. C., Li, Y. K., Guo, L., Xu, J., eds. (2021). Perturbation of RNA Binding Protein Regulation in Cancer. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-228-1

Table of Contents

frontiers in Genetics

EDITORIAL
published: 24 June 2021
doi: 10.3389/fgene.2021.693766

Editorial: Perturbation of RNA Binding Protein Regulation in Cancer

Yongsheng Li[1]* *and Juan Xu*[2]*

[1] Key Laboratory of Tropical Translational Medicine of Ministry of Education, College of Biomedical Information and Engineering, Hainan Medical University, Haikou, China, [2] College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China

Keywords: RNA binding protein, cancer, network, method, survival

Editorial on the Research Topic

Perturbation of RNA Binding Protein Regulation in Cancer

RNA-bind proteins (RBPs) are typically types of proteins that bind RNA to play critical roles in development or cancer (Pereira et al., 2017; Mohibi et al., 2019). Recent studies have identified thousands of RBPs and also revealed the dysregulation of RBPs in various kinds of cancer types, such as mutation (Chen et al., 2019), copy number variation (Xu et al., 2019), expression perturbation as well as perturbations of RBP-gene regulation (Zhang et al., 2020).

With the development of high throughput sequencing technology, some recent studies have highlighted precise dysregulated RBPs in specific cancers. Lung cancer is the leading cause of deaths worldwide and dysregulation of RBPs has been found in lung squamous cell carcinoma (LUSC). Li et al. analyzed the gene expression and clinical information from The Cancer Genome Atlas (TCGA) and observed 300 aberrantly expressed RBPs. These RBPs were mainly associated with mRNA metabolic processes, RNA modification and cancer-related signaling pathways. Moreover, they identified nine RBP genes for constructing a prognostic model in LUSC. In another study, Zhang et al. characterized the clinical relevance of RBPs in colorectal cancer. First, 242 differentially expressed RBPs were identified and eight RBPs were found to be related with the prognoses of colorectal cancer patients. Four RBPs (NOL3, PTRH1, UPF3B, and SMAD6) were used to construct the prognostic risk score model. In addition, Zhong et al. also constructed a prognostic model based on RBP expression in kidney renal clear cell carcinoma. Furthermore, potential drugs for cancer were predicted based on the Connectivity Map database. Moreover, RBPs were also play important roles during cancer progression (Wang et al.).

In addition, although some targets of RBPs were identified based on computational or experimental methods, the genome-wide RBP-gene regulatory network in cancer is largely unknown and little is known about the synergetic interaction between RBPs and other regulators. In recent studies, co-expression network analysis was applied to predict the function of RBPs (Wu et al.). In the past decade, these studies about RBPs mainly focused on mutations in RBPs or their target genes. However, it has been increasing appreciated that many driver mutations might perturb molecular interactions or regulatory networks (Mosca et al., 2015; Yi et al., 2017). Recently, a computational method Mutational Effect on RNA Interactome Topology (MERIT) was proposed to analyze the RBP-gene regulatory networks across cancer types (Li et al., 2019a). All these results provide insights into characterizing perturbed RBP-RNA regulatory networks in cancer, as well as the genotype-phenotype relationships underlying human cancers, and RBPs are potential biomarkers for precision medicine.

OPEN ACCESS

Edited and reviewed by:
Georgios D. Lianos, University Hospital of Ioannina, Greece

***Correspondence:**
Yongsheng Li
liyongsheng@hainmc.edu.cn
Juan Xu
xujuanbiocc@ems.hrbmu.edu.cn

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *12 April 2021*
Accepted: *31 May 2021*
Published: *24 June 2021*

Citation:
Li Y and Xu J (2021) Editorial: Perturbation of RNA Binding Protein Regulation in Cancer. Front. Genet. 12:693766. doi: 10.3389/fgene.2021.693766

The methylation of N6 adenosine (m6A) plays a critical role in diverse biological processes (Li et al., 2019b). Moreover, recent studies have revealed that RBPs also play important roles in RNA methylation. IGF2BP3 was identified as a potential oncogene across multiple cancer types and also play important roles in tissue development (Xu et al., 2021; Zhang et al., 2021). These studies provide another regulatory layer of RBPs in cancer.

In summary, RBPs play important roles in cancer development and progression. All these integrated analysis provided detailed knowledge of the function of the RBPs in cancer, which will facilitate the development of rational therapies for cancer.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was supported by Hainan Provincial Natural Science Foundation of China (Grant No. 820MS053), the National Natural Science Foundation of China (Grant No. 31970646, 61873075, 32060152, 32070673, 31871338), Heilongjiang Touyan Innovation Team Program.

REFERENCES

Chen, H. J., Topp, S. D., Hui, H. S., Zacco, E., Katarya, M., McLoughlin, C., et al. (2019). RRM adjacent TARDBP mutations disrupt RNA binding and enhance TDP-43 proteinopathy. *Brain* 142, 3753–3770. doi: 10.1093/brain/awz313

Li, Y. S., McGrail, D. J., Xu, J., Li, J. Y., Liu, N. N., Sun, M., et al. (2019a). MERIT: systematic analysis and characterization of mutational effect on RNA interactome topology. *Hepatology* 70, 532–546. doi: 10.1002/hep.30242

Li, Y. S., Xiao, J., Bai, J., Tian, Y., Qu, Y. W., Chen, X., et al. (2019b). Molecular characterization and clinical relevance of m A regulators across 33 cancer types. *Mol. Cancer* 18:137. doi: 10.1186/s12943-019-1066-3

Mohibi, S., Chen, X. B., and Zhang, J. (2019). Cancer the'RBP'eutics-RNA-binding proteins as therapeutic targets for cancer. *Pharmacol. Ther.* 203:107390. doi: 10.1016/j.pharmthera.2019.07.001

Mosca, R., Tenorio-Laranga, J., Olivella, R., Alcalde, V., Ceol, A., Soler-Lopez, M., et al. (2015). dSysMap: exploring the edgetic role of disease mutations. *Nat. Methods* 12, 167–168. doi: 10.1038/nmeth.3289

Pereira, B., Billaud, M., and Almeida, R. (2017). RNA-binding proteins in cancer: old players and new actors. *Trends Cancer* 3, 506–528. doi: 10.1016/j.trecan.2017.05.003

Xu, K., Cai, Y., Zhang, M., Zou, H., Chang, Z., Li, D., et al. (2021). Pan-cancer characterization of expression and clinical relevance of m6A-related tissue-elevated long non-coding RNAs. *Mol. Cancer* 20:31. doi: 10.1186/s12943-021-01324-8

Xu, X. D., Yu, Y., Zong, K., Lv, P. W., and Gu, Y. T. (2019). Up-regulation of IGF2BP2 by multiple mechanisms in pancreatic cancer promotes cancer proliferation by activating the PI3K/Akt signaling pathway. *J. Exp. Clin. Cancer Res.* 38:497. doi: 10.1186/s13046-019-1470-y

Yi, S., Lin, S. D., Li, Y. S., Zhao, W., Mills, G. B., and Sahni, N. (2017). Functional variomics and network perturbation: connecting genotype to phenotype in cancer. *Nat. Rev. Genet.* 18, 395–410. doi: 10.1038/nrg.2017.8

Zhang, J. W., Li, S. L., Zhang, L., Xu, J., Song, M. X., Shao, T. T., et al. (2020). RBP EIF2S2 promotes tumorigenesis and progression by regulating MYC-mediated inhibition via FHIT-related enhancers. *Mol. Ther.* 28, 1105–1118. doi: 10.1016/j.ymthe.2020.02.004

Zhang, Y., Xu, S. C., Xu, G., Gao, Y. Y., Li, S., Zhang, K., et al. (2021). Dynamic expression of m A regulators during multiple human tissue development and cancers. *Fronti. Cell Dev. Biol.* 8:629030. doi: 10.3389/fcell.2020.629030

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 15 January 2020
doi: 10.3389/fgene.2019.01284

Dysregulated lncRNA-miRNA-mRNA Network Reveals Patient Survival-Associated Modules and RNA Binding Proteins in Invasive Breast Carcinoma

Yu Dong[1], *Yang Xiao*[2,3], *Qihui Shi*[1] *and Chunjie Jiang*[2,3]*

[1] Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai, China, [2] Institute for Diabetes, Obesity, and Metabolism, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, United States, [3] Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, United States

OPEN ACCESS

Edited by:
Juan Xu,
Harbin Medical University, China

Reviewed by:
Meng Zhou,
Wenzhou Medical University, China
Kuixi Zhu,
University of Arizona, United States

***Correspondence:**
Chunjie Jiang
chunjie.jiang@pennmedicine.upenn.edu

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *03 September 2019*
Accepted: *21 November 2019*
Published: *15 January 2020*

Citation:
Dong Y, Xiao Y, Shi Q and Jiang C (2020) Dysregulated lncRNA-miRNA-mRNA Network Reveals Patient Survival-Associated Modules and RNA Binding Proteins in Invasive Breast Carcinoma. Front. Genet. 10:1284. doi: 10.3389/fgene.2019.01284

Breast cancer is the most common cancer in women, but few biomarkers are effective in clinic. Previous studies have shown the important roles of non-coding RNAs in diagnosis, prognosis, and therapy selection for breast cancer and have suggested the significance of integrating molecules at different levels to interpret the mechanism of breast cancer. Here, we collected transcriptome data including long non-coding RNA (lncRNA), microRNA (miRNA), and mRNA for ~1,200 samples, including 1079 invasive breast carcinoma samples and 104 normal samples, from The Cancer Genome Atlas (TCGA) project. We identified differentially expressed lncRNAs, miRNAs, and mRNAs that distinguished invasive carcinoma samples from normal samples. We further constructed an integrated dysregulated network consisting of differentially expressed lncRNAs, miRNAs, and mRNAs and found housekeeping and cancer-related functions. Moreover, 58 RNA binding proteins (RBPs) involved in biological processes that are essential to maintain cell survival were found in the dysregulated network, and 10 were correlated with overall survival. In addition, we identified two modules that stratify patients into high- and low-risk subgroups. The expression patterns of these two modules were significantly different in invasive carcinoma versus normal samples, and some molecules were high-confidence biomarkers of breast cancer. Together, these data demonstrated an important clinical application for improving outcome prediction for invasive breast cancers.

Keywords: lncRNAs, RNA binding protein, miRNAs, integrative analysis, invasive breast carcinoma, biomarker

INTRODUCTION

In women, breast cancer is the most commonly diagnosed cancer and accounts for ~30% of new cancer diagnoses (Siegel et al., 2017). Great improvements have been achieved in diagnosis, surgery, and medical treatment for breast cancer in the past decades. From 1989 to 2016, the death rate for breast cancer dropped by 40% for female breast cancers in the United States. However, it has still been the second leading cause of cancer death in women in the last ten years (Siegel et al., 2017). Invasive breast carcinoma accounts for about 80% of breast cancer (Weigelt et al., 2008) and exhibits high heterogeneity in terms of morphology, clinical features, and prognosis (Milanovic et al., 2013), and the regulatory mechanisms at the genomic level still thus need to be unearthed.

Many studies have investigated the pathogenesis underlying breast cancer and have discovered diagnostic and prognostic markers. In 2006, a study reported altered expression patterns of microRNAs (miRNAs) during initiation and progression and their relationship with cancer diagnosis, staging, and prognosis (Calin and Croce, 2006). Another study investigated the expression of deregulated miRNAs in breast cancer and found correlations of altered miRNA expression with estrogen receptor expression, vascular invasion, and other clinicopathological characteristics (Iorio et al., 2005). Long non-coding RNAs (lncRNAs) represent a new class of non-coding RNAs that are at least 200 nucleotides in length and do not possess a clearly defined open reading frame (Ponting et al., 2009). lncRNAs are critical regulatory factors in cancer initiation and progression (Li and Chen, 2013; Yang et al., 2014). The lncRNA DSCAM-AS1 holds a central position in estrogen receptor (ER)-regulated breast cancer and modulates tamoxifen resistance and tumor progression (Niknafs et al., 2016). Another lncRNA, MAGI2-AS3, can target the Fas/FasL signaling pathway to suppress cell growth in breast cancer (Yang et al., 2018b). Furthermore, a 12-lncRNA signature has been proposed that can be used to identify breast cancer patients at high risk of tumor recurrence, which could be utilized in clinic (Zhou et al., 2016b). Recently, some studies have shown that post-transcriptional regulatory networks can be regulated by molecules at multiple levels (Wei et al., 2017; Liu et al., 2019). By constructing a ceRNA network, a 10-lncRNA signature has been proposed that classified patients into high- and low-risk subgroups with significantly different survival outcomes, highlighting the value of integrating data sets from multiple levels (Zhou et al., 2016a). Mir-21 and lncRNA AWPPH regulate cancer cell chemosensitivity and proliferation in triple-negative breast cancer (Liu et al., 2019). Mir-223 promotes breast cancer cell proliferation by targeting FOXO1 and provides a new potential tumor marker (Wei et al., 2017). The above results imply the significance of integrating molecules at different regulatory levels for interpreting the mechanism of breast cancer, especially in invasive breast carcinoma.

RNA-binding proteins (RBPs) are a type of proteins that bind RNA through its globular RNA-binding domains (RBDs) (Hentze et al., 2018). RBPs can bind mRNA, pre-rRNA, tRNA, small nuclear RNA (snRNA), small nucleolar RNA (snoRNA) and residual ncRNA (Gerstberger et al., 2014) and can alter the fate or function of the bound RNAs during post-transcriptional gene regulation (PTGR), which correlates with the stability, transport, localization, and degradation of different RNAs. They act as important participants in gene regulation (Nishida et al., 2017) and play an important role in maintaining genome integrity (Gerstberger et al., 2014). RBPs have been found to be closely related to many human diseases and to be involved in a wide range of biological processes, such as tumorigenesis, proliferation, development, and apoptosis, by interacting with mRNA (Frisone et al., 2015; Grammatikakis et al., 2017), microRNA (Ciafre and Galardi, 2013), and lncRNA (Luo et al., 2013; Schmitt and Chang, 2016). There are ~20,000 protein-coding genes in humans, and 7.5% of genes are involved in RNA metabolism by binding to RNA (Hentze et al., 2018). But only a few RBPs have received intensive study.

The Cancer Genome Atlas (TCGA) project was launched in 2005 and has accelerated the comprehensive understanding of cancer genomic profiles, thus improving diagnostic methods, therapy standards, and preventive strategies. TCGA has released thousands of high-throughput molecular profiles at different levels, which help researchers better understand cancer pathogenesis, diagnosis, and prognosis. In this study, we integrated the expression profiles of breast cancer at multiple levels (lncRNA, miRNA, and mRNA) across ~1,200 samples, including 1079 invasive breast carcinoma samples as well as 104 normal samples. We identified differentially expressed lncRNAs, miRNAs, and mRNAs and then constructing a lncRNA-miRNA-mRNA dysregulated network, which is a power-law, small-world network. RBPs were found in the dysregulated network, and some of them are related to overall survival time. In addition, two modules were identified and exhibited a correlation with the overall survival time. Further analysis showed that these modules have significantly different expression patterns in cancer versus normal samples. To better understand these two modules, we mined the literature for the molecules in each module and found that some molecules play important roles in breast cancer biology.

MATERIALS AND METHODS

RNA-Seq Expression Data Sets and Pre-Processing

RNA-seq expression data sets of ~1200 patient samples were downloaded from TCGA (https://portal.gdc.cancer.gov/), comprising 1079 invasive breast carcinoma samples and 104 normal samples (**Table S1**). MRNA, lncRNA, and miRNA were included in each sample. Using Perl scripts, we combined ~1200 files into a single profile. The lncRNA expression profile was extracted from the profile based on the latest annotation from the Ensembl database. The biotypes of known lncRNAs are 3prime_overlapping_ncrna, ambiguous_orf, antisense, antisense_RNA, lincRNA, ncrna_host, non_coding, non_stop_decay, processed_transcript, retained_intron, sense_intronic, and sense_overlapping. The biotype of protein-coding genes is protein_coding. In total, 19951 mRNA, 15949

lncRNA, and 1881 miRNA were obtained from TCGA. Based on previously published papers (Yan et al., 2015; Li et al., 2018b; Pan et al., 2018), RNAs with expression 0 in more than 10% of normal samples were eliminated.

Analysis of Differential Expression Between Breast Carcinoma and Normal Samples

Differentially expressed molecules were identified through the use of previously reported methods (Li et al., 2015; Li et al., 2018b). Firstly, RNAs were divided into two groups. RNAs with an expression level equal to 0 in <30% tumor samples were subjected to a t-test, and RNAs with an expression level equal to 0 in >30% tumor samples were subjected to a Fisher's exact test. For the RNAs in the first group, RNAs with a fold change larger than 2 (or smaller than 0.5) and an adjusted p-value smaller than 0.01 were identified as differentially expressed. For RNAs in the second group, we determined their expression in binary fashion: ON (expressed, expression value larger than 0) and OFF (not detected, expression value equal 0). Firstly, the frequencies of ON and OFF in breast carcinoma and normal samples were calculated, respectively. RNAs expressed twice more frequently in cancer than in normal samples were marked as 'ON in cancer'; otherwise, RNAs were marked as 'OFF in Cancer.' Then, for each RNA, the significance of the contingency between ON/OFF and cancer/normal status was calculated by Fisher's exact test with adjustment for multiple testing *via* the Benjamini-Hochberg method. RNAs with FDR smaller than 0.01 was used. In total, 4269 differentially expressed protein-coding genes were identified, as well as 3057 differentially expressed lncRNAs and 367 differentially expressed miRNAs (**Tables S2**–**S5**). Validation of the differentially expressed RNAs was performed by extracting the expression values and normalized them based on Z-score. After that, based on the R package 'pheatmap,' samples were clustered using differentially expressed lncRNAs, miRNAs, and mRNAs, respectively. PCA was also used to cluster the samples.

Ago CLIP-Seq-Supported miRNA Target Sites

miRNA target sites were predicted using a target prediction algorithm from miRanda (Betel et al., 2010) with the default parameters. 3'UTR was used to predict target sites for mRNA, while for lncRNA, the full length of the lncRNA transcript was used. It has been reported that miRNAs function in the form of ribonucleoprotein complexes, RISCs (RNA-induced silencing complexes) (Fabian et al., 2010), and Argonaute (AGO)-family proteins represent the best-characterized protein components and are central to RISC function (Eulalio et al., 2008; Chekulaeva and Filipowicz, 2009). Ultraviolet (UV) crosslinking and immunoprecipitation (CLIP) was used to identify specific protein-RNA interactions (Konig et al., 2012). Hence the function of the Argonate-RNA-miRNA complex can be verified through CLIP technology (Chou et al., 2013). Here we downloaded AGO 1/2 CLIP-Seq datasets from starBase v2.0 (Li et al., 2014a) and identified AGO binding sites to filter candidate miRNA target sites. A target was reserved only if it overlapped with at least one AGO binding site. In total, 41632 miRNA-lncRNA regulatory relationships were predicted, including 1176 lncRNAs and 2509 miRNAs, and 1247237 miRNA-mRNA regulatory relationships were predicted, including 18252 protein-coding genes and 2511 miRNAs (**Tables S6** and **S7**).

Constructing the Dysregulated lncRNA-miRNA-mRNA Network

Based on the interactions of miRNA-lncRNA and miRNA-mRNA, we constructed an initial lncRNA-miRNA-mRNA network. A three-step filtering process was then performed: 1) RNAs that were not differentially expressed were filtered; 2) the expression of each RNA pair (miRNA-lncRNA or miRNA-mRNA) should be significantly correlated (p-value < 0.01 and | correlation coefficient| > 0.4) across samples based on Pearson correlation; 3) only miRNAs that were shared by both lncRNA and mRNA were considered. The dysregulated network was constructed based on 876 interactions and 539 differentially expressed molecules, including 75 miRNAs, 63 lncRNAs, and 401 protein-coding genes (**Table S8**). The network was visualized using Cytoscape (Shannon et al., 2003). CytoCluster (Li et al., 2017), a Cytoscape plugin for cluster analysis and visualization of biological networks, was used to identify modules, employing the graphically based IPC-MCE algorithm and adopting the default parameter values (0.6 as the Threshold).

Survival Analysis

The clinical data of all of the breast cancer patients were downloaded from TCGA. Perl scripts were used to extract the information regarding days to last follow up and vital status (alive or dead) for each invasive breast carcinoma patient. For each module, the average value in each sample was used. Modules that relate to the overall survival were identified by clustering the samples into two classes based on K-means Clustering. An R package, 'survival' was then used to 1) construct a surv object using the function 'Surv' based on the status and time, 2) create fitted survival curves with the Kaplan-Meier algorithm, using the function 'survfit' based on the surv object and class label, and 3) test for a difference between the two survival curves using a log-rank test. P-value < 0.05 was set as the cutoff. All reported p-values were two-sided.

Functional Enrichment Analysis

In order to investigate functional roles, GO and KEGG analyses were performed based on protein-coding genes in the network using the Database for Annotation, Visualization, and Integrated Discovery (DAVID, version 6.8) (Huang da et al., 2009; Huang et al., 2009). Cancer hallmarks related GO terms were identified by two previous studies (Subramanian et al., 2005; Plaisier et al., 2012). Additionally, PANTHER (Mi and Thomas, 2009) (https://reactome.org/) and REACTOME (Croft et al., 2011) (http://pantherdb.org/) pathway analysis were performed. To further investigate the functional roles, GAD, a database of genetic association data from complex diseases and disorders, was also used by DAVID (Huang da et al., 2009; Huang et al., 2009).

Cancer Genes

Two cancer gene lists were used to further validate the roles in cancer. The first one was compiled by by Mertins et al. (2016), who collected 415 oncogenes and tumor suppressors from UniProt (https://www.uniprot.org/) and published papers. Another list of 524 genes that had been implicated in malignant transformation according to a catalog of somatic mutations in cancer (COSMIC, http://cancer.sanger.ac.uk/cancergenome/projects/census) was collected by Uhlen et al. (Uhlen et al., 2015). In total, 724 potentially cancer-related genes were used (**Table S9**).

RESULTS

Differentially Expressed RNAs Distinguish Invasive Breast Carcinoma From Normal Tissues

We acquired the expression profiles of mRNA, lncRNA, and miRNA from TCGA, which contains 1183 samples, comprising 1079 invasive breast carcinoma samples and 104 normal samples (**Table S1**). Differentially expressed molecules were identified using the method detailed in Li et al. (2018b). RNAs with an expression level equal to 0 in <30% tumor samples were subjected to t-test, and RNAs with an expression level equal to 0 in >30% tumor samples were subjected to Fisher's exact test (see Methods). In total, 4269 protein-coding genes that were differentially expressed between invasive breast carcinoma and normal samples were identified, including 2349 up-regulated and 1920 down-regulated genes (**Tables S2** and **S3**). For lncRNAs, 3057 differentially expressed molecules were identified, of which 2033 were up-regulated and 1024 were down-regulated (**Tables S2** and **S4**). Additionally, 367 differentially expressed miRNAs were identified. 152 miRNAs were up-regulated, and 215 were down-regulated (**Tables S2** and **S5**).

We validated our differentially expressed molecules by performing unsupervised hierarchical cluster analyses for the 1179 invasive breast carcinoma samples and 104 normal samples using the R package 'pheatmap' with the default distance. The invasive breast carcinoma samples were clearly distinguished from normal samples in terms of differentially expressed lncRNAs, protein-coding genes, and miRNAs, respectively (**Figures 1A–C**). To further check these differentially expressed molecules, principle component analysis (PCA) analyses were performed using the R function 'prcomp.' Consistent with the unsupervised hierarchical clustering, the first two principal components could distinguish the tumor samples from normal samples (**Figures 1D–F**).

The Dysregulated Network Is a Biological Network Performing Housekeeping and Cancer-Related Functions

All of the differentially expressed molecules mentioned above were used to construct the dysregulated network. We predicted miRNA target sites for all protein-coding genes and lncRNAs based on the algorithm from miRanda (Betel et al., 2010), using the default parameters (see Methods). It has been reported that miRNA functions in the form of ribonucleoprotein complexes, RISCs (RNA-induced silencing complexes) (Fabian et al., 2010), and Argonaute (AGO)-family proteins represent the best-characterized protein components and are central to RISC function (Eulalio et al., 2008; Chekulaeva and Filipowicz, 2009). Ultraviolet (UV) crosslinking and immunoprecipitation (CLIP) was used to identify specific protein-RNA interactions (Konig et al., 2012). Hence, the function of the Argonate-RNA-miRNA complex can be verified through CLIP technology. Candidate miRNA and target site pairs were filtered by the AGO 1/2 CLIP-seq data from starBase (Li et al., 2014a) (see *Materials and Methods*). A total of 41632 interactions, including 2509 miRNAs and 1176 lncRNA targets, were predicted as well as 1247237 interactions between 2577 miRNAs and 18252 protein-coding genes (**Tables S6** and **S7**). Next, based on the three-step filtering process (see *Materials and Methods*), a dysregulated network was constructed from 876 interactions and 539 differentially expressed molecules, including 75 miRNAs, 63 lncRNAs, and 401 protein-coding genes (**Figure 2A**, **Table S8**).

It has been shown that many biological networks are small-world networks (Latora and Marchiori, 2001; Wagner and Fell, 2001), which have also been reported to be scale-free networks (Amaral et al., 2000). We tested whether our dysregulated network is a scale-free network by analyzing the degree distribution, which is one of the most important characteristics of a scale-free network and is defined as the number of edges incident to a node. As shown in **Figure S1A**, more than 88% of nodes had less than five edges, whereas only 6% of nodes had more than 10 edges. It fitted a power-law distribution with R^2 = 0.86 and correlation = 0.99, suggesting that our network is a scale-free network (Barabasi, 2009). In addition, most of the shortest paths were between 4 and 6 (**Figure S1B**), which is consistent with the property of a small-world network. Closeness is a measure of how close an individual is to other individuals in a network (Borgatti, 1995; Costenbader and Valente, 2003). The more central a node is, the closer it is to all other nodes. As shown in **Figure S1C**, the majority of nodes were highly central. Together, these data revealed that our dysregulated network is a scale-free and small-world network, indicating that our network is a canonical biological network.

The functions of the dysregulated network were investigated by using the protein-coding genes in this network to perform functional enrichment analysis (see Methods). All top-ten enriched gene ontology (GO) terms were related to cell cycle, mitotic nuclear division, and nuclear division (**Figure 2B**). These were all housekeeping functions for maintaining cell survival. We further acquired all of the housekeeping genes identified by Jiang et al. (2018) and found that 84 (21%) protein-coding genes in our dysregulated network were housekeeping genes (**Table S10**). In addition, based on a previous study (Salem et al., 2016), we obtained a list of GO terms related to hallmarks of cancer and found that these terms were also enriched in our network (**Figure 2C**). For example, signal transduction (GO:0007165) and positive regulation of cell proliferation (GO:0008284) are

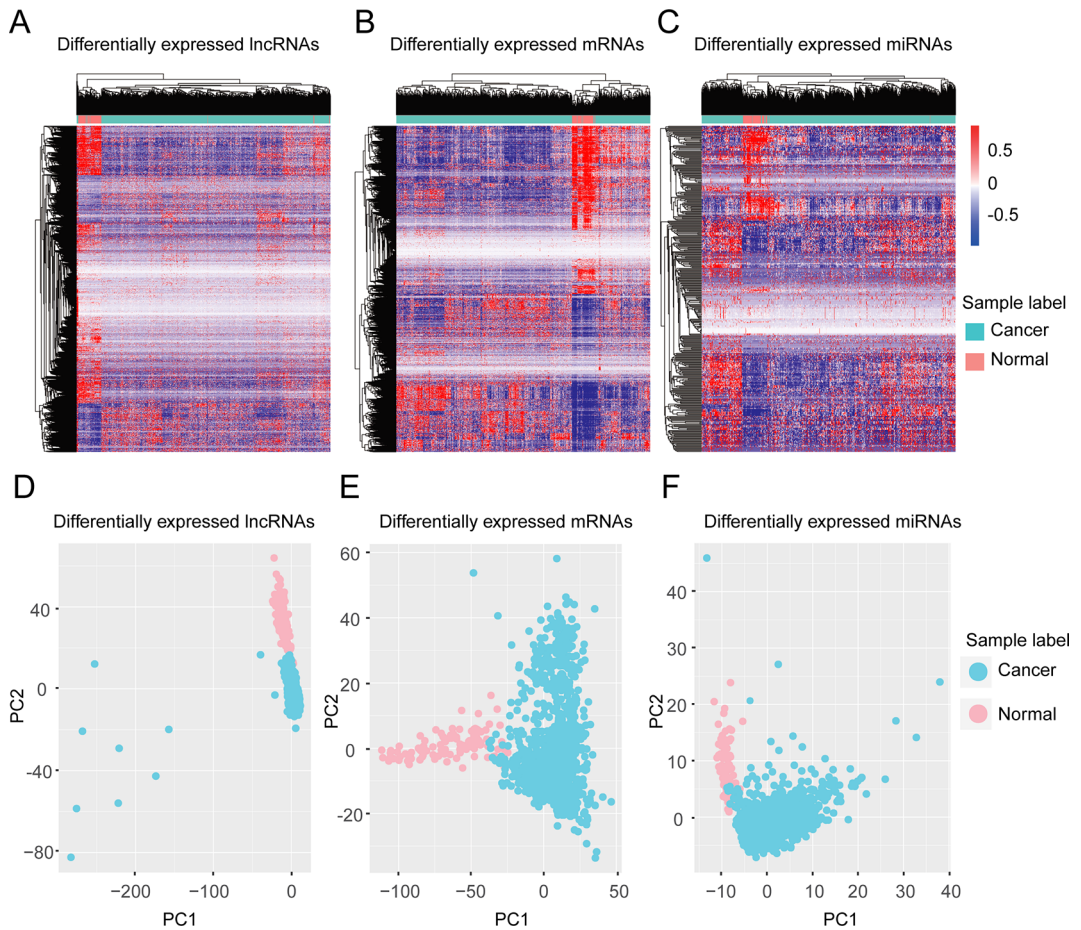


FIGURE 1 | Clustering based on differentially expressed molecules. Unsupervised hierarchical clustering of all samples based on differentially expressed lncRNAs **(A)** protein-coding genes **(B)** and miRNAs **(C)**. The unsupervised hierarchical clustering was performed using an R package, 'pheatmap' with the default distance setting, Euclidean distance. **(D–F)** PCA analysis based on differentially expressed lncRNAs **(D)**, protein-coding genes **(E)**, and miRNAs **(F)**. PCA analysis was performed by the R function 'prcomp'.

Self Sufficiency in Growth Signals, while negative regulation of cell proliferation (GO:0008285) and negative regulation of cell cycle (GO:0045786) are Insensitivity to Antigrowth Signals. Taken together, our dysregulated network demonstrated important and functional roles.

Moreover, we performed pathway enrichment analyses using three different pathway databases, the Kyoto Encyclopedia of Genes and Genomes (KEGG), PANTHER (Mi and Thomas, 2009), and REACTOME (Croft et al., 2011). For the KEGG pathway, housekeeping and cancer-related functions were again enriched (**Figure S2A**). The housekeeping functions were cell cycle and axon guidance, and the cancer-related functions were pathways in cancer, Melanoma, Colorectal cancer and Prostate cancer (**Figure S2A**). For the PANTHER and REACTOME pathway databases, most of the top terms were housekeeping functions (**Figures S2B**, **C**).

To further validate their important roles in cancer, we obtained 415 oncogenes and tumor suppressors from Mertins et al. (2016) and 524 genes that have been implicated in malignant transformation from Uhlen et al. (2015). In total, 724 cancer genes were used (**Table S9**). 656 of them were expressed in our dataset, and 31 were in our network (**Figure S2D**). Based on a hypergeometric test, the p-value was 7.46E-05, which suggested that our dysregulated network was significantly enriched in cancer-related genes. We further performed functional enrichment analysis using DAVID (Huang da et al., 2009; Huang et al., 2009) based on the Genetic Association Database (GAD), which is a database of genetic association data from complex diseases and disorders. Surprisingly, breast cancer was the most enriched term (**Figure 2D**), which corroborated the important roles of our dysregulated network in cancer biology.

RBPs in Our Dysregulated Network

Next, we investigated the RBPs in our dysregulated network. Based on published papers (Cook et al., 2011; Gerstberger et al., 2014; Fredericks et al., 2015; Hentze et al., 2018), 58 RBPs were found in our dysregulated network, of which 28 were upregulated and 30 were downregulated (**Table S11**). To improve our

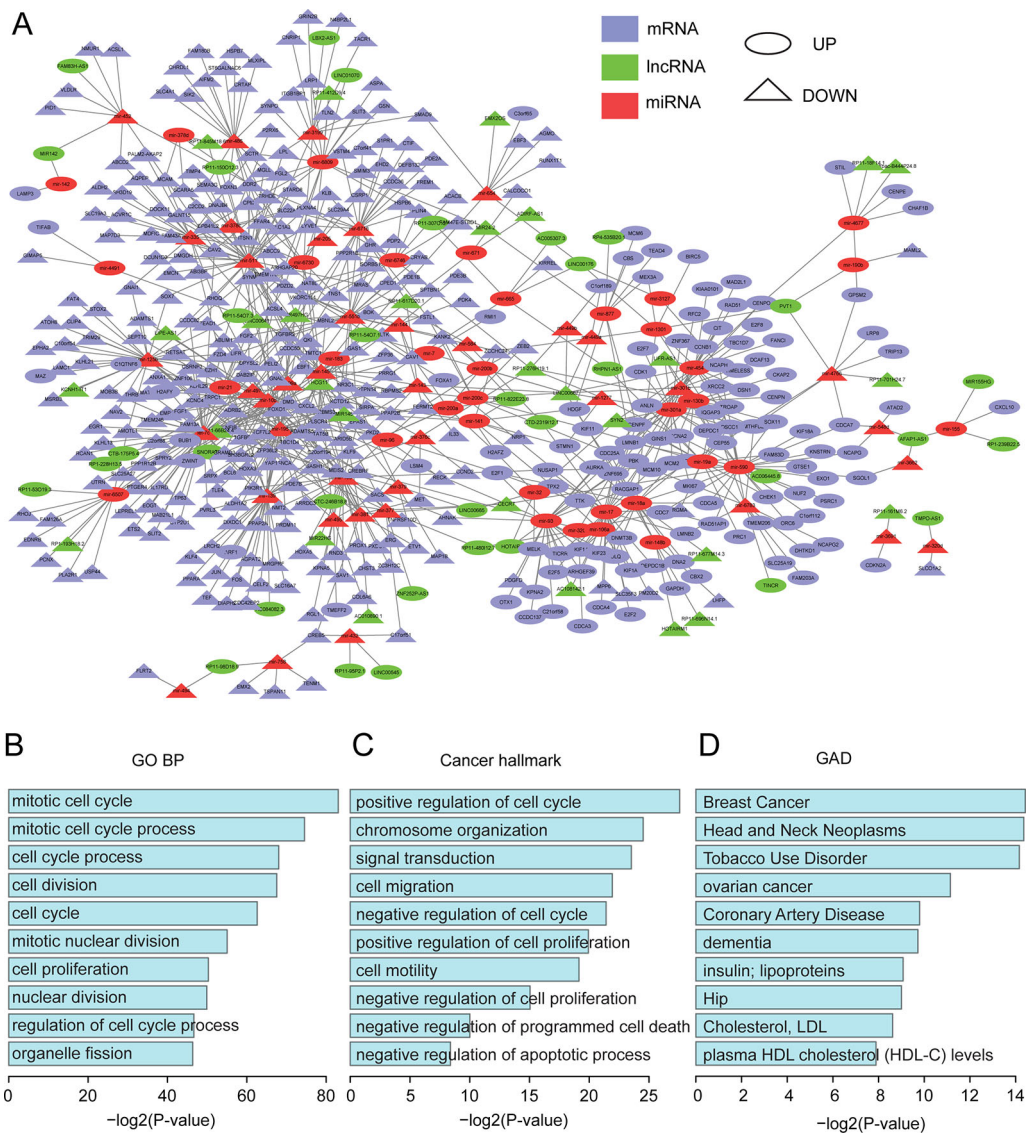


FIGURE 2 | Functional analysis for the dysregulated network. **(A)** The dysregulated lncRNA-miRNA-mRNA network. The network was visualized using Cytoscape. **(B)** The top 10 enriched GO terms. **(C)** The top 10 enriched cancer hallmark-related GO terms. **(D)** The top 10 enriched GAD terms.

understanding of the roles of RBP in invasive breast carcinoma, STRING (https://string-db.org/) was used to construct a protein–protein interaction (PPI) network (**Figure 3A**). Random networks of the same size were generated by STRING, which was used to assess whether the given network had more internal interactions than would be expected for a random set of the same size. A small PPI enrichment p-value indicates that the nodes are not random and that the observed number of edges is significant. Based on STRING, the PPI enrichment p-value was 1.0e-16, which means that these RBPs have more interactions than would occur in a random set. This enrichment indicated that these RBPs are at least partially biologically connected as a group. The GO analysis showed that all top 10 molecular function (MF) terms were binding-related functions and the top two were poly(A) RNA binding and RNA binding, which further confirms that they are RBPs (**Figure 3B**), and that these RBPs are involved biological processes that are essential to maintain cell survival like cell cycle, cell division, DNA packaging, and chromosome organization (**Figure 3C**). Moreover, GAD enrichment analysis was also performed, and it is worth noting that breast cancer was again the most enriched term (**Figure S3**).

To investigate whether these RBPs were associated with prognosis in invasive breast carcinoma patients, the overall survival for each RBP was calculated using the R package 'survival' (see Methods). Ten RBPs (CDKN2A, DCAF13, DNMT3B, EXO1, FANCI, KPNA2, RACGAP1, SORBS1, TP63, and ZNF106) were significantly associated with overall survival, including seven upregulated and three downregulated

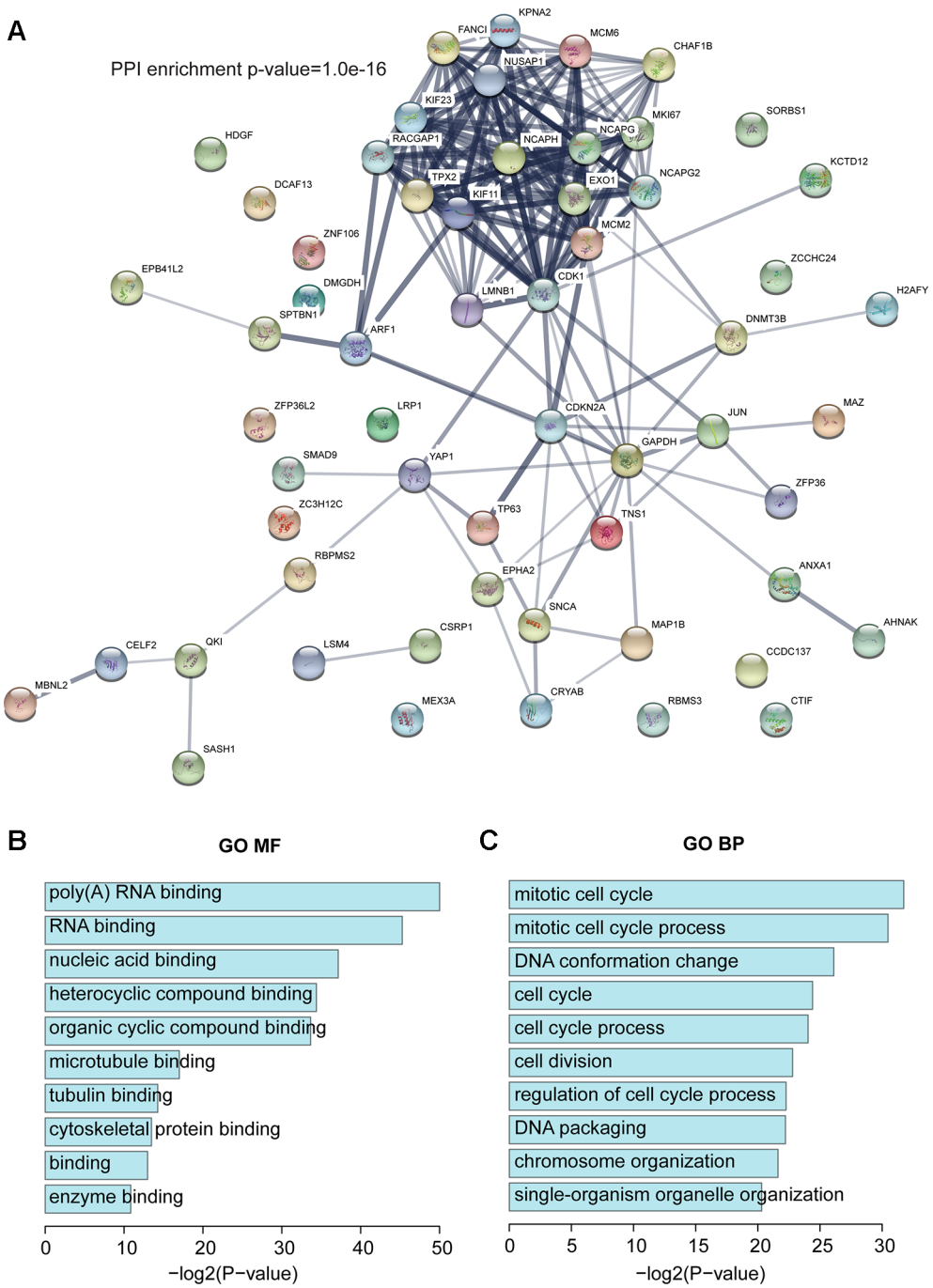


FIGURE 3 | RBPs in the dysregulated network. **(A)** Protein–protein interaction (PPI) network of RBPs based on STRING. **(B)** The top 10 enriched GO MF terms. **(C)** The top 10 enriched GO BP terms. BP, biological process.

RBPs (**Figures S4** and **S5**). Notably, some were reported to play roles in breast cancer (see Discussion). Overexpression of DCAF13, DNMT3B, KPNA2, EXO1, FANCI, RACGAP1, and ZNF106 in invasive breast carcinoma patients showed poor survival, while overexpression of CDKN2A, SORBS1, and TP63 showed better survival (**Figures S4** and **S5**).

Modules in the Dysregulated Network Relate to the Survival of Invasive Breast Carcinoma Patients

To further investigate the roles of our dysregulated network, CytoCluster (Li et al., 2017), a Cytoscape plugin for cluster analysis and visualization of biological networks, was used to

identify modules (see Methods). Subsequently, to explore the relationship between the modules and the prognosis of patients with invasive breast carcinoma, the overall survival for each module in invasive breast carcinoma patients was investigated (see Methods). We found that two modules were significantly ($p < 0.05$) correlated with overall survival (**Figures 4A–D**). Moreover, their expression patterns in normal and invasive breast carcinoma samples were assessed. These two modules

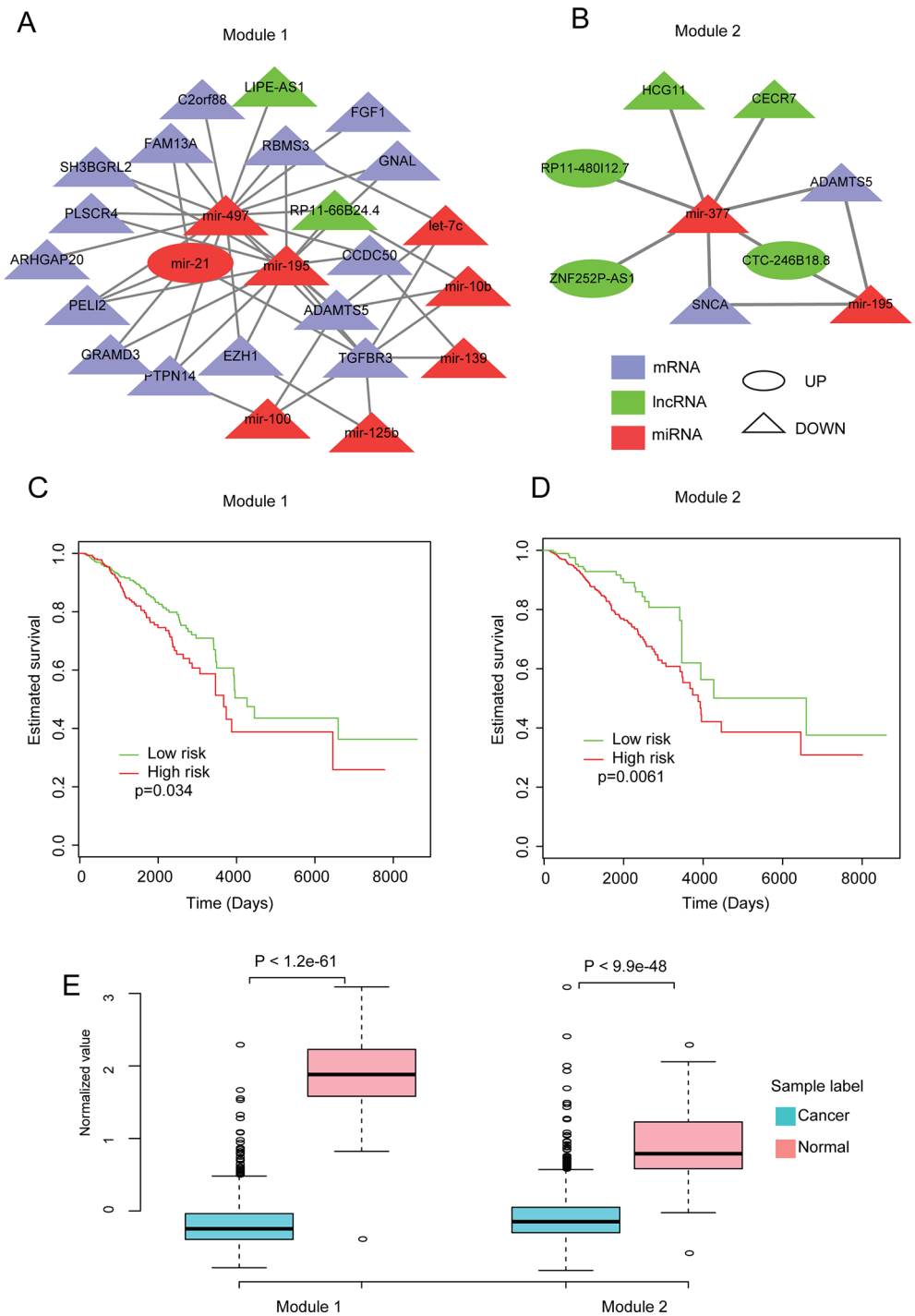


FIGURE 4 | Analysis of modules identified from the dysregulated network. **(A, B)** The two modules identified from the dysregulated network using Cytoscape with default parameters. **(C, D)** Kaplan-Meier plot of survival for these two modules. **(E)** Expression patterns of the modules in normal and cancer samples. The average expression value of each molecule crossing all normal/cancer samples was used.

showed significant differences in expression patterns between normal and invasive breast carcinoma samples (**Figure 4E**). Both showed significantly lower expression in invasive breast carcinoma samples, indicating that lower expression of these modules contributes to the development of invasive breast carcinoma.

In addition, to further investigate the functions of these two modules in breast cancer, literature-mining was used for the molecules in each module. Module 1 had 25 nodes, including eight miRNAs, two lncRNAs, and 15 protein-coding genes. Twenty-two of the molecules, including all of the miRNAs, have been shown to play important roles in breast cancer. For example, mir-195 inhibits tumor growth and metastasis in breast cancer cells (Singh et al., 2015; Wang et al., 2016c). Mir-497 contributes to cell proliferation, migration, and invasion of estrogen receptor alpha-negative breast cancer by targeting estrogen-related receptor alpha (Han et al., 2016; Wu et al., 2016b). TGFBR3 inhibits breast cancer progression through TGF-beta signaling (Lee et al., 2010). In addition, other molecules in module 1 such as ADAMTS5 (Fontanil et al., 2017), ARHGAP20 (Asaduzzaman et al., 2017), C2orf88 (Lo et al., 2015), EZH1(Liu et al., 2012), FAM13A (Goto-Yamaguchi et al., 2018), FGF1 (Slattery et al., 2013), GNAL (Yi et al., 2009), GRAMD3 (Boiles et al., 2015), PELI2 (Zang et al., 2017), PLSCR4 (Sahay et al., 2015), PTPN14 (Belle et al., 2015), RBMS3 (Zhu et al., 2019a), SH3BGRL2 (Alexe et al., 2007; Wen et al., 2018), let-7c (Fu et al., 2017), mir-100 (Jiang et al., 2016b), mir-10b (Wang et al., 2016b), mir-125b (Wang et al., 2019a), mir-139 (Dai et al., 2017), and mir-21 (Yan et al., 2008; Yanwirasti and Arisanty, 2017; Zhu et al., 2019b) have been reported to play important roles in breast cancer. Module 2 had nine nodes, including two miRNAs, five lncRNAs, and two protein-coding genes. Five of these molecules have been shown to play important roles in breast cancer. For example, Wang et al., (2019b) reported that overexpression of miR-377 correlates with better prognosis in triple-negative breast cancer. ADAMTS-5 may alter the cellular microenvironment, affecting the balance between protumor and antitumor effects (Fontanil et al., 2017). SNCA is the hub gene and is involved in promoting tumor invasion in breast cancer (Serra-Musach et al., 2012; Dang et al., 2016). Besides, molecules in module 2 like the lncRNAs (HCG11) (Liu et al., 2016) and mir-195 (Singh et al., 2015; Wang et al., 2016c) have also been reported to play roles in breast cancer. We also performed key driver analysis (KDA) (Bin Zhang , 2013) to identify key drivers in our network, and all of the miRNAs from our two modules were identified as key drivers. All of these results imply the important roles and vital functions of these two modules in breast cancer biology.

DISCUSSION

Breast cancer is a leading type of cancer in women worldwide (Stewart and Wild, 2014). Many improvements have been made in diagnostic techniques, surgical skills, and medical treatments relating to breast cancer in the past decades. However, it still caused 522,000 deaths in 2012 (Stewart and Wild, 2014). It is imperative to improve the diagnosis and treatment of breast cancer further. Therefore, the identification of cancer-related molecules and the exact regulatory mechanism of breast cancer initiation and development are attracting increasing attention.

It has been reported that lncRNAs and miRNAs play important roles in breast cancer, as do protein-coding genes (Cizkova et al., 2013; Li et al., 2014b; Kim et al., 2015; Yang et al., 2018b). Here we integrated the expression data of lncRNA, miRNA, and protein-coding genes based on ~1200 invasive breast carcinoma and normal samples from TCGA. A total of 4269 differentially expressed protein-coding genes, 3057 differentially expressed lncRNAs, and 367 differentially expressed miRNAs were identified. Based on unsupervised hierarchical clustering and PCA, the samples from invasive breast cancer were distinguished from the normal samples. To construct a dysregulated network, we predicted miRNA targets using an algorithm from miRanda (Betel et al., 2010) with the default parameters. As mature miRNA is part of an active RNA-induced silencing complex (RISC) (Rana, 2007) and the Ago family is central to RISC function (Tang, 2005), AGO CLIP-Seq data were applied to achieve highly convincing miRNA targets. Based on the differentially expressed lncRNAs, miRNAs, and protein-coding genes, an initial dysregulated lncRNA-miRNA-mRNA network was built. After three-step filtering, the final network was constructed, consisting of 876 interactions and 539 differentially expressed molecules.

Next, we analyzed this network through different aspects—the distribution of degree, shortest path, and closeness centrality—which showed that the dysregulated network is a scale-free, small-world network and a meaningful biological network. To further understand the function of the dysregulated network, functional enrichment analysis was performed. The top-10 GO terms showed housekeeping functions in our network. Furthermore, terms related to cancer hallmarks were also found, based on a previous study (Salem et al., 2016). Enrichment analysis with three different pathway databases supported the housekeeping and cancer-related functions in our dysregulated network. Based on two previous studies, 716 potential cancer genes were obtained, and further analysis showed enrichment in these cancer-related genes. Furthermore, we found that breast cancer was the most enriched term based on GAD, suggesting the important role of our dysregulated network in cancer biology.

It was known that RBPs play a central role in the regulation of gene expression, and dysregulated expression of RBPs has been related to the development of cancers (Galante et al., 2009; Bebee et al., 2014; Wang et al., 2015; Correa et al., 2016). In the present study, we identified 58 RBPs in our dysregulated network, and these were confirmed by GO BP analysis. These RBPs are involved in biological processes that are essential to maintaining cell survival. Based on STRING, we found that these RBPs had more interactions among themselves than what would be expected, indicating that they are at least partially biologically connected. Interestingly, GAD enrichment analysis again showed that breast cancer was the most enriched

term. In addition, 10 RBPs were found to be associated with the overall survival of invasive breast carcinoma patients, which suggested that they might be associated with tumor progression, invasion, and aggressiveness. Indeed, some have been reported to play roles in breast cancer. TP63 is a sequence-specific DNA binding transcriptional activator or repressor (Zhou et al., 2016b). In breast cancer, high expression of TP63 coupled with STAT6 has been shown to be associated with longer metastasis-free survival, indicating that TP63 could be involved in inhibiting the migration of breast cancer cells (Papageorgis et al., 2015). By silencing TP63 expression, breast cancer cells acquired increasing resistance to cisplatin, suggesting its role in drug reaction (Mendoza-Rodriguez et al., 2019). SORBS1 is an adaptor protein, and its overexpression inhibits the invasive capacity of tumor cells in breast cancer patients. Silencing SORBS1 promoted EMT and weakened chemotherapy sensitivity (Song et al., 2017). DCAF13, located in chromosome 8q22.3, has been shown to be amplified in breast cancer. Overexpression of DCAF13 was associated with worse prognosis and might be involved in regulating cell cycle progression (Chin et al., 2007; Cao et al., 2017). By targeting DNMT3b, miR-221 became involved in tumorigenicity through regulating the stemness of breast cancer cells (Roscigno et al., 2016). Additionally, DNMT3B helped maintain the CAF function of promoting breast cancer malignance (Tang et al., 2019). RBPs are important in tumor development, and their role still needs to be explored more.

Using CytoCluster (Li et al., 2017), we identified modules that were significantly related to the overall survival time. These two modules had significantly different expression patterns in cancer and in normal samples. Moreover, the literature mining revealed that some molecules in each module play important roles in breast cancer. In module 1, there were 25 nodes, including eight miRNAs, two lncRNAs, and 15 mRNAs. It had been reported that the upregulation of mir-497 inhibited cell proliferation, migration, and invasion in breast cancer (Han et al., 2016; Wang et al., 2016a; Wu et al., 2016b) and that mir-195 inhibited tumor growth, invasion, and metastasis by targeting other RNAs in breast cancer (Singh et al., 2015; Wang et al., 2016c). Importantly, mir-497 and mir-195 were the hub nodes in this module, indicating their essential role in the module. However, these two miRNAs were down-regulated in breast cancer, which means that the inhibition was lost, contributing to the development of breast cancer. Consistent with the tumor-suppressive role of these two miRNAs, TGFBR3 was reported to suppress breast cancer progression through TGF-beta signaling (Lee et al., 2010), and RBMS3 and PTPN14 were also shown to play roles in inhibiting metastasis (Belle et al., 2015; Zhu et al., 2019a). These data imply that the function of module 1 may be to inhibit cancer progress and metastasis and that these functional miRNAs may affect breast cancer through TGFBR3, RBMS3, and PTPN14. Additionally, LIPE-AS1 (lncRNA), RP11-66B24.4 (lncRNA), and CCDC50 (mRNA) have not been reported in BRCA, but LIPE-AS1 interacted with miRNA-497 and slightly correlated with overall survival ($p = 0.075$) and both RP11-66B24.4 and CCDC50 are regulated by the two hub miRNAs, which suggested that they might act as main or auxiliary regulators in the progression and metastasis of BRCA. In module 2, mRNA ADAMTS5 was reported to play roles during migration and invasion in breast cancer (Fontanil et al., 2017). It also functions as a tumor suppressor by inhibiting migration, invasion, and angiogenesis in human gastric cancer (Huang et al., 2019). Besides, two other studies have shown that the upregulation of ADAMTS5 promotes progression in colorectal cancer and drives metastasis in colon and non-small cell lung cancer (Gu et al., 2016; Yu et al., 2016). Another mRNA, SNCA, was also reported to be involved in tumor development by inhibiting invasion and inducing apoptosis (Li et al., 2018c; Yan et al., 2018). Thus, the function of module 2 might relate to cancer progression and survival. Previous studies have shown that miRNAs may function as tumor suppressors or oncogenes in tumor development, invasion, and metastasis. In module 2, mir-377 is the hub node and may be the core molecule involved in breast cancer due to its interactions with other molecules. Moreover, mir-377 has been reported to inhibit proliferation and metastasis in gastric cancer and pancreatic cancer (Chang et al., 2016; Wang et al., 2017). mir-195 was also important in BRCA, as it was shown that mir-195 could inhibit the invasion and metastasis of breast cancer (Singh et al., 2015; Wang et al., 2016c). lncRNA CECR7 interacts with mir-377 and had been reported to play a role in hepatocellular carcinoma (Zhang et al., 2015). In addition to the molecules reported to play roles in breast cancer, some novel candidate biomarkers, which may also be important to breast cancer, were found, but more evidence is needed in future.

Many studies have performed integrative analyses of TCGA breast cancer data through networks. For example, Yin et al. (2016) focused on identifying miRNA-mRNA pairs and constructed a miRNA target network in invasive breast carcinoma. Li et al. (2018a) found that some of the correlations between microRNA and target genes declined in cancer compared to normal across multiple cancers. Wu et al. (2016a) found two kinds of lncRNA-mRNA co-expression patterns: 1) correlations between lncRNA-mRNA in cancer were reversed compared to normal; 2) correlations between lncRNA-mRNA in cancer were similar to normal. Xiao et al. (2018) compared the differential genes between ER+ and ER- and constructed a ceRNA network and found that some molecules correlated with prognosis. Yang et al. (2018a) compared the differentially expressed genes in Triple-Negative Breast Cancer and also constructed a ceRNA network. Some molecules correlated with prognosis were identified and validated by qRT-PCR. Sun et al. (2019) identified eight lncRNAs as the prognosis signature for breast cancer using a ceRNA and WGCNA network. Gao et al. (2019) built a ceRNA and found some prognosis-related molecules (four lncRNAs, two miRNAs, and two mRNAs). Most studies built a ceRNA network, which contains molecules that are not differentially expressed. However, the integrated dysregulated network in this study consists of differentially expressed lncRNAs, miRNAs, and mRNAs only, and we identified RBPs and modules that can stratify patients into high- and low-risk subgroups. Moreover,

each module not only relates to prognosis but also contains RNAs that have been reported to play roles in breast cancer.

It is well known that the expression of non-coding RNAs is highly tissue- and cell-type specific, providing important clues about their specific functions in response to contextual demands (Mercer et al., 2008; Cabili et al., 2011; Jiang et al., 2016a). Here, we identified patient survival-associated modules including non-coding RNAs in invasive breast carcinoma, and this interpretation was supported in many ways. All molecules in the modules were differentially expressed in invasive breast carcinoma, indicating the potential roles of these molecules. The modules came from a scale-free biological network that performs functions that are related to housekeeping and are cancer hallmarks. More importantly, these two modules were significantly correlated with overall survival. Moreover, many papers have shown clues that molecules in our networks play roles in the progression of breast cancer, and KDA analysis also showed that the molecules in our networks are key drivers. Based on these strands of evidence, our results are credible. However, there are limitations to this study. Firstly, it is a network-based study. Secondly, our study is only based on bioinformatics analysis. Experiments are needed to support the identifications of functional roles.

CONCLUSIONS

In summary, using a network-based strategy, we provided a framework integrating miRNAs, mRNAs, and lncRNAs that are differentially expressed in breast cancer to identify biomarkers. Although further validation is still needed to support the potential roles of the RBPs and two modules, many strands of evidence show the correlations between our two modules and breast cancer. Overall, our dysregulated network provides new insights into outcome prediction for invasive breast cancers.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov/.

AUTHOR CONTRIBUTIONS

CJ and YD were responsible for the statistical analysis, contributed to the acquisition of data, and were major contributors to writing the manuscript. QS and YX assisted with data analysis and revised the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We sincerely thank the public database: TCGA.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01284/full#supplementary-material

REFERENCES

Alexe, G., Dalgin, G. S., Ganesan, S., Delisi, C., and Bhanot, G. (2007). Analysis of breast cancer progression using principal component analysis and clustering. *J. Biosci.* 32 (5), 1027–1039. doi: 10.1007/s12038-007-0102-4

Amaral, L. A. N., Scala, A., Barthelemy, M., and Stanley, H. E. (2000). Classes of small-world networks. *Proc. Natl. Acad. Sci. U. S. A.* 97 (21), 11149–11152. doi: 10.1073/pnas.200327197

Asaduzzaman, M., Constantinou, S., Min, H., Gallon, J., Lin, M. L., Singh, P., et al. (2017). Tumour suppressor EP300, a modulator of paclitaxel resistance and stemness, is downregulated in metaplastic breast cancer. *Breast Cancer Res. Treat* 163 (3), 461–474. doi: 10.1007/s10549-017-4202-z

Barabasi, A. L. (2009). Scale-free networks: a decade and beyond. *Science* 325 (5939), 412–413. doi: 10.1126/science.1173299

Bebee, T. W., Cieply, B. W., and Carstens, R. P. (2014). Genome-wide activities of RNA binding proteins that regulate cellular changes in the Epithelial to Mesenchymal Transition (EMT). *Syst. Biol. RNA Binding Proteins* 825, 267–302. doi: 10.1007/978-1-4939-1221-6

Belle, L., Ali, N., Lonic, A., Li, X., Paltridge, J. L., Roslan, S., et al. (2015). The tyrosine phosphatase PTPN14 (Pez) inhibits metastasis by altering protein trafficking. *Sci. Signal* 8 (364), ra18. doi: 10.1126/scisignal.2005547

Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* 11 (8), R90. doi: 10.1186/gb-2010-11-8-r90

Bin Zhang, J. Z. (2013). Identification of key causal regulators in gene networks by Bin Zhang and Jun Zhu. *Proceedings of the World Congress on Engineering II.*

Boiles, A. R., Prasanna, D., Shaker, A. B., Oswald, M., Mason, C., Keogh, M., et al. (2015). Integrated analysis of miRNAs expression in breast cancer patients to detect genes deregulation involved in malignant transformation. *J. Clin. Oncol.* 33 (15). doi: 10.1200/jco.2015.33.15_suppl.e12560

Borgatti, S. P. (1995). Centrality and AIDS. *Connections* 18 (1), 112–114.

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., et al. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25 (18), 1915–1927. doi: 10.1101/gad.17446611

Calin, G. A., and Croce, C. M. (2006). MicroRNA signatures in human cancers. *Nat. Rev. Cancer* 6 (11), 857–866. doi: 10.1038/nrc1997

Cao, J. Z., Hou, P. J., Chen, J. M., Wang, P. H., Wang, W. Q., Liu, W., et al. (2017). The overexpression and prognostic role of DCAF13 in hepatocellular carcinoma. *Tumor Biol.* 39 (6), 1010428317705753. doi: 10.1177/1010428317705753

Chang, W., Liu, M., Xu, J., Fu, H., Zhou, B., Yuan, T., et al. (2016). MiR-377 inhibits the proliferation of pancreatic cancer by targeting Pim-3. *Tumour Biol.* 37 (11), 14813–14824. doi: 10.1007/s13277-016-5295-4

Chekulaeva, M., and Filipowicz, W. (2009). Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. *Curr. Opin. Cell Biol.* 21 (3), 452–460. doi: 10.1016/j.ceb.2009.04.009

Chin, S. F., Teschendorff, A. E., Marioni, J. C., Wang, Y., Barbosa-Morais, N. L., Thorne, N. P., et al. (2007). High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biol.* 8 (10), R215. doi: 10.1186/gb-2007-8-10-r215

Chou, C. H., Lin, F. M., Chou, M. T., Hsu, S. D., Chang, T. H., Weng, S. L., et al. (2013). A computational approach for identifying microRNA-target

interactions using high-throughput CLIP and PAR-CLIP sequencing. *BMC Genomics* 14 (Suppl 1), S2. doi: 10.1186/1471-2164-14-S1-S2

Ciafre, S. A., and Galardi, S. (2013). microRNAs and RNA-binding proteins A complex network of interactions and reciprocal regulations in cancer. *RNA Biol.* 10 (6), 935–943. doi: 10.4161/rna.24641

Cizkova, M., Vacher, S., Meseure, D., Trassard, M., Susini, A., Mlcuchova, D., et al. (2013). PIK3R1 underexpression is an independent prognostic marker in breast cancer. *BMC Cancer* 13. doi: 10.1186/1471-2407-13-545

Cook, K. B., Kazan, H., Zuberi, K., Morris, Q., and Hughes, T. R. (2011). RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.* 39, D301–D308. doi: 10.1093/nar/gkq1069

Correa, B. R., de Araujo, P. R., Qiao, M., Burns, S. C., Chen, C., Schlegel, R., et al. (2016). Functional genomics analyses of RNA-binding proteins reveal the splicing regulator SNRPB as an oncogenic candidate in glioblastoma. *Genome Biol.* 17, 125. doi: 10.1186/s13059-016-0990-4

Costenbader, E., and Valente, T. W. (2003). The stability of centrality measures when networks are sampled. *Soc. Networks* 25 (4), 283–307. doi: 10.1016/S0378-8733(03)00012-1

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., et al. (2011). Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* 39 (Database issue), D691–D697. doi: 10.1093/nar/gkq1018

Dai, H., Gallagher, D., Schmitt, S., Pessetto, Z. Y., Fan, F., Godwin, A. K., et al. (2017). Role of miR-139 as a surrogate marker for tumor aggression in breast cancer. *Hum. Pathol.* 61, 68–77. doi: 10.1016/j.humpath.2016.11.001

Dang, T. T., Westcott, J. M., Maine, E. A., Kanchwala, M., Xing, C., and Pearson, G. W. (2016). DeltaNp63alpha induces the expression of FAT2 and Slug to promote tumor invasion. *Oncotarget* 7 (19), 28592–28611. doi: 10.18632/oncotarget8696

Eulalio, A., Huntzinger, E., and Izaurralde, E. (2008). Getting to the root of miRNA-mediated gene silencing. *Cell* 132 (1), 9–14. doi: 10.1016/j.cell.2007.12.024

Fabian, M. R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.* 79, 351–379. doi: 10.1146/annurev-biochem-060308-103103

Fontanil, T., Alvarez-Teijeiro, S., Villaronga, M. A., Mohamedi, Y., Solares, L., Moncada-Pazos, A., et al. (2017). Cleavage of Fibulin-2 by the aggrecanases ADAMTS-4 and ADAMTS-5 contributes to the tumorigenic potential of breast cancer cells. *Oncotarget* 8 (8), 13716–13729. doi: 10.18632/oncotarget.14627

Fredericks, A. M., Cygan, K. J., Brown, B. A., and Fairbrother, W. G. (2015). RNA-binding proteins: splicing factors and disease. *Biomolecules* 5 (2), 893–909. doi: 10.3390/biom5020893

Frisone, P., Pradella, D., Di Matteo, A., Belloni, E., Ghigna, C., and Paronetto, M. P. (2015). SAM68: signal transduction and RNA metabolism in human cancer. *BioMed. Res. Int.* 2015, 528954. doi: 10.1155/2015/528954

Fu, X., Mao, X., Wang, Y., Ding, X., and Li, Y. (2017). Let-7c-5p inhibits cell proliferation and induces cell apoptosis by targeting ERCC6 in breast cancer. *Oncol. Rep.* 38 (3), 1851–1856. doi: 10.3892/or.20175839

Galante, P. A. F., Sandhu, D., Abreu, R. D., Gradassi, M., Slager, N., Vogel, C., et al. (2009). A comprehensive in silico expression analysis of RNA binding proteins in normal and tumor tissue Identification of potential players in tumor formation. *RNA Biol.* 6 (4), 426–433. doi: 10.4161/rna.6.4.8841

Gao, C., Li, H., Zhuang, J., Zhang, H., Wang, K., Yang, J., et al. (2019). The construction and analysis of ceRNA networks in invasive breast cancer: a study based on the cancer genome atlas. *Cancer Manag. Res.* 11, 1–11. doi: 10.2147/CMAR.S182521

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. *Nat. Rev. Genet.* 15 (12), 829–845. doi: 10.1038/nrg3813

Goto-Yamaguchi, L., Yamamoto-Ibusuki, M., Yamamoto, Y., Fujiki, Y., Tomiguchi, M., Sueta, A., et al. (2018). Therapeutic predictors of neoadjuvant endocrine therapy response in estrogen receptor-positive breast cancer with reference to optimal gene expression profiling. *Breast Cancer Res. Treat* 172 (2), 353–362. doi: 10.1007/s10549-018-4933-5

Grammatikakis, I., Abdelmohsen, K., and Gorospe, M. (2017). Posttranslational control of HuR function. *Wiley Interdiscip. Reviews-RNA* 8 (1), 180239. doi: 10.1002/wrna.1372

Gu, J., Chen, J., Feng, J., Liu, Y. F., Xue, Q., Mao, G. X., et al. (2016). Overexpression of ADAMTS5 can regulate the migration and invasion of non-small cell lung cancer. *Tumor Biol.* 37 (7), 8681–8689. doi: 10.1007/s13277-015-4573-x

Han, L., Liu, B., Jiang, L., Liu, J., and Han, S. (2016). MicroRNA-497 downregulation contributes to cell proliferation, migration, and invasion of estrogen receptor alpha negative breast cancer by targeting estrogen-related receptor alpha. *Tumour Biol.* 37 (10), 13205–13214. doi: 10.1007/s13277-016-5200-1

Hentze, M. W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* 19 (5), 327–341. doi: 10.1038/nrm.2017.130

Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37 (1), 1–13. doi: 10.1093/nar/gkn923

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4 (1), 44–57. doi: 10.1038/nprot.2008.211

Huang, J., Sun, Y., Chen, H., Liao, Y., Li, S., Chen, C., et al. (2019). ADAMTS5 acts as a tumor suppressor by inhibiting migration, invasion and angiogenesis in human gastric cancer. *Gastric Cancer* 22 (2), 287–301. doi: 10.1007/s10120-018-0866-2

Iorio, M. V., Ferracin, M., Liu, C. G., Veronese, A., Spizzo, R., Sabbioni, S., et al. (2005). MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.* 65 (16), 7065–7070. doi: 10.1158/0008-5472.Can-05-1783

Jiang, C. J., Li, Y. S., Zhao, Z., Lu, J. P., Chen, H., Ding, N., et al. (2016a). Identifying and functionally characterizing tissue-specific and ubiquitously expressed human lncRNAs. *Oncotarget* 7 (6), 7120–7133. doi: 10.18632/oncotarget.6859

Jiang, Q., He, M., Guan, S., Ma, M., Wu, H., Yu, Z., et al. (2016b). MicroRNA-100 suppresses the migration and invasion of breast cancer cells by targeting FZD-8 and inhibiting Wnt/beta-catenin signaling pathway. *Tumour Biol.* 37 (4), 5001–5011. doi: 10.1007/s13277-015-4342-x

Jiang, C., Ding, N., Li, J., Jin, X., Li, L., Pan, T., et al. (2018). Landscape of the long non-coding RNA transcriptome in human heart. *Brief Bioinform.* 20 (5), 1812–1825. doi: 10.1093/bib/bby052

Kim, R. K., Suh, Y., Yoo, K. C., Cui, Y. H., Kim, H., Kim, M. J., et al. (2015). Activation of KRAS promotes the mesenchymal features of basal-type breast cancer. *Exp. Mol. Med.* 47, e137. doi: 10.1038/emm.2014.99

Konig, J., Zarnack, K., Luscombe, N. M., and Ule, J. (2012). Protein-RNA interactions: new genomic technologies and perspectives. *Nat. Rev. Genet.* 13 (2), 77–83. doi: 10.1038/nrg3141

Latora, V., and Marchiori, M. (2001). Efficient behavior of small-world networks. *Phys. Rev. Lett.* 87 (19), 198701. doi: 10.1103/PhysRevLett.87.198701

Lee, J. D., Hempel, N., Lee, N. Y., and Blobe, G. C. (2010). The type III TGF-beta receptor suppresses breast cancer progression through GIPC-mediated inhibition of TGF-beta signaling. *Carcinogenesis* 31 (2), 175–183. doi: 10.1093/carcin/bgp271

Li, C. H., and Chen, Y. C. (2013). Targeting long non-coding RNAs in cancers: progress and prospects. *Int. J. Biochem. Cell Biol.* 45 (8), 1895–1910. doi: 10.1016/j.biocel.2013.05.030

Li, J. H., Liu, S., Zhou, H., Qu, L. H., and Yang, J. H. (2014a). starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 42 (Database issue), D92–D97. doi: 10.1093/nar/gkt1248

Li, Q., Yao, Y., Eades, G., Liu, Z., Zhang, Y., and Zhou, Q. (2014b). Downregulation of miR-140 promotes cancer stem cell formation in basal-like early stage breast cancer. *Oncogene* 33 (20), 2589–2600. doi: 10.1038/onc.2013.226

Li, J., Han, L., Roebuck, P., Diao, L., Liu, L., Yuan, Y., et al. (2015). TANRIC: an interactive open platform to explore the function of lncRNAs in cancer. *Cancer Res.* 75 (18), 3728–3737. doi: 10.1158/0008-5472.CAN-15-0273

Li, M., Li, D., Tang, Y., Wu, F., and Wang, J. (2017). CytoCluster: a cytoscape plugin for cluster analysis and visualization of biological networks. *Int. J. Mol. Sci.* 18 (9), 1880. doi: 10.3390/ijms18091880

Li, X., Yu, X., He, Y., Meng, Y., Liang, J., Huang, L., et al. (2018a). Integrated analysis of MicroRNA (miRNA) and mRNA profiles reveals reduced correlation between microRNA and target gene in cancer. *BioMed. Res. Int.* 2018, 1972606. doi: 10.1155/2018/1972606

Li, Y., Li, L., Wang, Z., Pan, T., Sahni, N., Jin, X., et al. (2018b). LncMAP: pan-cancer atlas of long noncoding RNA-mediated transcriptional network perturbations. *Nucleic Acids Res.* 46 (3), 1113–1123. doi: 10.1093/nar/gkx1311

Li, Y. X., Yu, Z. W., Jiang, T., Shao, L. W., Liu, Y., Li, N., et al. (2018c). SNCA, a novel biomarker for Group 4 medulloblastomas, can inhibit tumor invasion and induce apoptosis. *Cancer Sci.* 109 (4), 1263–1275. doi: 10.1093/nar/gkx1311

Liu, S. L., Patel, S. H., Ginestier, C., Ibarra, I., Martin-Trevino, R., Bai, S. M., et al. (2012). MicroRNA93 regulates proliferation and differentiation of normal and malignant breast stem cells. *PloS Genet.* 8 (6), e1002751. doi: 10.1371/journal.pgen.1002751

Liu, H. R., Li, J., Koirala, P., Ding, X. F., Chen, B. H., Wang, Y. H., et al. (2016). Long non-coding RNAs as prognostic markers in human breast cancer. *Oncotarget* 7 (15), 20584–20596. doi: 10.18632/oncotarget.7828

Liu, A. N., Qu, H. J., Gong, W. J., Xiang, J. Y., Yang, M. M., and Zhang, W. (2019). LncRNA AWPPH and miRNA-21 regulates cancer cell proliferation and chemosensitivity in triple-negative breast cancer by interacting with each other. *J. Cell Biochem.* 120 (9), 14860–14866. doi: 10.1002/jcb.28747

Lo, P. H., Tanikawa, C., Katagiri, T., Nakamura, Y., and Matsuda, K. (2015). Identification of novel epigenetically inactivated gene PAMR1 in breast carcinoma. *Oncol. Rep.* 33 (1), 267–273. doi: 10.3892/or.20143581

Luo, M., Li, Z. W., Wang, W., Zeng, Y. G., Liu, Z. H., and Qiu, J. X. (2013). Long non-coding RNA H19 increases bladder cancer metastasis by associating with EZH2 and inhibiting E-cadherin expression. *Cancer Lett.* 333 (2), 213–221. doi: 10.1016/j.juro.2013.08.057

Mendoza-Rodriguez, M. G., Ayala-Sumuano, J. T., Garcia-Morales, L., Zamudio-Meza, H., Perez-Yepez, E. A., and Meza, I. (2019). IL-1 beta inflammatory cytokine-induced TP63 isoform delta NP63 alpha signaling cascade contributes to cisplatin resistance in human breast cancer cells. *Int. J. Mol. Sci.* 20 (2), 270. doi: 10.3390/ijms20020270

Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F., and Mattick, J. S. (2008). Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl. Acad. Sci. U.S.A.* 105 (2), 716–721. doi: 10.1073/pnas.0706729105

Mertins, P., Mani, D. R., Ruggles, K. V., Gillette, M. A., Clauser, K. R., Wang, P., et al. (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* 534 (7605), 55–62. doi: 10.1038/nature18003

Mi, H., and Thomas, P. (2009). PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods Mol. Biol.* 563, 123–140. doi: 10.1007/978-1-60761-175-2_7

Milanovic, R., Roje, Z., Korusic, A., Deno, I. T., Baric, A., and Stanec, Z. (2013). Clinical and patohistological factors affecting the 5 year survival rate in a population of croatian women with invasive ductal breast carcinoma. *Collegium Antropologicum* 37 (2), 459–464.

Niknafs, Y. S., Han, S., Ma, T., Speers, C., Zhang, C., Wilder-Romans, K., et al. (2016). The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression. *Nat. Commun.* 7, 12791. doi: 10.1038/ncomms12791

Nishida, K., Kuwano, Y., Nishikawa, T., Masuda, K., and Rokutan, K. (2017). RNA binding proteins and genome integrity. *Int. J. Mol. Sci.* 18 (7), 1341. doi: 10.3390/ijms18071341

Pan, D. Z., Garske, K. M., Alvarez, M., Bhagat, Y. V., Boocock, J., Nikkola, E., et al. (2018). Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS. *Nat. Commun.* 9, 1512. doi: 10.1038/s41467-018-03554-9

Papageorgis, P., Ozturk, S., Lambert, A. W., Neophytou, C. M., Tzatsos, A., Wong, C. K., et al. (2015). Targeting IL13Ralpha2 activates STAT6-TP63 pathway to suppress breast cancer lung metastasis. *Breast Cancer Res.* 17, 98. doi: 10.1186/s13058-015-0607-y

Plaisier, C. L., Pan, M., and Baliga, N. S. (2012). A miRNA-regulatory network explains how dysregulated miRNAs perturb oncogenic processes across diverse cancers. *Genome Res.* 22 (11), 2302–2314. doi: 10.1101/gr.133991.111

Ponting, C. P., Oliver, P. L., and Reik, W. (2009). Evolution and functions of long noncoding RNAs. *Cell* 136 (4), 629–641. doi: 10.1016/j.cell.2009.02.006

Rana, T. M. (2007). Illuminating the silence: understanding the structure and function of small RNAs. *Nat. Rev. Mol. Cell Biol.* 8 (1), 23–36. doi: 10.1038/nrm2085

Roscigno, G., Quintavalle, C., Donnarumma, E., Puoti, I., Diaz-Lagares, A., Iaboni, M., et al. (2016). MiR-221 promotes stemness of breast cancer cells by targeting DNMT3b. *Oncotarget* 7 (1), 580–592. doi: 10.18632/oncotarget.5979

Sahay, D., Leblanc, R., Grunewald, T. G. P., Ambatipudi, S., Ribeiro, J., Clezardin, P., et al. (2015). The LPA1/ZEB1/miR-21-activation pathway regulates metastasis in basal breast cancer. *Oncotarget* 6 (24), 20604–20620. doi: 10.18632/oncotarget.3774

Salem, O., Erdem, N., Jung, J., Munstermann, E., Worner, A., Wilhelm, H., et al. (2016). The highly expressed 5'isomiR of hsa-miR-140-3p contributes to the tumor-suppressive effects of miR-140 by reducing breast cancer proliferation and migration. *BMC Genomics* 17, 566. doi: 10.1186/s12864-016-2869-x

Schmitt, A. M., and Chang, H. Y. (2016). Long noncoding RNAs in cancer pathways. *Cancer Cell* 29 (4), 452–463. doi: 10.1016/j.ccell.2016.03.010

Serra-Musach, J., Aguilar, H., Iorio, F., Comellas, F., Berenguer, A., Brunet, J., et al. (2012). Cancer develops, progresses and responds to therapies through restricted perturbation of the protein-protein interaction network. *Integr. Biol. (Camb)* 4 (9), 1038–1048. doi: 10.1039/c2ib20052j

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi: 10.1101/gr.1239303

Siegel, R. L., Miller, K. D., and Jemal, A. (2017). Cancer Statistics, 2017. *Ca-a Cancer J. Clin.* 67 (1), 7–30. doi: 10.3322/caac.21387

Singh, R., Yadav, V., Kumar, S., and Saini, N. (2015). MicroRNA-195 inhibits proliferation, invasion and metastasis in breast cancer cells by targeting FASN, HMGCR, ACACA and CYP27B1. *Sci. Rep.* 5, 17454. doi: 10.1038/srep17454

Slattery, M. L., John, E. M., Stern, M. C., Herrick, J., Lundgreen, A., Giuliano, A. R., et al. (2013). Associations with growth factor genes (FGF1, FGF2, PDGFB, FGFR2, NRG2, EGF, ERBB2) with breast cancer risk and survival: the breast cancer health disparities study. *Breast Cancer Res. Treat* 140 (3), 587–601. doi: 10.1007/s10549-013-2644-5

Song, L. L., Chang, R. X., Dai, C., Wu, Y. J., Guo, J. Y., Qi, M. Y., et al. (2017). SORBS1 suppresses tumor metastasis and improves the sensitivity of cancer to chemotherapy drug. *Oncotarget* 8 (6), 9108–9122. doi: 10.18632/oncotarget.12851

Stewart, B. W., and Wild, C. (2014). International agency for research on cancer, and world health organization *World cancer report 2014.*

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102 (43), 15545–15550. doi: 10.1073/pnas.0506580102

Sun, M., Wu, D., Zhou, K., Li, H., Gong, X., Wei, Q., et al. (2019). An eight-lncRNA signature predicts survival of breast cancer patients: a comprehensive study based on weighted gene co-expression network analysis and competing endogenous RNA network. *Breast Cancer Res. Treat* 175 (1), 59–75. doi: 10.1007/s10549-019-05147-6

Tang, X., Tu, G., Yang, G. L., Wang, X., Kang, L. M., Yang, L. P., et al. (2019). Autocrine TGF-beta 1/miR-200s/miR-221/DNMT3B regulatory loop maintains CAF status to fuel breast cancer cell proliferation. *Cancer Lett.* 452, 79–89. doi: 10.1016/j.canlet.2019.02.044

Tang, G. L. (2005). siRNA and miRNA: an insight into RISCs. *Trends Biochem. Sci.* 30 (2), 106–114. doi: 10.1016/j.tibs.2004.12.007

Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Proteomics. tissue-based map of the human proteome. *Science* 347 (6220), 1260419. doi: 10.1126/science.1260419

Wagner, A., and Fell, D. A. (2001). The small world inside large metabolic networks. *Proc. R. Soc. B-Biolog. Sci.* 268 (1478), 1803–1810. doi: 10.1098/rspb.20011711

Wang, J., Liu, Q., and Shyr, Y. (2015). Dysregulated transcription across diverse cancer types reveals the importance of RNA-binding protein in carcinogenesis. *BMC Genomics* 16, S5. doi: 10.1186/1471-2164-16-S7-S5

Wang, L., Jiang, C. F., Li, D. M., Ge, X., Shi, Z. M., Li, C. Y., et al. (2016a). MicroRNA-497 inhibits tumor growth and increases chemosensitivity to 5-fluorouracil treatment by targeting KSR1. *Oncotarget* 7 (3), 2660–2671. doi: 10.18632/oncotarget6545

Wang, N., Chen, P., Huang, L. P., and Wang, T. Z. (2016b). Prognostic significance of microRNA-10b overexpression in breast cancer: a meta-analysis. *Genet. Mol. Res.* 15 (2), gmr7350. doi: 10.4238/gmr.15027350

Wang, Y., Zhang, X., Zou, C., Kung, H. F., Lin, M. C., Dress, A., et al. (2016c). miR-195 inhibits tumor growth and angiogenesis through modulating IRS1 in breast cancer. *BioMed. Pharmacother.* 80, 95–101. doi: 10.1016/j.biopha.2016.03.007

Wang, C. Q., Chen, L., Dong, C. L., Song, Y., Shen, Z. P., Shen, W. M., et al. (2017). MiR-377 suppresses cell proliferation and metastasis in gastric cancer *via* repressing the expression of VEGFA. *Eur. Rev. Med. Pharmacol. Sci.* 21 (22), 5101–5111. doi: 10.26355/eurrev_201711_13826

Wang, D. Y., Gendoo, D. M. A., Ben-David, Y., Woodgett, J. R., and Zacksenhaus, E. (2019a). A subgroup of microRNAs defines PTEN-deficient, triple-negative breast cancer patients with poorest prognosis and alterations in RB1, MYC, and Wnt signaling. *Breast Cancer Res.* 21 (1), 18. doi: 10.1186/s13058-019-1098-z

Wang, X., Chen, T., Zhang, Y., Zhang, N., Li, C., Li, Y., et al. (2019b). Long noncoding RNA Linc00339 promotes triple-negative breast cancer progression through miR-377-3p/HOXC6 signaling pathway. *J. Cell Physiol.* 234 (8), 13303–13317. doi: 10.1002/jcp.28007

Wei, Y. T., Guo, D. W., Hou, X. Z., and Jiang, D. Q. (2017). miRNA-223 suppresses FOXO1 and functions as a potential tumor marker in breast cancer. *Cell Mol. Biol. (Noisy-le-grand)* 63 (5), 113–118. doi: 10.14715/cmb/2017.63.5.21

Weigelt, B., Horlings, H. M., Kreike, B., Hayes, M. M., Hauptmann, M., Wessels, L. F. A., et al. (2008). Refinement of breast cancer classification by molecular characterization of histological special types. *J. Pathol.* 216 (2), 141–150. doi: 10.1002/path2407

Wen, J. X., Li, X. Q., and Chang, Y. (2018). Signature gene identification of cancer occurrence and pattern recognition. *J. Comput. Biol.* 25 (8), 907–916. doi: 10.1089/cmb.20170261

Wu, W., Wagner, E. K., Hao, Y., Rao, X., Dai, H., Han, J., et al. (2016a). Tissue-specific co-expression of long non-coding and coding RNAs associated with breast cancer. *Sci. Rep.* 6, 32731. doi: 10.1038/srep32731

Wu, Z., Cai, X., Huang, C., Xu, J., and Liu, A. (2016b). miR-497 suppresses angiogenesis in breast carcinoma by targeting HIF-1alpha. *Oncol. Rep.* 35 (3), 1696–1702. doi: 10.3892/or.20154529

Xiao, B., Zhang, W., Chen, L., Hang, J., Wang, L., Zhang, R., et al. (2018). Analysis of the miRNA-mRNA-lncRNA network in human estrogen receptor-positive and estrogen receptor-negative breast cancer based on TCGA data. *Gene* 658, 28–35. doi: 10.1016/j.gene.2018.03.011

Yan, L. X., Huang, X. F., Shao, Q., Huang, M. Y., Deng, L., Wu, Q. L., et al. (2008). MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. *RNA* 14 (11), 2348–2360. doi: 10.1261/rna.1034808

Yan, X., Hu, Z., Feng, Y., Hu, X., Yuan, J., Zhao, S. D., et al. (2015). Comprehensive genomic characterization of long non-coding RNAs across human cancers. *Cancer Cell* 28 (4), 529–540. doi: 10.1016/j.ccell.2015.09.006

Yan, Y. L., Xu, Z. J., Hu, X. F., Qian, L., Li, Z., Zhou, Y. Y., et al. (2018). SNCA is a functionally low-expressed gene in lung adenocarcinoma. *Genes* 9 (1), 16. 10.3390/genes9010016

Yang, G. D., Lu, X. Z., and Yuan, L. J. (2014). LncRNA: a link between RNA and cancer. *Biochim. Et Biophys. Acta-Gene Regul. Mech.* 1839 (11), 1097–1109. doi: 10.1016/j.bbagrm.2014.08.012

Yang, R., Xing, L., Wang, M., Chi, H., Zhang, L., and Chen, J. (2018a). Comprehensive analysis of differentially expressed profiles of lncRNAs/mRNAs and miRNAs with associated ceRNA networks in triple-negative breast cancer. *Cell Physiol. Biochem.* 50 (2), 473–488. doi: 10.1159/000494162

Yang, Y., Yang, H., Xu, M., Zhang, H., Sun, M., Mu, P., et al. (2018b). Long non-coding RNA (lncRNA) MAGI2-AS3 inhibits breast cancer cell growth by targeting the Fas/FasL signalling pathway. *Hum. Cell* 31 (3), 232–241. doi: 10.1007/s13577-018-0206-1

Yanwirasti, W.,. A. H., and Arisanty, D. (2017). Evaluation of MiR-21 and MiR-10b expression of human breast cancer in west sumatera. *Pak J. Biol. Sci.* 20 (4), 189–196. doi: 10.3923/pjbs.2017.189.196

Yi, C. H., Zheng, T., Leaderer, D., Hoffman, A., and Zhu, Y. (2009). Cancer-related transcriptional targets of the circadian gene NPAS2 identified by genome-wide ChIP-on-chip analysis. *Cancer Lett.* 284 (2), 149–156. doi: 10.1016/j.canlet.2009.04.017

Yin, Y., Shen, C., Xie, P., Cheng, Z., and Zhu, Q. (2016). Construction of an initial microRNA regulation network in breast invasive carcinoma by bioinformatics analysis. *Breast* 26, 1–10. doi: 10.1016/j.breast.2015.11.008

Yu, L., Lu, Y., Han, X., Zhao, W., Li, J., Mao, J., et al. (2016). microRNA -140-5p inhibits colorectal cancer invasion and metastasis by targeting ADAMTS5 and IGFBP5. *Stem Cell Res. Ther.* 7 (1), 180. doi: 10.1186/s13287-016-0438-5

Zang, H., Li, N., Pan, Y., and Hao, J. (2017). Identification of upstream transcription factors (TFs) for expression signature genes in breast cancer. *Gynecol. Endocrinol.* 33 (3), 193–198. doi: 10.1080/09513590.2016.1239253

Zhang, J., Fan, D., Jian, Z., Chen, G. G., and Lai, P. B. (2015). Cancer specific long noncoding RNAs show differential expression patterns and competing endogenous RNA potential in hepatocellular carcinoma. *PloS One* 10 (10), e0141042. doi: 10.1371/journal.pone.0141042

Zhou, M., Wang, X. J., Shi, H. B., Cheng, L., Wang, Z. Z., Zhao, H. Q., et al. (2016a). Characterization of long non-coding RNA-associated ceRNA network to reveal potential prognostic lncRNA biomarkers in human ovarian cancer. *Oncotarget* 7 (11), 12598–12611. doi: 10.18632/oncotarget7181

Zhou, M., Zhong, L., Xu, W., Sun, Y., Zhang, Z., Zhao, H., et al. (2016b). Discovery of potential prognostic long non-coding RNA biomarkers for predicting the risk of tumor recurrence of breast cancer patients. *Sci. Rep.* 6, 31038. doi: 10.1038/srep31038

Zhu, L., Xi, P. W., Li, X. X., Sun, X., Zhou, W. B., Xia, T. S., et al. (2019a). The RNA binding protein RBMS3 inhibits the metastasis of breast cancer by regulating Twist1 expression. *J. Exp. Clin. Cancer Res.* 38 (1), 105. doi: 10.1186/s13046-019-1111-5

Zhu, M., Wang, X., Gu, Y., Wang, F., Li, L., and Qiu, X. (2019b). MEG3 overexpression inhibits the tumorigenesis of breast cancer by downregulating miR-21 through the PI3K/Akt pathway. *Arch. Biochem. Biophys.* 661, 22–30. doi: 10.1016/j.abb.2018.10.021

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 January 2020
doi: 10.3389/fgene.2019.01394

Systematically Dissecting the Function of RNA-Binding Proteins During Glioma Progression

Jianjun Wang[1], *Jianfeng Qi*[1,2] *and Xianzeng Hou*[1]*

[1] Department of Neurosurgery, The First Hospital Affiliated with Shandong First Medical University, Shandong Provincial Qianfoshan Hospital, Jinan, China, [2] College of Medicine, Shandong First Medical University, Taian, China

RNA-binding proteins (RBPs) play important roles in regulating gene expression and dysregulation of RBPs have been observed in various types of cancer. However, the role of RBPs during glioma progression, and particular in Chinese patients, is only starting to be unveiled. Here, we systematically analyzed the somatic mutation, gene expression patterns of 2949 RBPs during glioma progression. Our comprehensive study reveals several of highly mutated genes (such as ATRX, TTN and SETD2) and differentially expressed genes (such as KIF4A, TTK and CEP55). Integration of the expression of RBPs and genes, we constructed a regulatory network in glioma and revealed the functional links between RBPs and cancer-related genes. Moreover, we identified the prognosis spectrum of RBPs during glioma progression. The expression of a number of RBPs, such as SNRPN and IGF2BP3, are significantly associated with overall survival of patients in all grades. Taken together, our analyses provided a valuable RBP resource during glioma progression, and revealed several candidates that potentially contribute to development of therapeutic targets for glioma.

Keywords: glioma progression, RNA-binding protein, mutations, regulatory network, prognosis

OPEN ACCESS

Edited by:
Yongsheng Kevin Li,
Harbin Medical University, China

Reviewed by:
Shengli Li,
University of Texas Health Science Center at Houston, United States
Zhe Su,
The University of Texas at Austin, United States

***Correspondence:**
Xianzeng Hou
houxianzeng@sina.com

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *29 October 2019*
Accepted: *19 December 2019*
Published: *28 January 2020*

Citation:
Wang J, Qi J and Hou X (2020) Systematically Dissecting the Function of RNA-Binding Proteins During Glioma Progression. Front. Genet. 10:1394. doi: 10.3389/fgene.2019.01394

INTRODUCTION

RNA-binding proteins (RBPs) play crucial roles in post-transcriptional events and perturbations in RBP activity have been associated with various types of cancer (Pereira et al., 2017; Hentze et al., 2018). Understanding the function of RBPs in cancer will help identifying potential prognostic and response biomarkers for design of therapeutic targets (Bonnal et al., 2012; Kudinov et al., 2017). Glioma is a common and aggressive type of brain tumor, which was with poor outcome and no effective treatment by far (Ostrom et al., 2014; Reifenberger et al., 2017). Systematical dissection of RBP functions during glioma progression will provide new insights into the underlying mechanisms of glioma.

Comprehensive identification and annotation of human RBPs are the primary step for investigating their functions. With the development of high throughput sequencing, numbers of RBPs have been identified (Gerstberger et al., 2014). Several databases have curated a number of RBPs. For example, RBPDB is a database for collection of experimentally validated RBPs (Cook et al., 2011). ATtRACT also manually curated approximate 370 RBPs (Giudice et al., 2016). Recently, EuRBPDB has been constructed, which is a widely-used resource for RBPs (Liao et al.,

2019). Moreover, number of studies have found a lot of alterations in RBPs during cancer development and progression. Wang et al., revealed the importance of RBPs in carcinogenesis by large-scale transcriptional profiling studies (Wang et al., 2015). The mutational spectrum of RBPs had been analyzed and identified a number of RBPs exhibited significantly mutation in cancer (Neelamraju et al., 2018). Li et al., have also decoded the genome-wide RBP mutational and transcriptomic landscape and yielded valuable insights into the function of RBPs (Li et al., 2019a). All these results suggest that there are prevalent alterations of RBPs in cancer development and progression.

RBPs have also been demonstrated to play important roles in neurodegeneration and glioma progression (Pereira et al.,, 2017). Numbers of RBPs have also been identified in glioma. Correa et al. revealed the splicing regulator SNRPB as an oncogenic candidate in glioblastoma (GBM) through functional genomics analyses (Correa et al., 2016). RNA-binding protein PCBP2 has been identified to modulate glioma growth by regulating FHL3 (Han et al., 2013). Moreover, Musashi1 was found to be a central regulator of adhesion pathways in GBM (Uren et al., 2015). The RBP IMP2 can preserve GBM stem cells by preventing let-7 target gene silencing (Pereira et al.,, 2017). Although a number of regulatory networks have been analyzed during glioma progression, such as microRNA-gene regulatory network (Li et al., 2013), transcriptional regulatory network (Li et al., 2015), competitive endogenous RNAs (ceRNAs) network (Xu et al., 2015), we are still lack of knowledge about RBP regulatory network during glioma progression.

To address these questions, we systematically analyzed the genetic and transcriptomic alterations of RBPs during glioma progression. We identified a number of RBPs with somatic mutations, differentially expressed during glioma progression. In addition, several RBPs associated with patient overall survival were also identified. These RBPs regulated a number of cancer-related genes and played important roles during glioma progression in Chinese patients. All these results provide novel insights into the function of RBPs in glioma.

MATERIALS AND METHODS

Collection of Human RNA-Binding Proteins

All the human RBPs were downloaded from the EuRBPDB database, which is a comprehensive resource for annotation of eukaryotic RBPs. In total, there were 2,949 RBPs and these RBPs were further classified into canonical and non-canonical RBPs.

Genetic Alteration Profiles of Glioma

The genetic alterations for Chinese glioma patients were downloaded from Chinese Glioma Genome Atlas (CGGA). There were 286 patients sequenced by whole-exome sequencing (Hu et al., 2018). We directly downloaded the gene-level mutation datasets. In this table, each row represents one gene and the columns represent the glioma patient. The one values indicated that this gene was mutated in corresponding patient while zeros indicated not mutated. In addition, we also downloaded the clinical information of these patients from CGGA (**Table S1**). The grade, gender, age, overall survival time, censor status and isocitrate dehydrogenase (IDH) mutation status were included. In addition, we downloaded the somatic mutations of low-grade glioma (LGG) and GBM patients from The Cancer Genome Atlas (TCGA) project. The overall survival and disease-free survival time of these patients were also downloaded.

Genome-Wide Gene Expression Profile of Glioma

Genome-wide gene expression of glioma patients were also downloaded from CGGA (September 9, 2019). mRNA-Seq data were used in our analyses, which included 693 patients in total. The reads were aligned by STAR (Dobin et al., 2013) and the expression were evaluated as RSEM (Li and Dewey, 2011). There were 185 patients with both somatic mutations and RNA-Seq data (**Figure S1**). We also downloaded the clinical information for these 693 patients (**Table S2**).

Identification of Top Mutated Genes in Glioma

To identify the genes with high mutation frequency, we separately ranked each gene in grade II, III, and IV glioma. The gene mutation frequency was defined as:

$$F(g) = \frac{n_g}{N}$$

where n_g was the number of patients with gene g mutated and N was the total number of patients in specific grade.

Identified the Genes With Perturbed Expression in Glioma

We used Wilcoxon's rank sum test to evaluate the difference of gene expression between two adjacent grades. For example, the fold changes for comparison between grade II and III were defined as the (average expression of genes in grade III)/(average expression of genes in grade II). The p-values of Wilcoxon's rank sum test were adjusted by Benjamini–Hochberg (BH) method. Genes with fold changes > 2 and adjusted p-values < 0.05 were defined as up-regulated genes and those with fold changes < 0.5 and adjusted p-values < 0.05 were defined as down-regulated genes. The comparisons were performed between grade II vs. III, and III vs. IV.

Construction of RNA-Binding Protein–Gene Regulatory Network in Glioma

RNA-binding proteins are key regulators of gene expression, yet only a small fraction have been functionally characterized (Pereira et al., 2017). It is still difficult to identify the target genes for the majority of RBPs. Increasing studies have demonstrated that the regulators are likely to co-express with their target genes. Thus, we identified the co-expressed genes of RBPs and constructed the RBP–gene regulatory network in glioma. Here, only the cancer genes were considered and the

cancer-related genes were downloaded from COSMIC Cancer Gene Census (Oct 25, 2019) (Sondka et al., 2018). There are 723 genes in total. For each RBP–gene pair, we calculated the Pearson correlation coefficient (PCC) as follows:

$$R_{ij} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma_X}\right)\left(\frac{Y_i - \bar{Y}}{\sigma_Y}\right)$$

where $\bar{X}$nd $\bar{Y}$where the average expression of RBP X and gene Y, σ_X and $\sigma_Y$where the standard error of expression of RBP and gene. There were n patients in the analysis. All the RBP–gene pairs with PCCs > 0.70 and p-values < 0.05 were identified to construct the regulatory network. The network was visualized by Cytoscape (version 3.7.1) (Shannon et al., 2003).

Identifying Clinical-Associated RNA-Binding Proteins in Glioma

To identify the RBPs whose expression was potentially correlated with glioma patient survival, we first divided patients in each grade into two groups based on the median expression of each RBP. The survival difference between two groups were evaluated by log-rank test. The hazard ratio (HR) was also calculated. This procedure was performed by the R package (version 3.6.1) (https://cran.r-project.org/web/packages/survival/index.html). RBPs with HR > 1 and p < 0.05 were defined as risky factors and those with HR < 1 and p < 0.05 were defined as protective factors.

RESULTS

High-Grade Glioma Patients Exhibit Poor Prognosis and Less Isocitrate Dehydrogenase Mutation

High-grade glioma remains incurable despite number of genetic alterations have been revealed (Chen et al., 2016). Here, we analyzed the 286 glioma patients sequenced by whole-exome sequencing. We found that glioma patients in high-grade exhibited poor prognosis (**Figure 1A**, log-rank p < 0.001). Moreover, we explored the 693 patients with mRNA-Seq data. We also found that the patients in high grade were with significantly poorer survival (**Figure 1B**, log-rank p < 0.001). Particularly, the patients in grade IV (glioblastoma, also known GBM) were with the poorest survival (**Figures 1A, B**). These results were consistent with the current knowledge, that GBM is the most aggressive cancer.

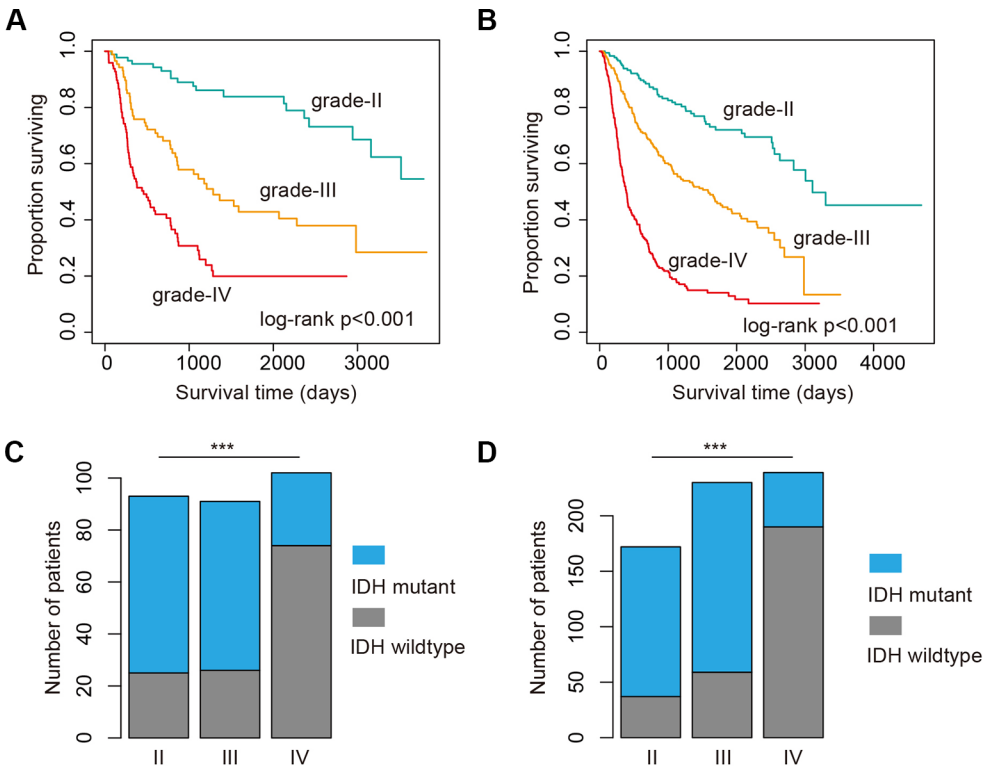


FIGURE 1 | High grade glioma patients with poor survival and less isocitrate dehydrogenase (IDH) mutation. **(A)** Kaplan–Meier plot indicating survival of glioma patients with mutation data in different grades. **(B)** Kaplan–Meier plot indicating survival of glioma patients with expression data in different grades. **(C)** The proportion of patients with IDH mutation or wild type in different grades. Those patients were with mutation data. **(D)** The proportion of patients with IDH mutation or wild type in different grades. Those patients were with expression data. ***P < 0.001.

Next, we investigated whether there are somehow difference in the clinical information for patients in different grades. We first compared the ages of patients with mutation data. The average ages for patients in grade II and III were 38.21 and 39.53 years. The average ages for patients in grade IV were 47.56, which were significantly older than II and III (p-values < 0.01, Wilcoxon's rank sum tests). Moreover, we got the similar results in the patients with mRNA data. However, there were no significant difference between grade II and III. IDH1 is the most commonly mutated gene in glioma (Philip et al., 2018). We thus investigated the mutation frequency of IDH1 in glioma patients. We found that GBM patients were with less IDH mutation, either in the exome sequencing cohort or the mRNA-Seq cohort (**Figures 1C, D**, p-values < 0.001, Fisher's exact test). Moreover, we analyzed the data from TCGA project and found that patients with IDH1 mutation exhibit better overall survival than the wide type ones in LGG and GBM (**Figures 2A, C**). When considering the disease-free survival time, we found that patients with IDH1 mutation also show better survival in LGG and GBM (**Figures 2B, D**). All these results suggest that high-grade glioma patients were older, were not likely with IDH1 mutation and exhibited poor survival.

Prevalent Somatic Mutations of RNA-Binding Protein During Glioma Progression

RBPs have been found to play critical roles in glioma. We thus next investigated the genetic alterations of 2,949 RBPs in glioma (**Figure 3A**). There were 1,826 (61.92%) canonical RBPs with specific RNA binding domains, and 1,123 (38.08%) non-canonical RBPs. Next, we calculated the number of RBPs with different binding domains. We found that there were more than 150 RBPs with RRM_1 domains (**Figure 3B**). We explored whether each patient was with RBP mutation and found that approximate 87.10% patients in grade II, 94.50% patients in grade III and 84.31% patients in grade IV were with RBP mutations (**Figure 3C**). These results suggest that there were prevalent somatic mutations in RBPs during glioma progression.

We next explored which RBPs were with higher mutation frequency in different grades glioma patients. Top ranked 10 mutated genes in each grade were shown in **Figure 3D**. We found that ATRX was ranked top 1 in all grades. The ATRX status has been found to be one of the critical markers that define the molecular classification of gliomas (Nandakumar et al., 2017).

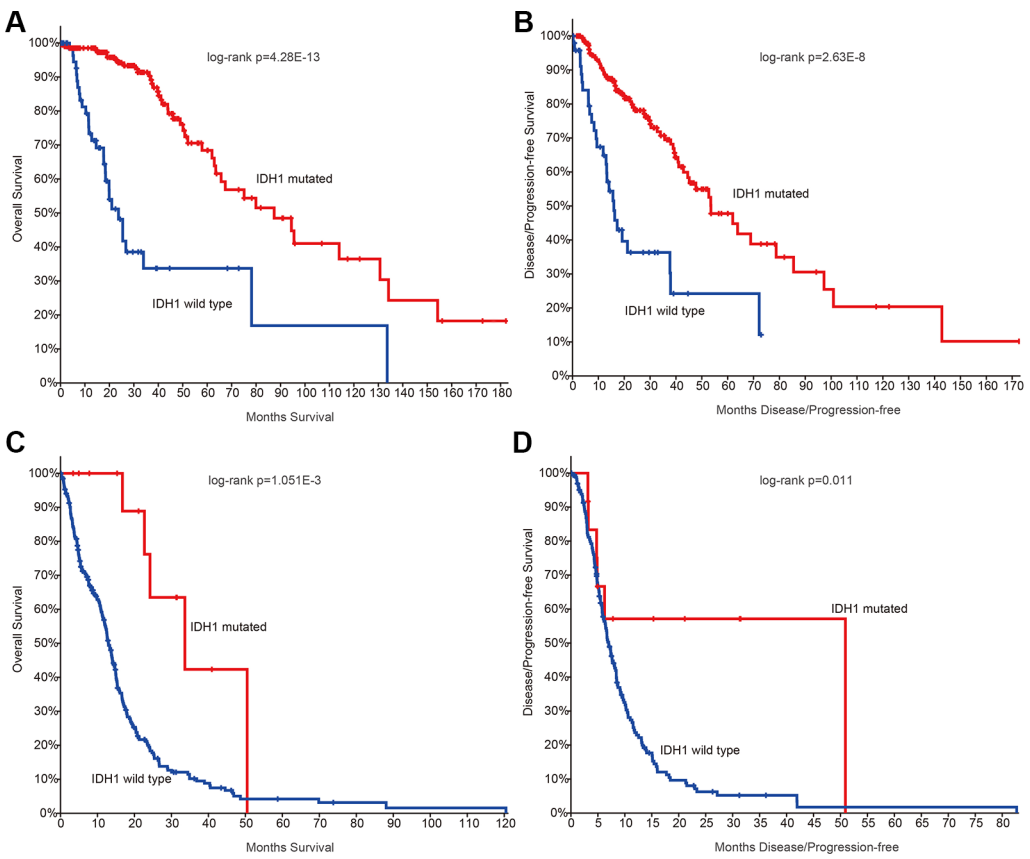


FIGURE 2 | The survival plots for glioma patients in The Cancer Genome Atlas (TCGA) project. **(A)** Kaplan–Meier plot indicating overall survival of low-grade glioma patients with isocitrate dehydrogenase 1 (IDH1) mutation or not. **(B)** Kaplan–Meier plot indicating disease-free survival of low-grade glioma patients with IDH1 mutation or not. **(C)** Kaplan–Meier plot indicating overall survival of glioblastoma (GBM) patients with IDH1 mutation or not. **(D)** Kaplan–Meier plot indicating disease-free survival of GBM patients with IDH1 mutation or not.

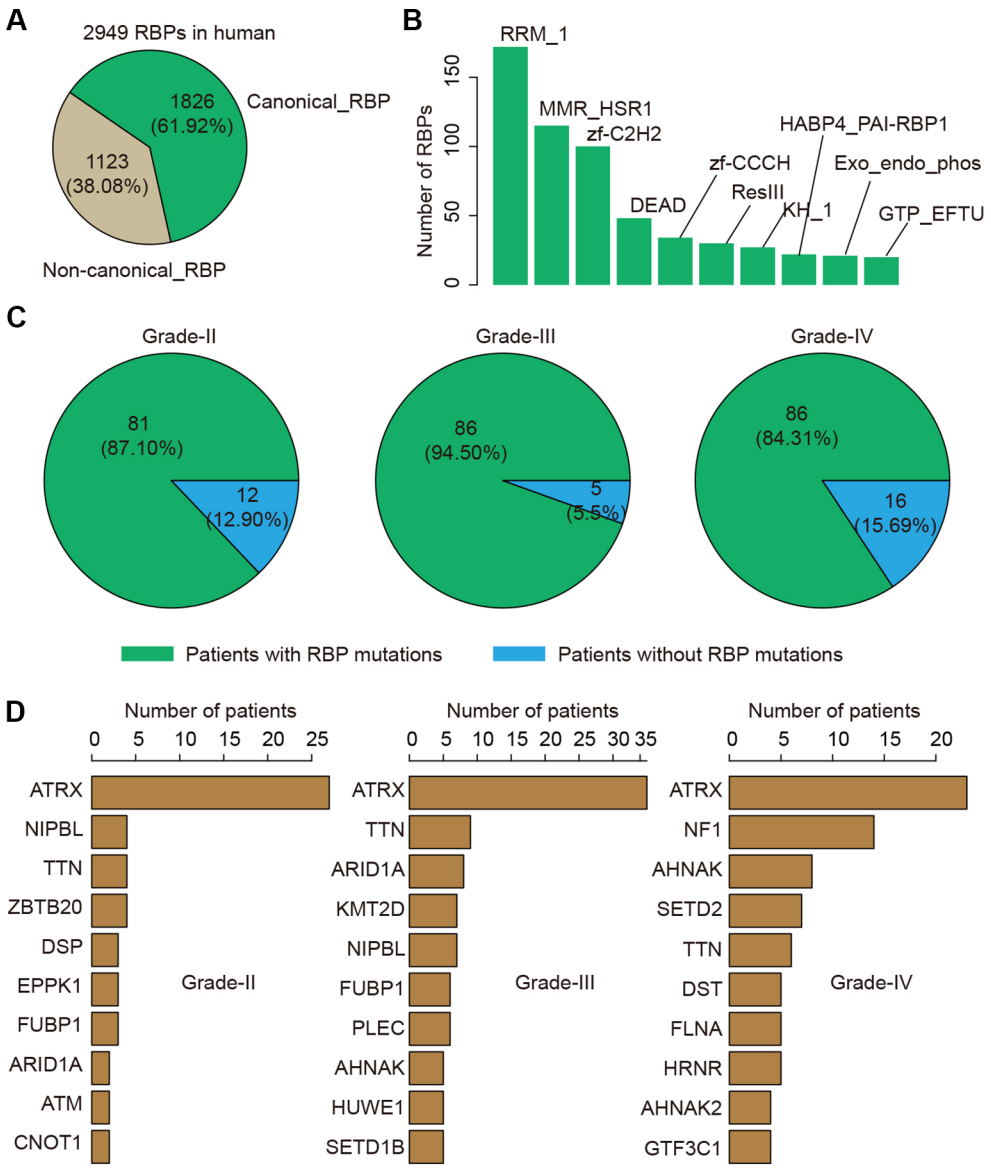


FIGURE 3 | The mutation spectrum of RNA-binding protein (RBP) mutations in glioma. **(A)** The pie chart shows the proportion of canonical and non-canonical RBPs. **(B)** The bar charts shows the number of RBPs in top ranked RBP families. **(C)** The pie charts show the proportion of patients in different grade with RBP mutations. Left for grade II, middle for grade III and right for grade IV. **(D)** Top ranked 10 genes by mutation frequency in different grades. Left for grade II, middle for grade III and right for grade IV.

ATRX loss can promote tumor growth and impair DNA repair in glioma (Koschmann et al., 2016). In total, we found that ATRX1 was mutated in approximate 30% of all glioma patients (**Figure 4**). Another frequently mutated gene was TTN in all grades (**Figure 3D**), which was also identified previously in glioma (Panossian et al., 2018). TTN was mutated in 6% of all glioma patients (**Figure 4**). Moreover, we found that NF1 were with higher mutation frequency in GBM, which has been used to define the mesenchymal subtype of GBM (Verhaak et al., 2010). We also identified several candidate genes, such as ARID1A, SETD2, FLNA and KMT2D. In addition, we queried the PubMed and found that numbers of these genes were co-occurred with "glioma" or "glioblastoma" in literature (**Figure S2** and **S3**). These results provided candidate RBPs for further functional investigation in glioma.

Expression Perturbations of RNA-Binding Proteins During Glioma Progression

Besides genetic alterations, evidence have suggested that the expression of RBPs were also perturbed in cancer (Sebestyen et al., 2016; Li et al., 2017). We next systematically analyzed the RBP transcriptome in different grades of glioma. We identified

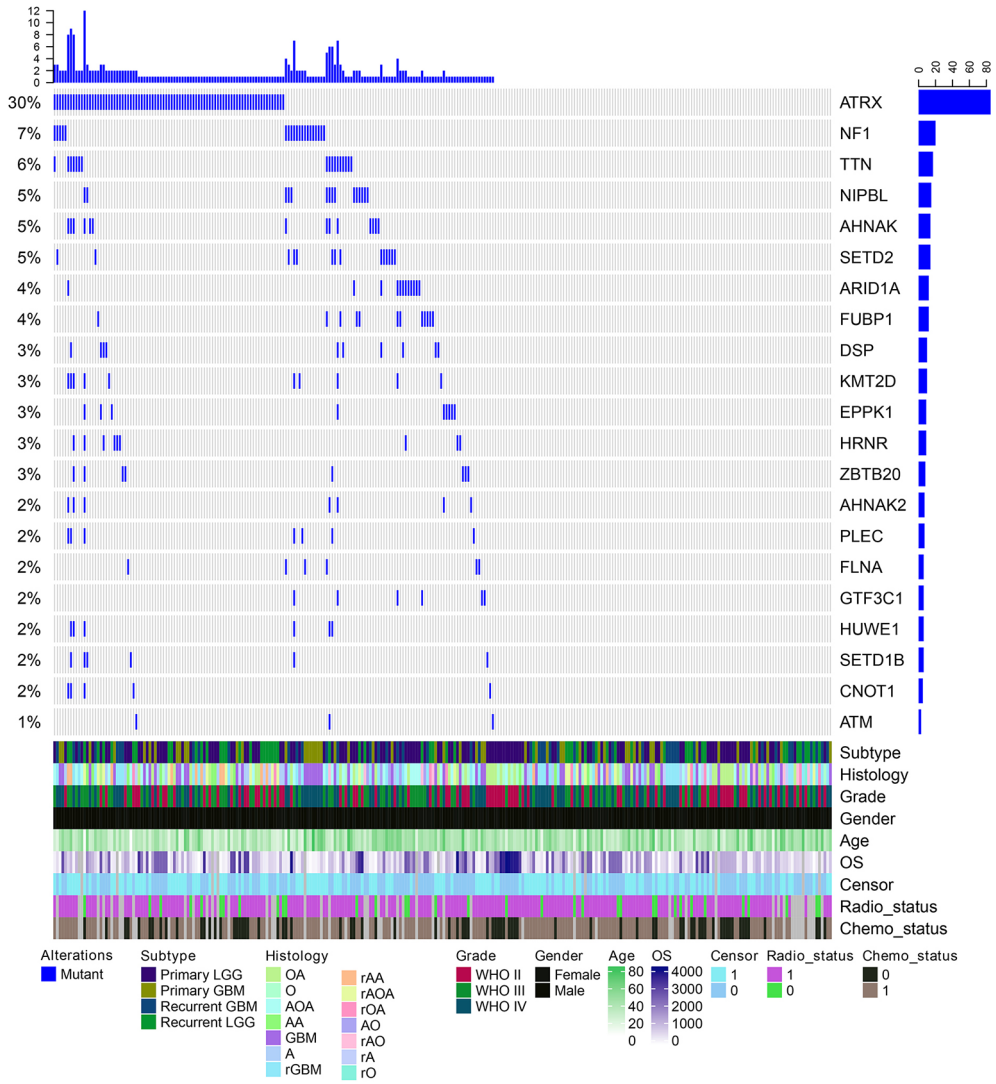


FIGURE 4 | The mutation spectrum of top ranked RNA-binding proteins (RBPs) in glioma. Each column is one glioma patient. The blue lines indicate whether the gene is mutated or not. The clinical information of these patients are shown on the bottom panels.

32 up-regulated RBPs in comparison between grade II vs. III (**Figure 5A**). There are more RBPs exhibited expression perturbations when comparison between grade III and IV (**Figure 5B** and **Table S3**). These results suggested that the transcriptome were likely to be perturbed during the progression from low grade to high grade. Among the top up-regulated genes in comparison between grade II and III, we identified four important genes, such as IGF2BP2, TTK, KIF4A and CEP55 (**Figure 5C**). It has been shown that IGF2BP2 was a direct target of miR−188 in glioma, and IGF2BP2 under−expression served tumor−suppressive roles in glioma growth and metastasis (Ding et al., 2017). CEP55 has been found to promote cell proliferation and inhibits cell apoptosis in glioma (Li et al., 2018b).

When we compared the transcriptome of patients in grade III vs. grade IV, we identified 56 up-regulated genes and 12 down-regulated genes (**Figure 5B**). Among the up-regulated genes, we identified NNMT, LGALS3, PDLIM4, TUBA1C and ANXA2 as top five (**Figure 5D**). NNMT silencing had been shown to activate tumor suppressor PP2A and inhibits tumor forming (Palanichamy et al., 2017). This was consistent with our result that it was up-regulated in GBM. LGALS3 was also found to promote GBM and was associated with tumor risk and prognosis (Wang et al., 2019). PDLIM4 had been identified as a gene signature associated with the clinical outcome in high-grade gliomas (de Tayrac et al., 2011). For the down-regulated genes, we have identified CSDC2, ZNF804B, LIN28A and SRRM3 as candidates. However, few of these were investigated in current studies. These results suggested that the tumor suppressors need to be paid attention in future glioma studies. Taken together, function analysis of these

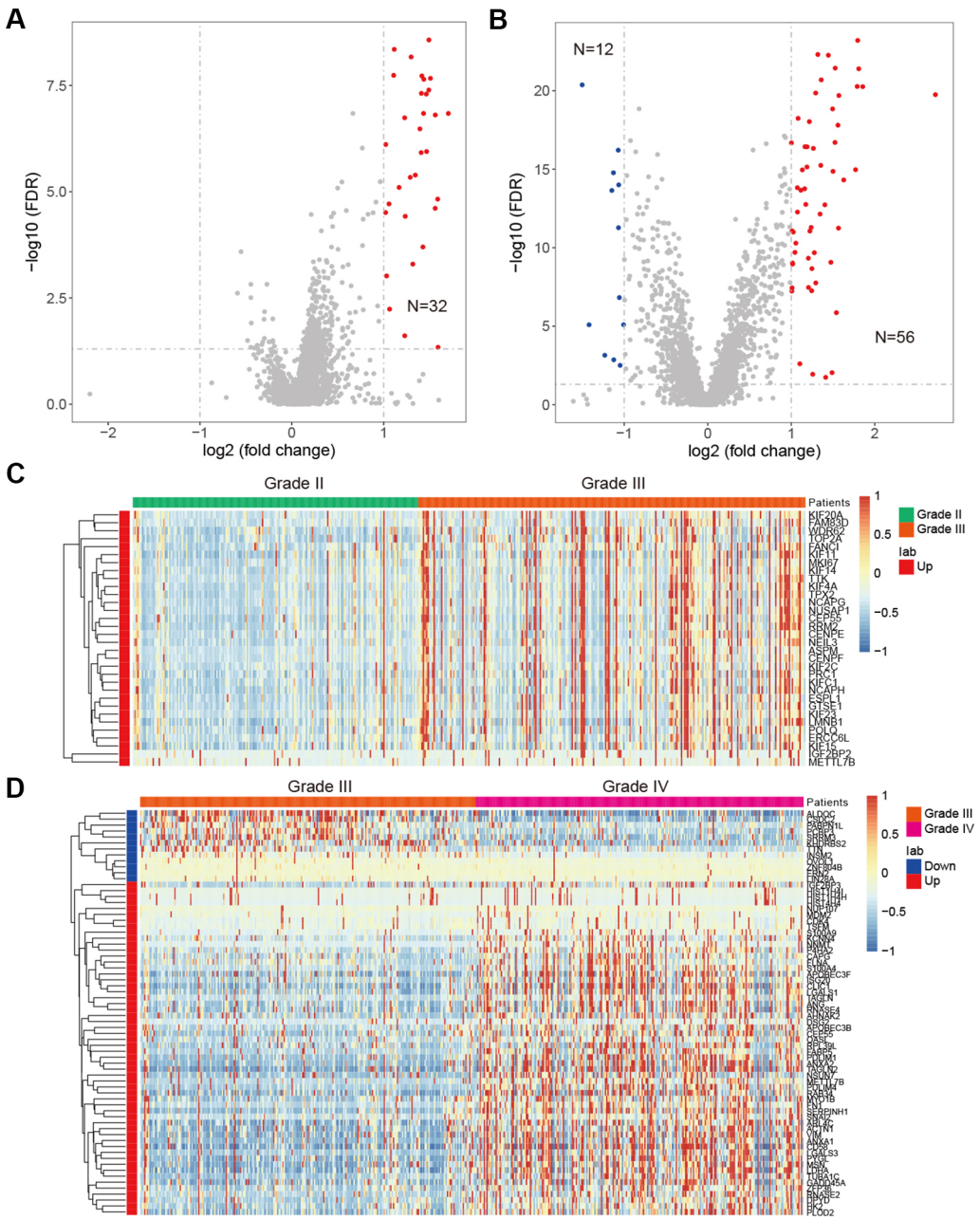


FIGURE 5 | Differentially expressed genes in glioma. **(A)** Volcano plot shows the differentially expressed genes between grade II and III. Red for up-regulated genes. **(B)** Volcano plot shows the differentially expressed genes between grade III and IV. Red for up-regulated genes and blue for down-regulated genes. **(C)** Heat map for genes that perturbed between grade II and III. **(D)** Heat map for genes that perturbed between grade III and IV.

RBPs provide insight into the transcriptome perturbations of glioma progression.

RNA-Binding Protein Regulatory Network During Glioma Progression

Proteins do not function isolatedly but interact with other molecules in complex cellular networks for signal transduction (Xu et al., 2017; Yi et al., 2017). Understanding the RBP regulatory network during glioma progression will get deep insights into their functions. We thus identified the co-expressed genes in glioma. Here, we focused on the cancer-related genes (**Table S4**). At the PCC > 0.70 and *p*-adjusted < 0.05, we identified 368 regulatory interactions among 55 RBPs and 69 genes (**Figure 6A** and **Table S5**). In this regulatory network, several RBPs and genes were correlated with the development and progression of glioma. For example, KIF4A

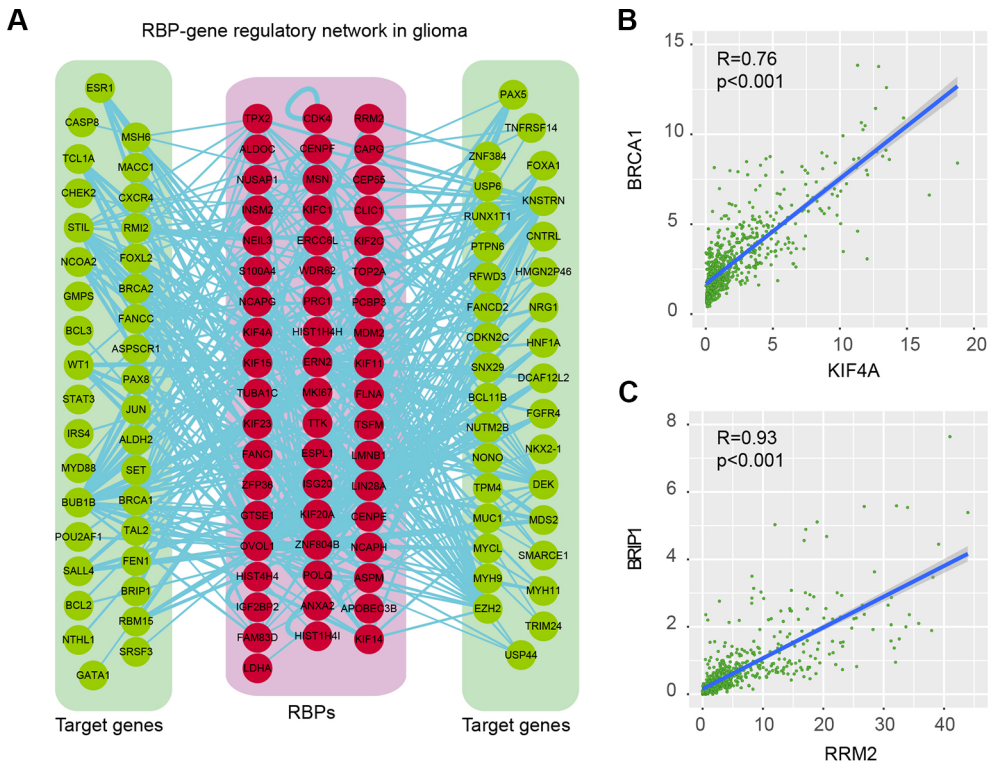


FIGURE 6 | RNA-binding protein (RBP)–gene regulatory network in glioma. **(A)** The RBP–gene regulatory network in glioma. Red for RBPs and green for target genes that were related to cancer. **(B)** Scatter plot shows the correlation between KIF4A and BRCA1 in glioma. **(C)** Scatter plot shows the correlation between RRM2 and BRIP1 in glioma.

was correlated with cell cycle, G2M checkpoint (Cho et al., 2019). Abrogation of BRCA1 had also been found to play roles in tumor growth in glioma (Rasmussen et al., 2016). We found that the expression of KIF4A and BRCA1 was significantly correlated with each other in glioma (**Figure 6B**, $R = 0.76$ and p-value < 0.001), providing a functional link between RBP and gene.

Another example was RRM2 and BRIP1, which was significantly correlated with each other in expression (**Figure 6C**, $R = 0.93$ and p-value < 0.001). RRM2 had been found to promote the progression of human GBM and was a potential prognostic biomarker in glioma (Li et al., 2018a; Sun et al., 2019). BRIP1 was found to be an independent signature, which was correlated with worse prognosis in glioma (de Sousa et al., 2017). Our results provided a way to functionally explain the signaling of RRM2 during glioma progression. Moreover, we also identified the functional association between TTK and BUB1B, KIF23 and POLQ. All these RBP–gene correlations provide suitable ways for functional characterization of RBPs in glioma.

Prognostic Potential of RNA-Binding Protein Regulators

RBPs are essential modulators of transcription and numbers of RBPs have being found to be associated with the survival of patients (Frau et al., 2013). We next identified the RBPs that were associated with the survival of patients in different grades. We found that there were more RBPs were associated with survival in grade III, compared with other two grades (**Figure 7** and **Table S6**). For the protective RBPs in glioma, 25, 425 and 28 RBPs were only associated survival in grade II, III and IV. Three RBPs (ARPP21, SNRPN and GLRX3) were associated with patient overall survival in all grades (**Figure 7A**). SNRPN had been found as a autism-related gene by regulating cortical and spine development via nuclear receptor (Li et al., 2016). Hypermethylation of SNRPN increased as the cellular origin of the tumors advanced in oogenesis and was closely correlated in individual teratomas (Miura et al., 1999). We found that the high expression of SNRPN was correlated with better overall survival in all grade glioma patients (**Figures 7B–D**, log-rank p-values < 0.05).

In addition, we also identified numbers of risky RBPs in glioma (**Figure 7E**). There were 8, 1,220, 12 RBPs were associated with survival in specific grade. In total, 27 RBPs were identified as risky factors in all grades, including BRCA1, MCM2, IGF2BP3, KIF2C, VIM and PLOD3. High MCM2 expression was found to be strongly associated with poor overall survival in patients with high-grade glioma in our current study, as well as previous studies (Hua et al., 2014). Recently, IGF2BP3 has been identified as a potential oncogene across multiple cancer types (Li et al., 2019b). We found that the high expression of IGF2BP3 was significantly associated with poor prognosis in all three grades (**Figure 7F–H**, log-rank p-values < 0.05). These results provided

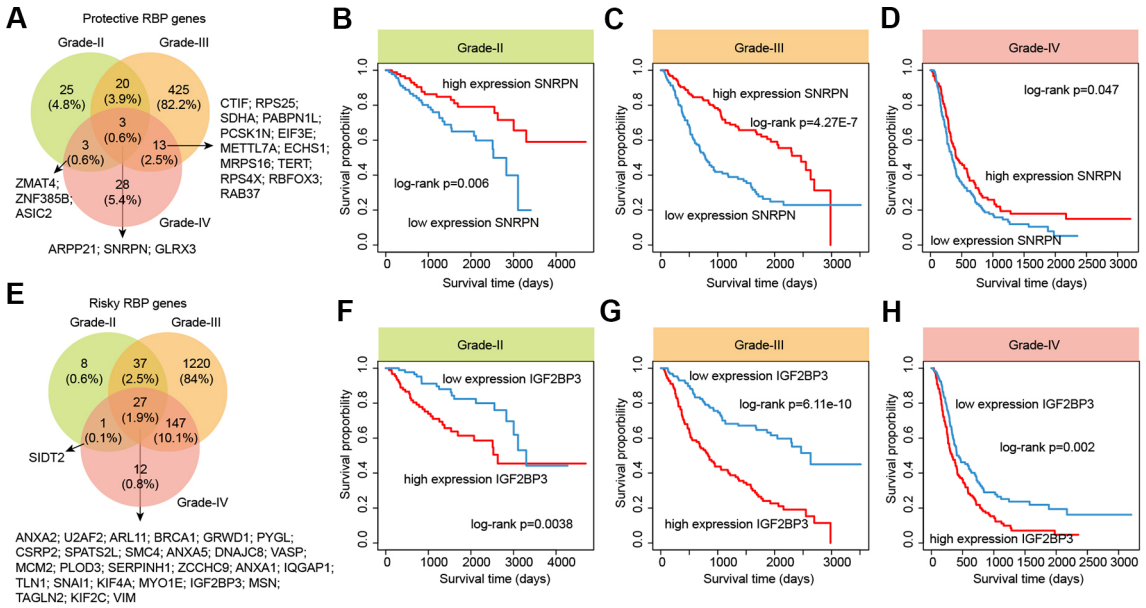


FIGURE 7 | The survival landscape of RNA-binding proteins (RBPs) in glioma. **(A)** Venny plots show the protective RBPs in different grade glioma patients. **(B–D)** Kaplan–Meier plots indicating overall survival of glioma patients with low and high expression of SNRPN. **(E)** Venny plots show the risky RBPs in different grade glioma patients. **(F–H)** Kaplan–Meier plots indicating overall survival of glioma patients with low and high expression of IGF2BP3.

more evidence for the oncogene roles of IGF2BP3 in cancer. Taken together, our analyses provided a prognostic spectrum for RBPs during glioma progression.

DISCUSSION

In this study, we systematically analyzed the genetic and transcriptome alterations of RBPs during glioma progression. The top mutated RBPs in different grades of glioma patients were identified and several of them had been found to play important roles in glioma or other cancers. Moreover, we compared the transcriptome and identified the differentially expressed RBPs. We found that there were more RBPs exhibited expression perturbations during the transition from grade III to IV. These results suggested that the transcriptome was greatly perturbed in the progression of high-grade glioma. Our regulatory network and prognosis analysis also revealed several important candidates for functional characterization in future experiments.

Although several candidate RBPs were identified in our current study, there are a lot of work need to do for investigating the detail functional ways of these RBPs. RBPs have been found to regulate alterative splicing (AS) and influences the expression of genes (Fei et al., 2017). Alterations of AS are emerging as important signatures in cancer (Liu et al., 2017). The mutation of RBPs generally impair the recognition of regulatory sites, and affecting the splicing of multiple genes. However, it is still difficult to determine the targets for the majority of RBPs. The best method for identifying the targets of RBPs is CLIP-Seq, but there are limited number of data currently. With the development of sequencing technology, such as CLIP-Seq, we will get more details about the function of RBPs. Moreover, RBPs can also interact with noncoding RNAs. Identifying the cell type specific RBP interactome will yield novel insight into the function of RBPs.

Moreover, we identified an important RBP IGF2BP3 during glioma progression. Beside RBP, this genes also an epigenetic regulator, which can affect the fates of mRNA in an m6A-dependent manner (Bi et al., 2019). These results suggest that m6A might also play important roles during glioma progression. Several studies have emerged to reveal the function of m6A in glioma (Dixit et al., 2017; Zhang et al., 2017). But we are still lack of knowledge about the landscape of m6A alterations in glioma, particular in Chinese cohort. Moreover, we are not sure to what extent RBPs can regulate m6A. Is it just one case or general regulation? With the increasing data of RBP regulation as well as other epigenetic data, we will get deep insight into this regulatory layer in cancer.

In summary, our comprehensive analyses dissect the potential function of RBPs during glioma progression. Understanding the functions of candidate RBPs identified in this study will provide insight into the underlying mechanisms of glioma progression.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://www.cgga.org.cn/download.jsp.

AUTHOR CONTRIBUTIONS

JW contributed to data analysis and paper writing, JQ contributed to figure construction, XH contributed to the design, data analysis and paper writing.

FUNDING

This paper was supported by grant 2015BSB14042 from the Natural Science Foundation of Shandong Province, China.

ACKNOWLEDGMENTS

We would like to thank the CGGA project for providing the datasets of glioma.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01394/full#supplementary-material

REFERENCES

Bi, Z., Liu, Y., Zhao, Y., Yao, Y., Wu, R., Liu, Q., et al. (2019). A dynamic reversible RNA N(6) -methyladenosine modification: current status and perspectives. *J. Cell. Physiol.* 234, 7948–7956. doi: 10.1002/jcp.28014

Bonnal, S., Vigevani, L., and Valcarcel, J. (2012). The spliceosome as a target of novel antitumour drugs. *Nat. Rev. Drug Discovery* 11, 847–859. doi: 10.1038/nrd3823

Chen, R., Cohen, A. L., and Colman, H. (2016). Targeted therapeutics in patients with high-grade gliomas: past, present, and future. *Curr. Treat. Options Oncol.* 17 (8), 42. doi: 10.1007/s11864-016-0418-0

Cho, S. Y., Kim, S., Kim, G., Singh, P., and Kim, D. W. (2019). Integrative analysis of KIF4A, 9, 18A, and 23 and their clinical significance in low-grade glioma and glioblastoma. *Sci. Rep.* 9, 4599. doi: 10.1038/s41598-018-37622-3

Cook, K. B., Kazan, H., Zuberi, K., Morris, Q., and Hughes, T. R. (2011). RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.* 39, D301–D308. doi: 10.1093/nar/gkq1069

Correa, B. R., de Araujo, P. R., Qiao, M., Burns, S. C., Chen, C., Schlegel, R., et al. (2016). Functional genomics analyses of RNA-Binding Proteins reveal the splicing regulator SNRPB as an oncogenic candidate in glioblastoma. *Genome Biol.* 17, 125. doi: 10.1186/s13059-016-0990-4

de Sousa, J. F., Torrieri, R., Serafim, R. B., Di Cristofaro, L. F., Escanfella, F. D., Ribeiro, R., et al. (2017). Expression signatures of DNA repair genes correlate with survival prognosis of astrocytoma patients. *Tumour Biol. J. Int. Soc. Oncodev. Biol. Med.* 39 (4), 1010428317694552. doi: 10.1177/1010428317694552

de Tayrac, M., Aubry, M., Saikali, S., Etcheverry, A., Surbled, C., Guenot, F., et al. (2011). A 4-gene signature associated with clinical outcome in high-grade gliomas. *Clin. cancer Res. an Off. J. Am. Assoc. Cancer Res.* 17, 317–327. doi: 10.1158/1078-0432.CCR-10-1126

Ding, L., Wang, L., and Guo, F. (2017). microRNA188 acts as a tumour suppressor in glioma by directly targeting the IGF2BP2 gene. *Mol. Med. Rep.* 16, 7124–7130. doi: 10.3892/mmr.2017.7433

Dixit, D., Xie, Q., Rich, J. N., and Zhao, J. C. (2017). Messenger RNA methylation Regulates glioblastoma tumorigenesis. *Cancer Cell.* 31, 474–475. doi: 10.1016/j.ccell.2017.03.010

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Fei, T., Chen, Y., Xiao, T., Li, W., Cato, L., Zhang, P., et al. (2017). Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing. *Proc. Nat. Acad. Sci. U. S. A.* 114, E5207–E5215. doi: 10.1073/pnas.1617467114

Frau, M., Feo, F., and Pascale, R. M. (2013). Pleiotropic effects of methionine adenosyltransferases deregulation as determinants of liver cancer progression and prognosis. *J. Hepatol.* 59, 830–841. doi: 10.1016/j.jhep.2013.04.031

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-Binding Proteins. *Nat. Rev. Genet.* 15, 829–845. doi: 10.1038/nrg3813

Giudice, G., Sanchez-Cabo, F., Torroja, C., and Lara-Pezzi, E. (2016). ATtRACT-a database of RNA-Binding Proteins and associated motifs. *Database. J. Biol. Databases Curation*. doi: 10.1093/ database/baw035

Han, W., Xin, Z., Zhao, Z., Bao, W., Lin, X., Yin, B., et al. (2013). RNA-binding protein PCBP2 modulates glioma growth by regulating FHL3. *J. Clin. Invest.* 123, 2103–2118. doi: 10.1172/JCI61820

Hentze, M. W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-Binding Proteins. *Nat. Rev. Mol. Cell Biol.* 19, 327–341. doi: 10.1038/nrm.2017.130

Hu, H., Mu, Q., Bao, Z., Chen, Y., Liu, Y., Chen, J., et al. (2018). Mutational landscape of secondary glioblastoma guides MET-targeted trial in brain tumor. *Cell* 175, 1665–1678 e1618. doi: 10.1016/j.cell.2018.09.038

Hua, C., Zhao, G., Li, Y., and Bie, L. (2014). Minichromosome Maintenance (MCM) family as potential diagnostic and prognostic tumor markers for human gliomas. *BMC Cancer* 14, 526. doi: 10.1186/1471-2407-14-526

Koschmann, C., Calinescu, A. A., Nunez, F. J., Mackay, A., Fazal-Salom, J., Thomas, D., et al. (2016). ATRX loss promotes tumor growth and impairs nonhomologous end joining DNA repair in glioma. *Sci. Transl. Med.* 8, 328ra328. doi: 10.1126/scitranslmed.aac8228

Kudinov, A. E., Karanicolas, J., Golemis, E. A., and Boumber, Y. (2017). Musashi RNA-Binding Proteins as cancer drivers and novel therapeutic targets. *Clin. Cancer Res. an Off. J. Am. Assoc. Cancer Res.* 23, 2143–2153. doi: 10.1158/1078-0432.CCR-16-2728

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* 12, 323. doi: 10.1186/1471-2105-12-323

Li, Y., Xu, J., Chen, H., Bai, J., Li, S., Zhao, Z., et al. (2013). Comprehensive analysis of the functional microRNA-mRNA regulatory network identifies miRNA signatures associated with glioma malignant progression. *Nucleic Acids Res.* 41, e203. doi: 10.1093/nar/gkt1054

Li, Y., Shao, T., Jiang, C., Bai, J., Wang, Z., Zhang, J., et al. (2015). Construction and analysis of dynamic transcription factor regulatory networks in the progression of glioma. *Sci. Rep.* 5, 15953. doi: 10.1038/srep15953

Li, H., Zhao, P., Xu, Q., Shan, S., Hu, C., Qiu, Z., et al. (2016). The autism-related gene SNRPN regulates cortical and spine development *via* controlling nuclear receptor Nr4a1. *Sci. Rep.* 6, 29878. doi: 10.1038/srep29878

Li, Y., Sahni, N., Pancsa, R., McGrail, D. J., Xu, J., Hua, X., et al. (2017). Revealing the determinants of widespread alternative splicing perturbation in cancer. *Cell Rep.* 21, 798–812. doi: 10.1016/j.celrep.2017.09.071

Li, C., Zheng, J., Chen, S., Huang, B., Li, G., Feng, Z., et al. (2018a). RRM2 promotes the progression of human glioblastoma. *J. Cell. Physiol.* 233, 6759–6767. doi: 10.1002/jcp.26529

Li, F., Jin, D., Tang, C., and Gao, D. (2018b). CEP55 promotes cell proliferation and inhibits apoptosis via the PI3K/Akt/p21 signaling pathway in human glioma U251 cells. *Oncol. Lett.* 15, 4789–4796. doi: 10.3892/ol.2018.7934

Li, Y., McGrail, D. J., Xu, J., Li, J., Liu, N. N., Sun, M., et al. (2019a). MERIT: systematic analysis and characterization of mutational effect on RNA interactome topology. *Hepatology* 70, 532–546. doi: 10.1002/hep.30242

Li, Y., Xiao, J., Bai, J., Tian, Y., Qu, Y., Chen, X., et al. (2019b). Molecular characterization and clinical relevance of m(6)A regulators across 33 cancer types. *Mol. Cancer* 18, 137. doi: 10.1186/s12943-019-1066-3

Liao, J. Y., Yang, B., Zhang, Y. C., Wang, X. J., Ye, Y., Peng, J. W., et al. (2019). EuRBPDB: a comprehensive resource for annotation, functional and oncological investigation of eukaryotic RNA Binding Proteins (RBPs). *Nucleic Acids Res.* doi: 10.1093/nar/gkz823

Liu, J., Liu, T., Wang, X., and He, A. (2017). Circles reshaping the RNA world: from waste to treasure. *Mol. Cancer* 16, 58. doi: 10.1186/s12943-017-0630-y

Miura, K., Obama, M., Yun, K., Masuzaki, H., Ikeda, Y., Yoshimura, S., et al. (1999). Methylation imprinting of H19 and SNRPN genes in human benign ovarian teratomas. *Am. J. Hum. Genet.* 65, 1359–1367. doi: 10.1086/302615

Nandakumar, P., Mansouri, A., and Das, S. (2017). The role of ATRX in glioma biology. *Front. Oncol.* 7, 236. doi: 10.3389/fonc.2017.00236

Neelamraju, Y., Gonzalez-Perez, A., Bhat-Nakshatri, P., Nakshatri, H., and Janga, S. C. (2018). Mutational landscape of RNA-Binding Proteins in human cancers. *RNA Biol.* 15, 115–129. doi: 10.1080/15476286.2017.1391436

Ostrom, Q. T., Bauchet, L., Davis, F. G., Deltour, I., Fisher, J. L., Langer, C. E., et al. (2014). The epidemiology of glioma in adults: a "state of the science" review. *Neuro-oncology* 16, 896–913. doi: 10.1093/neuonc/nou087

Palanichamy, K., Kanji, S., Gordon, N., Thirumoorthy, K., Jacob, J. R., Litzenberg, K. T., et al. (2017). NNMT silencing activates tumor suppressor PP2A, inactivates oncogenic STKs, and inhibits tumor forming ability. *Clin. Cancer Res. an Off. J. Am. Assoc. Cancer Res.* 23, 2325–2334. doi: 10.1158/1078-0432.CCR-16-1323

Panossian, A., Seo, E. J., and Efferth, T. (2018). Novel molecular mechanisms for the adaptogenic effects of herbal extracts on isolated brain cells using systems biology. *Phytomed. Int. J. Phytother. Phytopharmacol.* 50, 257–284. doi: 10.1016/j.phymed.2018.09.204

Pereira, B., Billaud, M., and Almeida, R. (2017). RNA-Binding Proteins in cancer: old players and new actors. *Trends Cancer* 3, 506–528. doi: 10.1016/j.trecan.2017.05.003

Philip, B., Yu, D. X., Silvis, M. R., Shin, C. H., Robinson, J. P., Robinson, G. L., et al. (2018). Mutant IDH1 promotes glioma formation *in vivo*. *Cell Rep.* 23, 1553–1564. doi: 10.1016/j.celrep.2018.03.133

Rasmussen, R. D., Gajjar, M. K., Tuckova, L., Jensen, K. E., Maya-Mendoza, A., Holst, C. B., et al. (2016). BRCA1-regulated RRM2 expression protects glioblastoma cells from endogenous replication stress and promotes tumorigenicity. *Nat. Commun.* 7, 13398. doi: 10.1038/ncomms13398

Reifenberger, G., Wirsching, H. G., Knobbe-Thomsen, C. B., and Weller, M. (2017). Advances in the molecular genetics of gliomas - implications for classification and therapy. *Nat. Rev. Clin. Oncol.* 14, 434–452. doi: 10.1038/nrclinonc.2016.204

Sebestyen, E., Singh, B., Minana, B., Pages, A., Mateo, F., Pujana, M. A., et al. (2016). Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res.* 26, 732–744. doi: 10.1101/gr.199935.115

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Sondka, Z., Bamford, S., Cole, C. G., Ward, S. A., Dunham, I., and Forbes, S. A. (2018). The COSMIC cancer gene census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* 18, 696–705. doi: 10.1038/s41568-018-0060-1

Sun, H., Yang, B., Zhang, H., Song, J., Zhang, Y., Xing, J., et al. (2019). RRM2 is a potential prognostic biomarker with functional significance in glioma. *Int. J. Biol. Sci.* 15, 533–543. doi: 10.7150/ijbs.30114

Uren, P. J., Vo, D. T., de Araujo, P. R., Potschke, R., Burns, S. C., Bahrami-Samani, E., et al. (2015). RNA-Binding Protein musashi1 is a central regulator of adhesion pathways in glioblastoma. *Mol. Cell. Biol.* 35, 2965–2978. doi: 10.1128/MCB.00410-15

Verhaak, R. G., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., et al. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell.* 17, 98–110. doi: 10.1016/j.ccr.2009.12.020

Wang, J., Liu, Q., and Shyr, Y. (2015). Dysregulated transcription across diverse cancer types reveals the importance of RNA-binding protein in carcinogenesis. *BMC Genomics* 16 Suppl 7, S5. doi: 10.1186/1471-2164-16-S7-S5

Wang, H., Song, X., Huang, Q., Xu, T., Yun, D., Wang, Y., et al. (2019). LGALS3 promotes treatment resistance in glioblastoma and is associated with tumor risk and prognosis. *Cancer Epidemiol. Biomarkers Prev. Publ. Am. Assoc. Cancer Res. cosponsored by Am. Soc. Prev. Oncol.* 28, 760–769. doi: 10.1158/1055-9965.EPI-18-0638

Xu, J., Li, Y., Lu, J., Pan, T., Ding, N., Wang, Z., et al. (2015). The mRNA related ceRNA-ceRNA landscape and significance across 20 major cancer types. *Nucleic Acids Res.* 43, 8169–8182. doi: 10.1093/nar/gkv853

Xu, J., Shao, T., Ding, N., Li, Y., and Li, X. (2017). miRNA-miRNA crosstalk: from genomics to phenomics. *Brief. Bioinform.* 18, 1002–1011. doi: 10.1093/bib/bbw073

Yi, S., Lin, S., Li, Y., Zhao, W., Mills, G. B., and Sahni, N. (2017). Functional variomics and network perturbation: connecting genotype to phenotype in cancer. *Nat. Rev. Genet.* 18, 395–410. doi: 10.1038/nrg.2017.8

Zhang, S., Zhao, B. S., Zhou, A., Lin, K., Zheng, S., Lu, Z., et al. (2017). m(6)A Demethylase ALKBH5 maintains tumorigenicity of glioblastoma stem-like cells by sustaining FOXM1 expression and cell proliferation program. *Cancer Cell.* 31, 591–606 e596. doi: 10.1016/j.ccell.2017.02.013

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 05 March 2020
doi: 10.3389/fgene.2020.00185

Integrated Analysis of the Functions and Prognostic Values of RNA Binding Proteins in Lung Squamous Cell Carcinoma

*Wei Li, Xue Li, Li-Na Gao and Chong-Ge You**

Laboratory Medicine Center, Lanzhou University Second Hospital, Lanzhou, China

Lung cancer is the leading cause of cancer-related deaths worldwide. Dysregulation of RNA binding proteins (RBPs) has been found in a variety of cancers and is related to oncogenesis and progression. However, the functions of RBPs in lung squamous cell carcinoma (LUSC) remain unclear. In this study, we obtained gene expression data and corresponding clinical information for LUSC from The Cancer Genome Atlas (TCGA) database, identified aberrantly expressed RBPs between tumors and normal tissue, and conducted a series of bioinformatics analyses to explore the expression and prognostic value of these RBPs. A total of 300 aberrantly expressed RBPs were obtained, comprising 59 downregulated and 241 upregulated RBPs. Functional enrichment analysis indicated that the differentially expressed RBPs were mainly associated with mRNA metabolic processes, RNA processing, RNA modification, regulation of translation, the TGF-beta signaling pathway, and the Toll-like receptor signaling pathway. Nine RBP genes (*A1CF, EIF2B5, LSM1, LSM7, MBNL2, RSRC1, TRMU, TTF2*, and *ZCCHC5*) were identified as prognosis-associated hub genes by univariate, least absolute shrinkage and selection operator (LASSO), Kaplan–Meier survival, and multivariate Cox regression analyses, and were used to construct the prognostic model. Further analysis demonstrated that high risk scores for patients were significantly related to poor overall survival according to the model. The area under the time-dependent receiver operator characteristic curve of the prognostic model was 0.712 at 3 years and 0.696 at 5 years. We also developed a nomogram based on nine RBP genes, with internal validation in the TCGA cohort, which showed a favorable predictive efficacy for prognosis in LUSC. Our results provide novel insights into the pathogenesis of LUSC. The nine-RBP gene signature showed predictive value for LUSC prognosis, with potential applications in clinical decision-making and individualized treatment.

Keywords: lung squamous cell carcinoma, RNA-binding proteins, prognostic signature, survival analysis, bioinformatics

OPEN ACCESS

Edited by:
Li Guo,
Nanjing University of Posts and Telecommunications, China

Reviewed by:
Wenhua Liang,
First Affiliated Hospital of Guangzhou Medical University, China
Yuming Jiang,
Stanford University, United States

***Correspondence:**
Chong-Ge You
youchg@lzu.edu.cn

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *27 October 2019*
Accepted: *17 February 2020*
Published: *05 March 2020*

Citation:
Li W, Li X, Gao L-N and You C-G (2020) Integrated Analysis of the Functions and Prognostic Values of RNA Binding Proteins in Lung Squamous Cell Carcinoma. Front. Genet. 11:185. doi: 10.3389/fgene.2020.00185

INTRODUCTION

Lung cancer is one of the most commonly diagnosed diseases and the leading cause of cancer-related deaths worldwide (Siegel et al., 2019). Lung squamous cell carcinoma (LUSC) accounts for 30% of lung cancer cases, resulting in about 0.4 million deaths each year worldwide (Siegel et al., 2013). Despite advances in diagnosis and treatment of lung cancer over the past few decades, there remains a lack of effective therapies for patients, underscoring the demand for novel treatment methods. Owing to differences in genetic and epigenetic changes among different subtypes of lung cancer, effective treatment targets of adenocarcinoma may not be suitable for LUSC (Wang et al., 2019). Therefore, a systematic study to explore the differentially expressed genes in LUSC is required to identify potential diagnostic markers and therapeutic targets for LUSC.

RNA binding proteins (RBPs) are proteins that interact with various types of RNA and are ubiquitously expressed in cells (Masuda and Kuwano, 2019; New et al., 2019; Otsuka et al., 2019). A total of 1542 RBPs have been identified by high-throughput screening in human cells, representing 7.5% of all protein coding genes (Gerstberger et al., 2014). These RBPs affect post-transcriptional events in cells and modulate cell physiology, and are therefore involved in multiple biological processes including RNA splicing, mRNA stability, export to the cytoplasm, localization, and protein translation (Masuda and Kuwano, 2019; Nahalka, 2019). Given that RPBs perform various critical functions in post-transcriptional events, it is unsurprising that alterations in RBPs are closely related to the initiation and progression of many human diseases. However, the roles of RBPs in the origin and development of cancer remain relatively unexplored.

In recent years, genome-wide analysis has indicated that many RBPs show dysregulated expression in tumors relative to adjacent normal tissues, and that their expression is associated with patient prognosis (Chen et al., 2019; Cooke et al., 2019; Zhang et al., 2019). It is well-known that the dysregulation of RBPs in cancer cells is mainly caused by genomic alterations, microRNA-mediated regulation, epigenetic mechanisms, and post-translational modifications (Gerstberger et al., 2014). Previous studies have linked known cancer drivers to RBP dysregulation. For example, the oncogene *crabp2* interacts with the RBP HuR to promote metastasis of lung cancer cells by regulating integrin β1/FAK/ERK signaling (Wu et al., 2019). Transforming growth factor-β (TGF-β) induces the expression of RNA-binding motif protein 38 (RBM38) in breast cancer, which promotes epithelial-to-mesenchymal transition by regulating the zonula occludens-1 transcript (Wu et al., 2017). The forkhead box K2 protein (FOXK2) promotes colorectal cancer metastasis by upregulating mRNA expression of zinc finger E-box binding homeobox 1 (ZEB1) (Du et al., 2019). Taken together, these studies indicate that the RBPs are closely related to the occurrence and development of human tumors. However, only a small fraction of RBPs have been studied intensively and found to have key roles in cancers to date. Therefore, we collected all relevant LUSC data from The Cancer Genome Atlas (TCGA) database and performed the present systematic analysis to examine the potential molecular functions and clinical significance of RBPs in LUSC. We identified multiple differentially expressed RBPs associated with LUSC, which provide new insight into the pathogenesis of the disease. Some of them may serve as potential biomarkers for diagnosis and prognosis.

MATERIALS AND METHODS

Data Preprocessing and Identification of Differentially Expressed RBPs

RNA sequencing data of 501 LUSC samples and 49 normal lung tissue samples with corresponding clinical information were downloaded from TCGA.[1] The raw data were preprocessed using the DESeq2 package.[2] Differentially expressed RBPs were identified based on a false discovery rate < 0.05 and $|\log_2$ fold change (FC)$| \geq 1$. All differentially expressed RBPs had an average count value more than 1.

GO and KEGG Functional Enrichment Analyses

The biological functions of these differently expressed RBPs were systematically investigated by gene ontology (GO) enrichment, which comprised three terms: molecular function, biological process, and cellular component. The Kyoto Encyclopedia of Genes and Genomes database (KEGG) was used to detect potential biological pathways of differentially expressed RBPs. All GO and KEGG pathway enrichment analyses were carried out using the WebGestalt (WEB-based Gene SeT AnaLysis Toolkit[3]) (Liao et al., 2019) with a *P*-value less than 0.05 and gene number more than 5.

Protein–Protein Interaction Network Construction and Module Screening

The protein–protein interactions (PPIs) among all differentially expressed RBPs were detected using the START (Search Tool for the Retrieval of Interacting Genes[4]) (Szklarczyk et al., 2019), and their network was constructed with the Cytoscape 3.7.0 software. Subsequently, the key modules were screened from the PPI network with scores >7 and node counts >5 by using the MCODE (Molecular Complex Detection) plug-in in Cytoscape. The cytoHubba plug-in was used to select hub genes. $P < 0.05$ was considered to indicate a significant difference.

Hub RBPs Expression Validation and Efficacy Evaluation

The Human Protein Atlas (HPA) database[5] (Uhlen et al., 2017) was used to detect the expression of 10 hub genes at a translational level. Receiver operating characteristic (ROC) curves were constructed with the GraphPad Prism 7.0

[1]https://portal.gdc.cancer.gov/

[2]http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html

[3]http://www.webgestalt.org/

[4]http://www.string-db.org/

[5]http://www.proteinatlas.org/

software to calculate the ability to discriminate between normal and tumor tissue.

Copy-Number Alterations and Mutation Analysis of Hub RBPs

The copy-number alteration and mutation data for all hub RBPs from the PPI network were identified using segmentation analysis and the GISTIC algorithm in cBioPortal[6] (Gao et al., 2013). Next, we carried out a co-expression analysis of all hub RBPs. Then we constructed a network including all hub genes and the 50 most frequently altered neighbor genes.

Prognosis-Related RBP Selection

The differentially expressed RBPs were subjected to a univariate Cox regression analysis using the survival package in R. A log-rank test was used to select the significant prognosis-related candidate RBPs, and the least absolute shrinkage and selection operator (LASSO), a widely used machine learning algorithm, was used to further predict the prognostic significance of candidate RBPs (iteration equal 1000) (Jiang et al., 2018). We also used a Kaplan–Meier test to evaluate the prognostic value of each candidate RBP identified by LASSO; the RBPs with *P*-value less than 0.05 were considered to be true prognosis-related RBPs.

[6]https://www.cbioportal.org/

Prognostic Model Construction and Evaluation

Based on the selected prognosis-related RBPs genes, we developed a multivariate Cox proportional hazards regression model to predict the prognosis of LUSC patients (Jiang et al., 2017). In this model, the risk score of each sample was calculated according to the following formula:

$$Risk\ score\ =\ \sum_{i=1}^{n} Expi\,\beta i,$$

where β represents the regression coefficient, and *Exp* represents the gene expression value.

To evaluate the performance of this prognostic model, LUSC patients from the TCGA with a survival time greater than 1 month were divided into low- and high-risk subgroups according to the median risk score, and the difference in overall survival (OS) between the two subgroups was compared by a log-rank

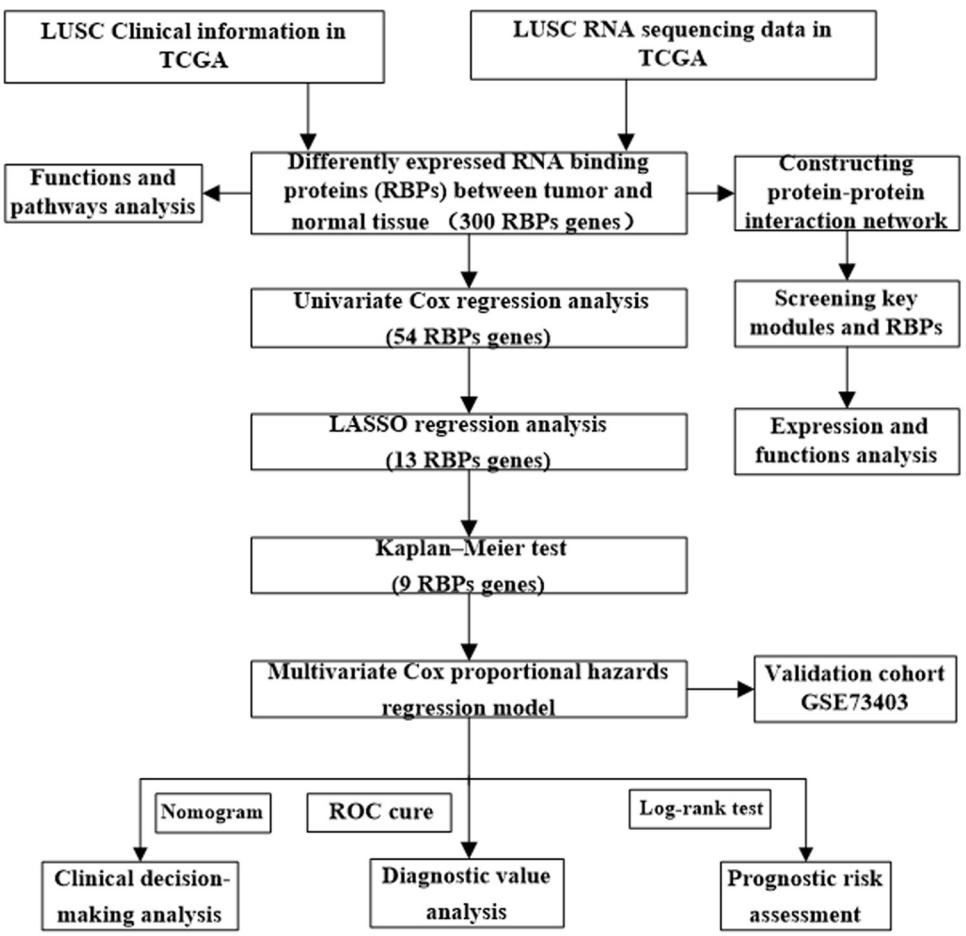


FIGURE 1 | Framework for analyzing the RBPs in LUSC.

TABLE 1 | GO and KEGG pathway analysis results for differentially expressed RBPs.

Expression	GO term	*P*-value
Up-regulated RBPs		
Biological processes	Cellular amide metabolic process	<0.001
	RNA processing	<0.001
	ncRNA metabolic process	<0.001
	RNA modification	<0.001
	Ribonucleoprotein complex biogenesis	<0.001
Molecular function	RNA binding	<0.001
	Catalytic activity, acting on RNA	<0.001
	Structural constituent of ribosome	8.34e-14
	Nuclease activity	1.33e-9
Cellular component	Nucleolus	<0.001
	Mitochondrial matrix	<0.001
	Sm-like protein family complex	<0.001
	Ribonucleoprotein complex	<0.001
KEGG pathway	RNA degradation	2.34e-12
	Ribosome biogenesis in eukaryotes	1.27e-11
	mRNA surveillance pathway	2.33e-10
	Spliceosome	<0.001
Down-regulated RBPs		
Biological processes	mRNA metabolic process	5.93e-10
	RNA processing	3.31e-7
	Defense response to virus	3.57-7
	Regulation of translation	7.47e-7
Molecular function	RNA binding	<0.001
	Poly-pyrimidine tract binding	5.66e-8
	Translation regulator activity	4.28e-8
Cellular component	Ribonucleoprotein complex	7.9078e-8
	Endolysosome membrane	0.0000033273
	RNA cap binding complex	0.0000041529
KEGG pathway	TGF-beta signaling pathway	0.001
	Toll-like receptor signaling pathway	0.0018
	mRNA surveillance pathway	0.02

test. Besides, the SurvivalROC R package was used to construct a ROC curve for prognostic performance of this model, and we drew a nomogram plot to forecast the likelihood of OS using the rms R package. Additionally, 69 LUSC patient samples from the GSE73403 dataset[7] were used as a validation cohort to confirm the predictive value of the prognostic model.

RESULTS

Selection of Differentially Expressed RBPs in LUSC

The workflow of this study is illustrated in **Figure 1**. RNA sequencing data for LUSC and corresponding clinical information were downloaded from the TCGA database. A total of 501 LUSC samples and 49 normal lung samples were analyzed. The DESEq2 software packages were used to preprocess these data and detect the differentially expressed RBPs. In total, 1542 RBPs (Gerstberger et al., 2014) were analyzed in this study, of which 300 met our inclusion criteria (adj $P < 0.05$, $|\log_2 FC| \geq 1.0$), comprising 59 downregulated and 241 upregulated RBPs. The expression distribution of these differentially expressed RBPs is shown in **Supplementary Figure S1**.

[7]https://www.ncbi.nlm.nih.gov/gds/?term=GSE73403

Functional Enrichment Analysis of the Differentially Expressed RBPs

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 and pathway analysis for these differentially expressed RBPs using the online tool WebGestalt. Upregulated differentially expressed RBPs were significantly enriched in biological processes associated with the cellular amide metabolic process, RNA processing, RNA metabolic process, RNA modification, and ribonucleoprotein complex biogenesis (**Table 1**). The downregulated differentially expressed RBPs were notably enriched in the mRNA metabolic process, RNA processing, defense response to virus, and regulation of translation (**Table 1**). 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, structural constituent of ribosome, and nuclease activity (**Table 1**), whereas the downregulated RBPs were significantly enriched in RNA binding, poly-pyrimidine tract binding, and translation regulator activity (**Table 1**). In regard to the cellular component, the upregulated RBPs were mainly enriched in the nucleolus, mitochondrial matrix, Sm-like protein family complex, and ribonucleoprotein complex, and downregulated RBPs were mainly enriched in the ribonucleoprotein complex, endolysosome membrane, and RNA cap binding complex (**Table 1**). Moreover, we found that downregulated differentially expressed RBPs were mainly enriched in the TGF-beta signaling pathway, Toll-like receptor signaling pathway, and mRNA surveillance pathway, whereas upregulated RBPs were significantly enriched for RNA degradation, ribosome biogenesis in eukaryotes, mRNA surveillance pathway, and the spliceosome (**Table 1**).

PPI Network Construction and Key Module Screening

We constructed a protein–protein co-expression network using Cytoscape software and the STRING database, in order to better understand the potential molecular functions of these differentially expressed RBPs in LUSC. This PPI network contained a total of 167 nodes and 771 edges (**Figure 2A**). Then we screened the hub genes by computing degree and betweenness, and obtained 10 candidate genes*: MRPL15, MRPL13, MRPL4, MRPL3, MRPL24, MRPS12, MRPL11, MRPL21, MRPL36,* and *MRPL47*. Subsequently, we further analyzed the co-expression network to detect potential critical modules by using the plug-in MODE in Cytoscape, and determined the top two significant modules. Module 1 included 18 nodes and 147 edges (**Figure 2B**), and module 2 consisted of 14 nodes and 91 edges (**Figure 2C**). The GO and pathway

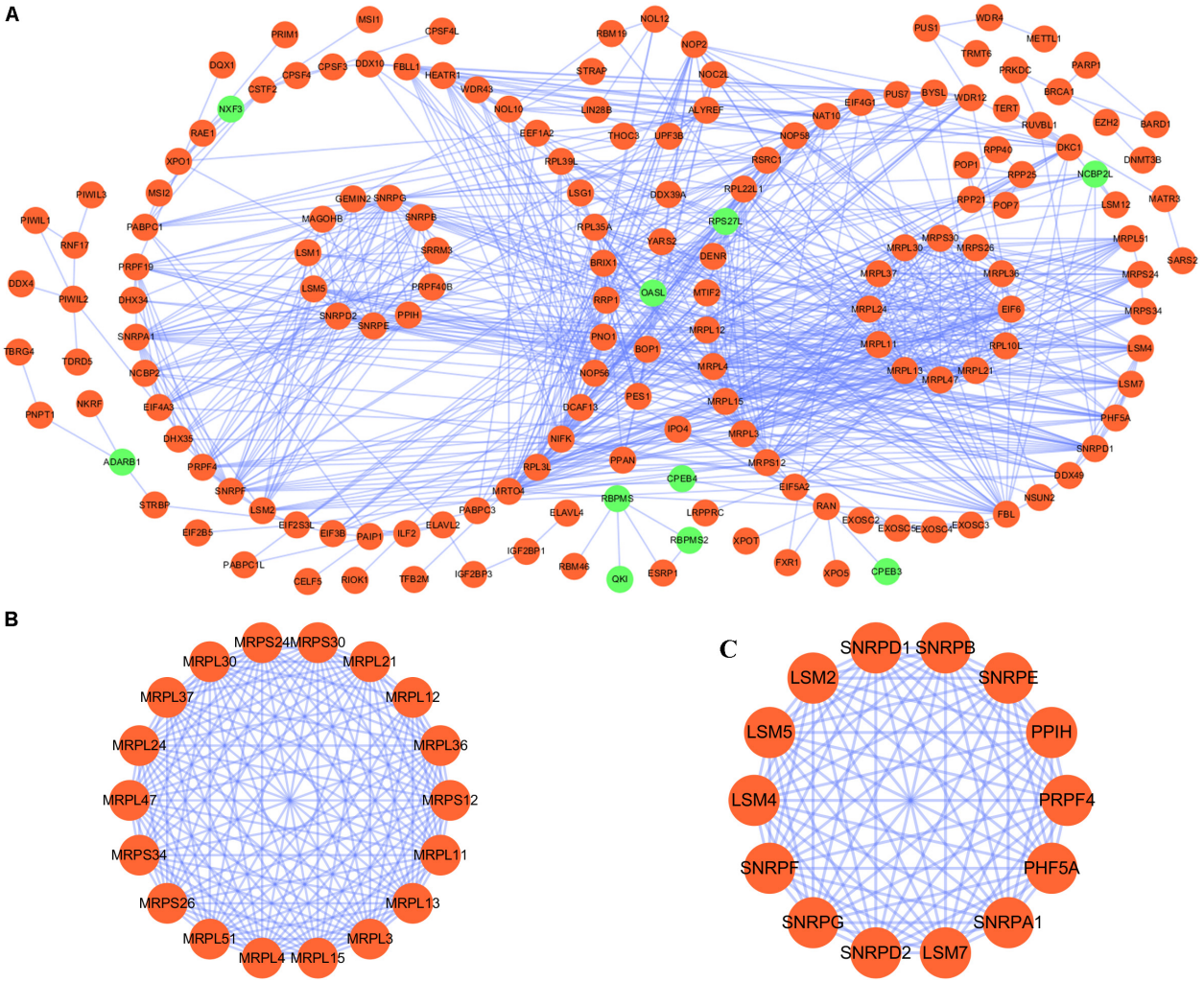


FIGURE 2 | PPI network and module analysis. **(A)** PPI network for RBPs; **(B)** critical module 1 in PPI network; **(C)** critical module 2 in PPI network.

analyses showed that the genes from module 1 were mainly enriched in mitochondrial translation, mitochondrial gene expression, and cellular protein complex disassembly, whereas the genes in module 2 were significantly enriched in spliceosomal snRNP assembly, mRNA splicing, mRNA metabolic process, and RNA processing.

Hub Gene Expression Validation

To further determine the expression of these hub genes in LUSC, we used immunohistochemistry results from the Human Protein Atlas database to show that MRPL15, MRPL13, MRPL4, MRPL3, MRPL24, MRPS12, MRPL11, MRPL21, MRPL36, and MRPL47 were significantly increased in lung cancer compared with normal lung tissue (**Figure 3**). Furthermore, we used ROC curve analysis to evaluate the efficacy of 10 hub genes to discriminate between carcinoma tissue and normal lung tissue. The area under the curve (AUC) of hub genes MRPL15 (AUC = 0.9585, 95% CI: 0.9376–0.9795, $P < 0.0001$), MRPL13 (AUC = 0.9480, 95% CI: 0.9111–0.9849, $P < 0.0001$), MRPL4 (AUC = 0.9578, 95% CI: 0.9407–0.9749, $P < 0.0001$), MRPL3 (AUC = 0.9943, 95% CI: 0.9896–0.9991, $P < 0.0001$), MRPL24 (AUC = 0.9415, 95% CI: 0.9158–0.9672, $P < 0.0001$), MRPS12 (AUC = 0.9862, 95% CI: 0.9758–0.9966, $P < 0.0001$), MRPL11 (AUC = 0.9393, 95% CI: 0.9062–0.9724, $P < 0.0001$), MRPL21 (AUC = 0.934, 95% CI: 0.9074–0.9608, $P < 0.0001$), MRPL36 (AUC = 0.9835, 95% CI: 0.9718–0.9953, $P < 0.0001$), and MRPL47 (AUC = 0.9845, 95% CI: 0.9751–0.9939, $P < 0.0001$) were all greater than 0.9, indicating that the hub genes had higher diagnostic accuracy for LUSC (**Figure 4**).

Mutation and Copy-Number Alteration Analysis of Candidate Hub Genes in LUSC Patients

Mutation and copy-number alteration (CNA) analyses of the hub genes MRPL15, MRPL13, MRPL4, MRPL3, MRPL24, MRPS12, MRPL11, MRPL21, MRPL36, and MRPL47 were carried out using the cBioPortal online tool for LUSC (TCGA, Provisional).

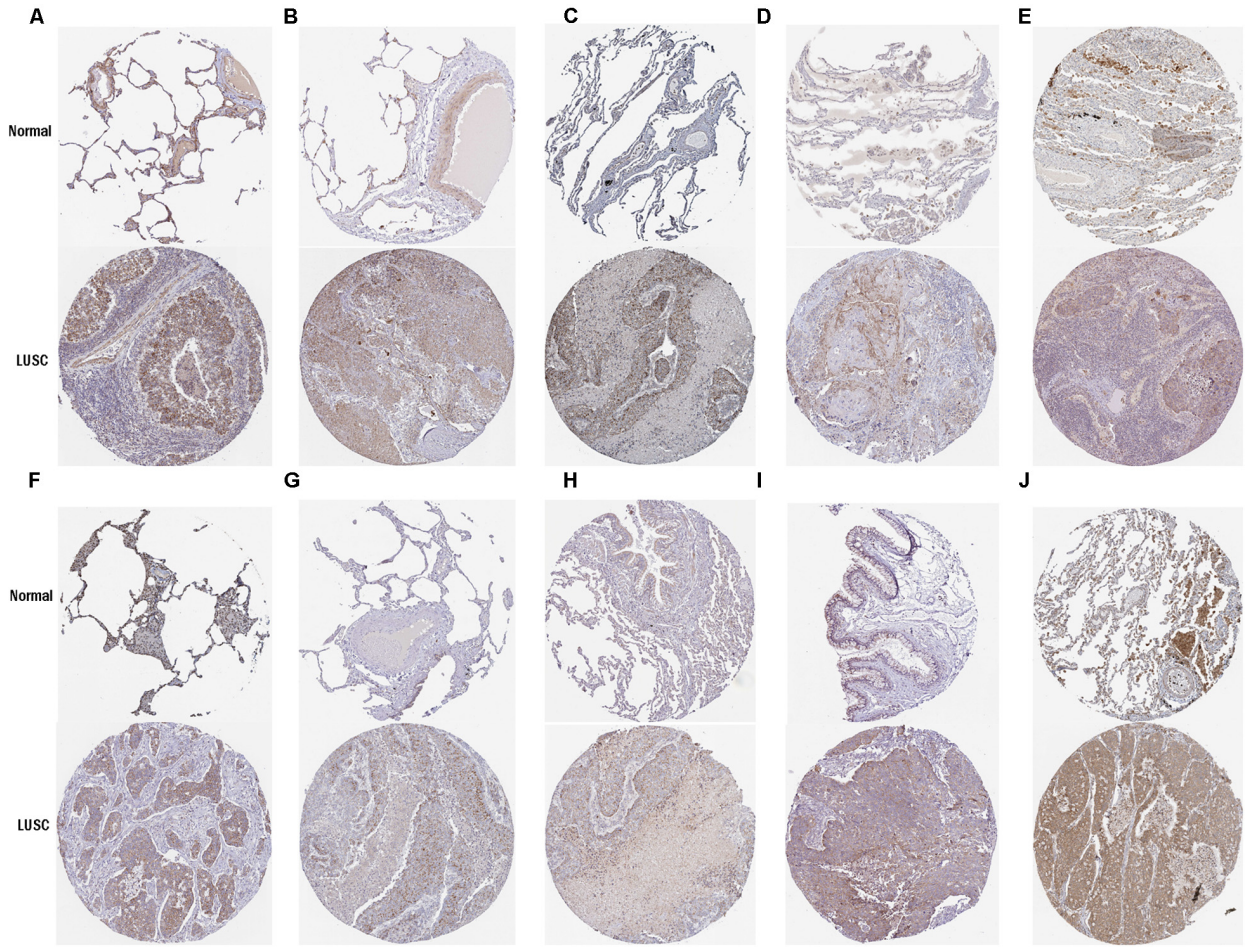


FIGURE 3 | Validation of protein expression of hub genes in normal lung tissue and LUSC using the HPA database. **(A)** MRPL15; **(B)** MRPL13; **(C)** MRPL4; **(D)** MRPL3; **(E)** MRPL24; **(F)** MRPS12; **(G)** MRPL11; **(H)** MRPL21; **(I)** MRPL36; **(J)** MRPL47.

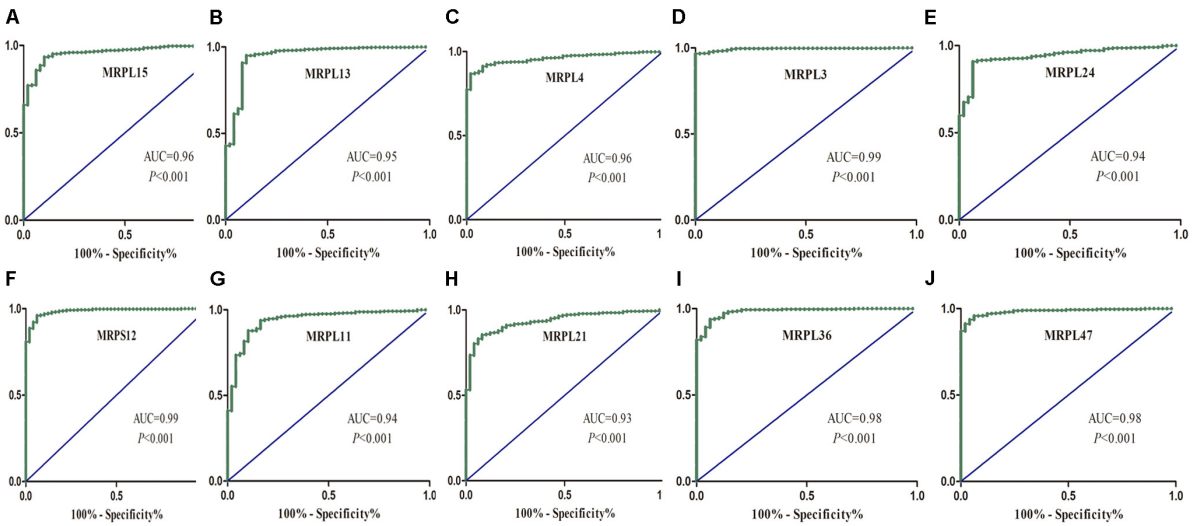


FIGURE 4 | ROC analysis of 10 hub RBPs based on the TCGA dataset. **(A)** MRPL15; **(B)** MRPL13; **(C)** MRPL4; **(D)** MRPL3; **(E)** MRPL24; **(F)** MRPS12; **(G)** MRPL11; **(H)** MRPL21; **(I)** MRPL36; **(J)** MRPL47.

The results showed that these 10 hub genes were altered in 178 samples out of 511 LUSC patients (35%). Two or more alterations were found in 68% of the LUSC samples (121 samples) (**Figures 5A,B**). The amplification of *MRPL47* was the most frequent copy-number alteration among these 10 hub genes. Then we constructed an interaction network containing 60 nodes, which comprised 10 hub genes and the 50 most frequently altered neighbor genes (**Figure 5C**). We also found that mitochondrial translation-related genes, including *GFM1, MTIF2, MTRF1, MRPS10, MRPS11, MRPL1, MRPL9*, and *PTCD3*, were closely associated with alterations of the 10 hub genes.

Prognosis-Related RBP Screening

Of the 300 differentially expressed RBPs, 54 were associated with prognosis as confirmed by univariate Cox regression analysis (**Supplementary Table S1**). Then we conducted a LASSO regression analysis to obtain the RBP genes with the best potential prognostic significance; 13 RBP genes, *A1CF*, *F4*, *DQX1*, *EIF2B5*, *GEMIN2*, *LSM1*, *LSM7*, *MBNL2*, *PABPC3*, *RSRC1*, *TRMU*, *TTF2*, and *ZCCHC5*, were selected (**Supplementary Figure S2**). To further determine the RBPs with the greatest potential prognosis ability, a Kaplan–Meier test for OS was used to identify nine RBP-coding genes, *A1CF*, *EIF2B5*, *LSM1*, *LSM7*, *MBNL2*, *RSRC1*, *TRMU*, *TTF2*, and *ZCCHC5* (**Figure 6**).

Prognosis-Related Genetic Risk Score Model Construction and Validation

The nine RBPs were analyzed by multiple stepwise Cox regression to construct a predictive model (**Table 2**). The risk score of each LUSC patient was computed according to the following formula:

$$Risk\ score = (0.0218 * \mathrm{ExpMBNL2}) + (-0.0134 * \mathrm{ExpLSM1}) + (2.4069 * \mathrm{ExpA1CF}) + (-0.0067 * \mathrm{ExpEIF2B5})$$

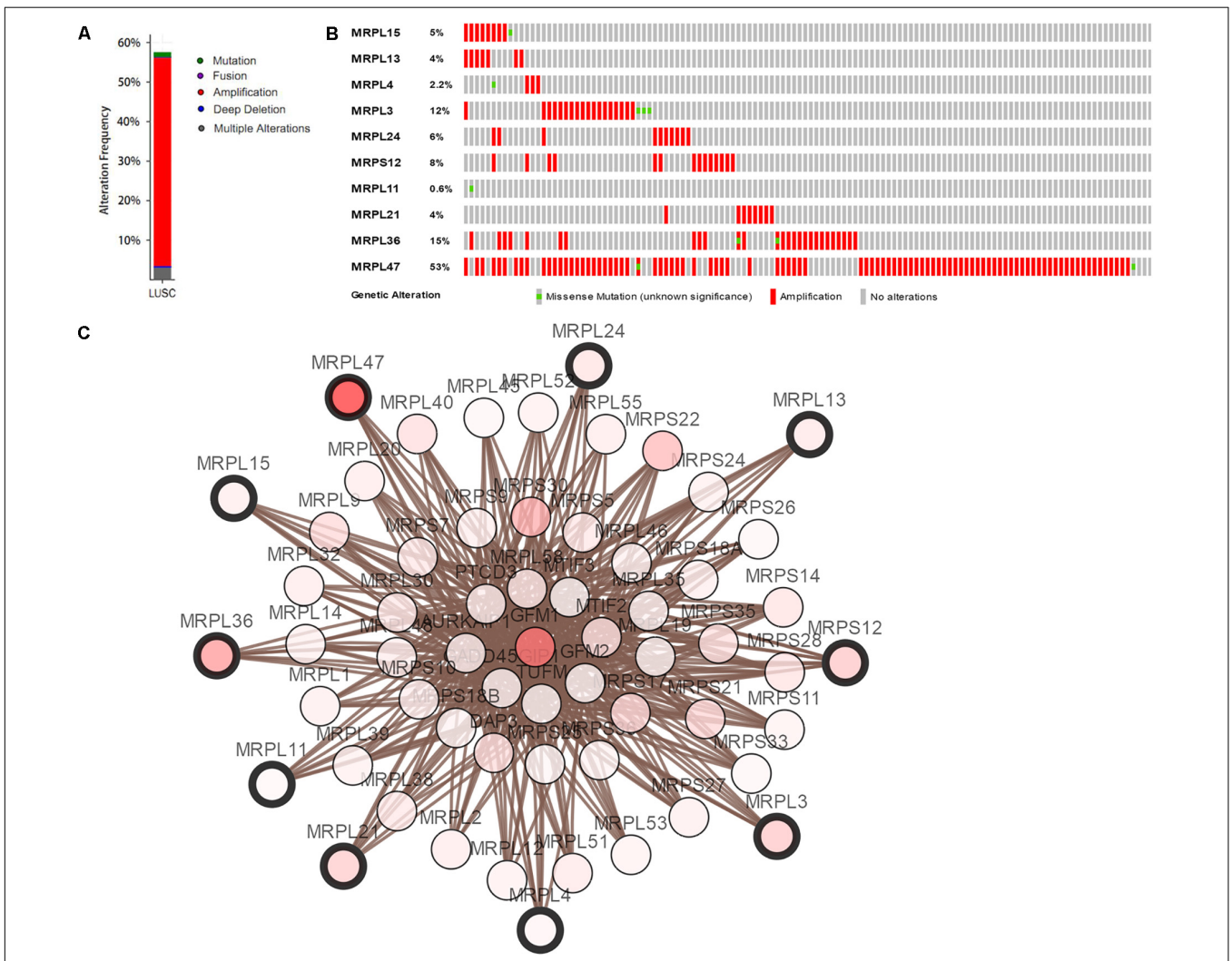


FIGURE 5 | Hub RBP expression and alteration analysis in LUSC. **(A)** Mutation frequency of hub genes; **(B)** mutation frequency of each gene; **(C)** interaction network.

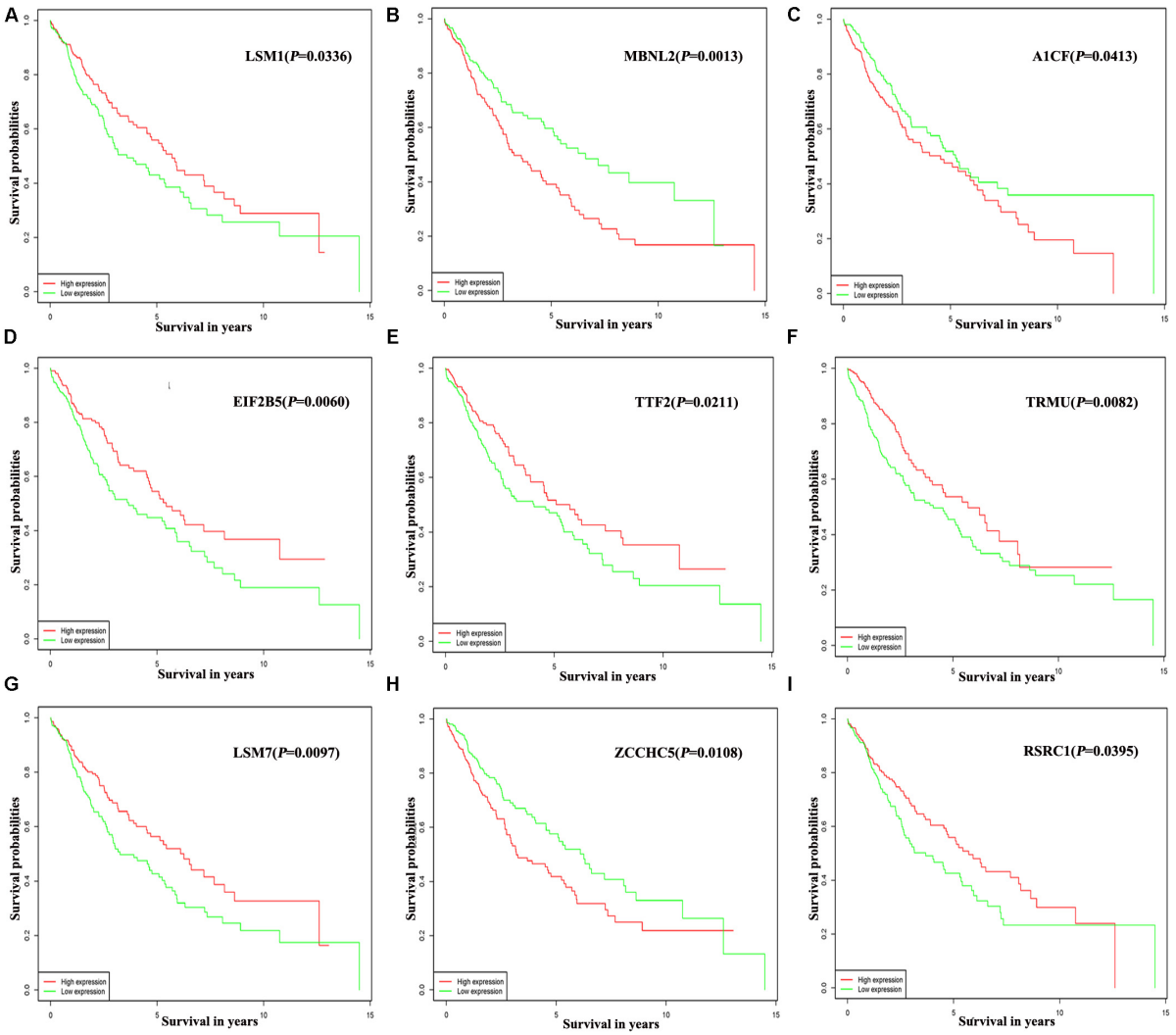


FIGURE 6 | Prognostic value of key nine RBPs in LUSC. **(A)** LSM1; **(B)** MBNL2; **(C)** A1CF; **(D)** EIF2B5; **(E)** TTF2; **(F)** TRMU; **(G)** LSM7; **(H)** ZCCHC5; **(I)** RSRC1.

$$+ (-0.0550 * \mathrm{ExpTTF2}) + (-0.0557 * \mathrm{ExpTRMU}) + (-0.0066 * \mathrm{ExpLSM7}) + (1.3639 * \mathrm{ExpZCCHC5}) + (-0.0132 * \mathrm{ExpRSRC1})$$

To assess the predictive ability of this model, we divided 424 LUSC patients into high- and low-risk groups for survival analysis according to the median risk score. Patients in the high-risk subgroup had a significantly lower OS rate than those in the low-risk subgroup (**Figure 7A**). Then we performed a time-dependent ROC analysis to further evaluate the prognostic performance of the nine-RBP gene signature; the AUC of the ROC curve for OS was 0.712 at 3 years and 0.696 at 5 years (**Figure 7B**). The expression heat map and survival status of patients with the nine-RBP gene biomarker in the low- and high-risk subgroups are shown in **Figure 7C**. These results reveal that our prognostic model had moderate sensitivity and specificity. Furthermore, we assessed whether the nine-RBP gene signature predictive model has similar prognostic ability in other LUSC patient cohorts; the same risk assessment formula was utilized to the GSE73403 datasets. The results indicated that patients with

TABLE 2 | Multivariate Cox regression analysis to identify prognosis-related hub RBPs.

Gene	Coef	Exp (coef)	Se (coef)	*z*	*Pr*(>| z|)
LSM1	−0.0134	0.9867	0.0062	−2.1820	0.0291
MBNL2	0.0218	1.0220	0.0114	1.9120	0.0558
A1CF	2.4069	11.0993	1.8701	1.2870	0.1981
EIF2B5	−0.0067	0.9933	0.0054	−1.2340	0.2171
TTF2	−0.0550	0.9465	0.0513	−1.0710	0.2844
TRMU	−0.0557	0.9458	0.0532	−1.0460	0.2955
LSM7	−0.0066	0.9934	0.0079	−0.8410	0.4004
ZCCHC5	1.3639	3.9115	1.7583	0.7760	0.4379
RSRC1	−0.0132	0.9869	0.0223	−0.5920	0.5536

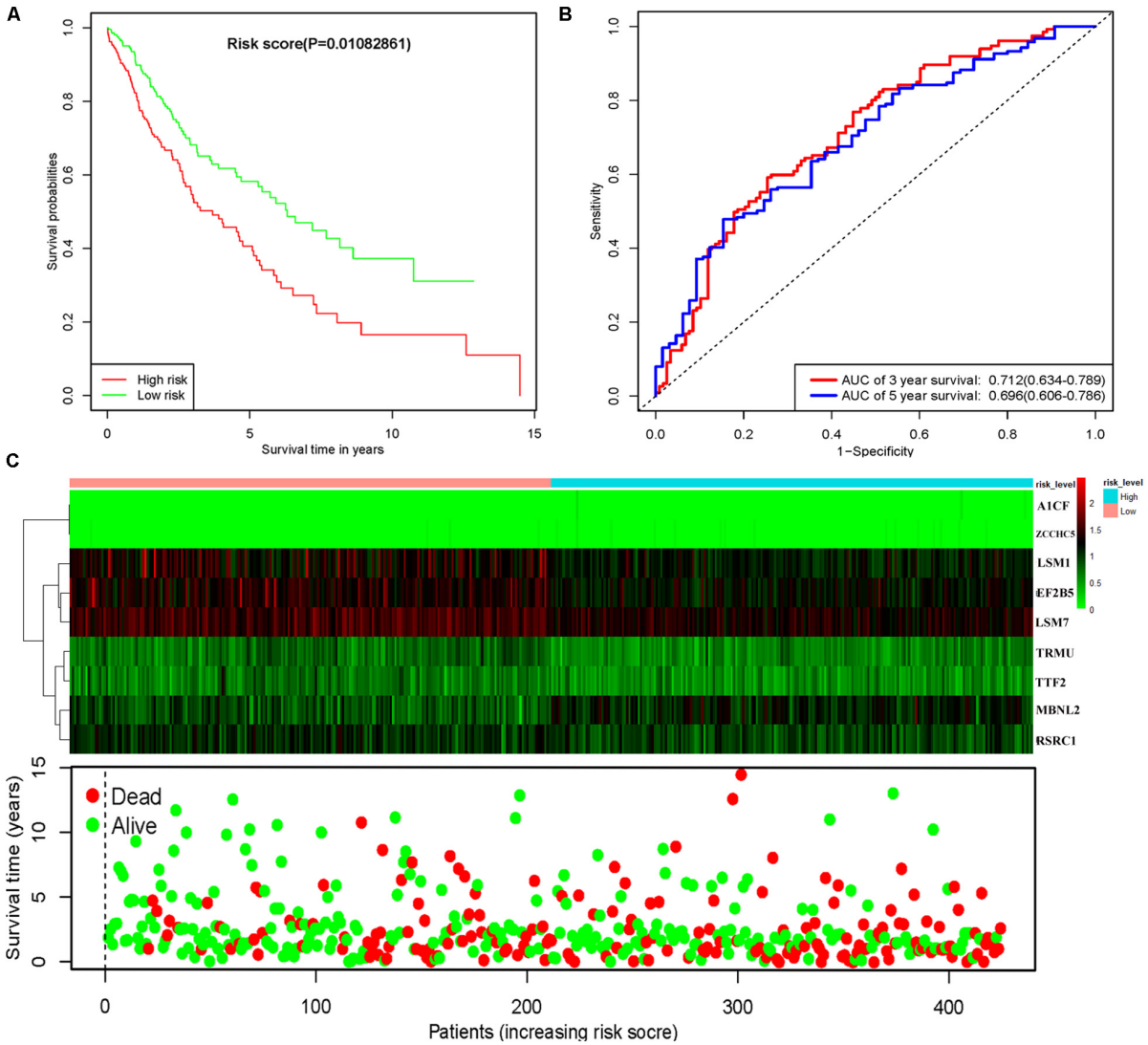


FIGURE 7 | Risk score analysis of nine-gene prognostic model in TCGA LUSC cohort. **(A)** Survival analysis according to risk score; **(B)** ROC analysis; **(C)** heat map and survival status of patients.

high-risk score had poorer OS than those with low-risk score in the GSE73403 cohorts (**Figures 8A–C**).

In order to construct a quantitative model for LUSC prognosis, we combined the nine-RBP marker to build a nomogram plot (**Figure 9A**). This allowed us to calculate the estimated survival probabilities of LUSC patients at 3 and 5 years by plotting a vertical line between the total point axis and each prognosis axis. We constructed calibration plots, which showed that there was good conformity between the predicted and observed outcomes (**Figures 9B,C**). We also calculated the concordance index for OS in the TCGA and GSE16011 cohorts, which were 0.69 and 0.66 respectively. In addition, we evaluated the prognostic value of different clinical features in 335 patients with LUSC by conducting a univariate regression analysis. The results indicated that age, smoking, stage, distant metastasis, and risk score were related to OS of LUSC patients ($P < 0.01$) (**Table 3**). However, we only found that age, smoking, and risk score were independent prognostic factors related to OS through multiple regression analysis (**Table 3**).

DISCUSSION

Malignant tumors are characterized by uncontrolled cell growth, which is mainly due to the dysregulated expression of cancer driver genes that regulate cell proliferation and differentiation. This includes gain of function mutations of oncogenes and functional deletion alterations of tumor-suppressor genes, or disabling of genome maintenance genes (Masuda and Kuwano, 2019; Zhou et al., 2019). Many studies have reported that RBPs show dysregulated expression in various human cancers (Dong et al., 2019; Soni et al., 2019; Velasco et al., 2019). However, little

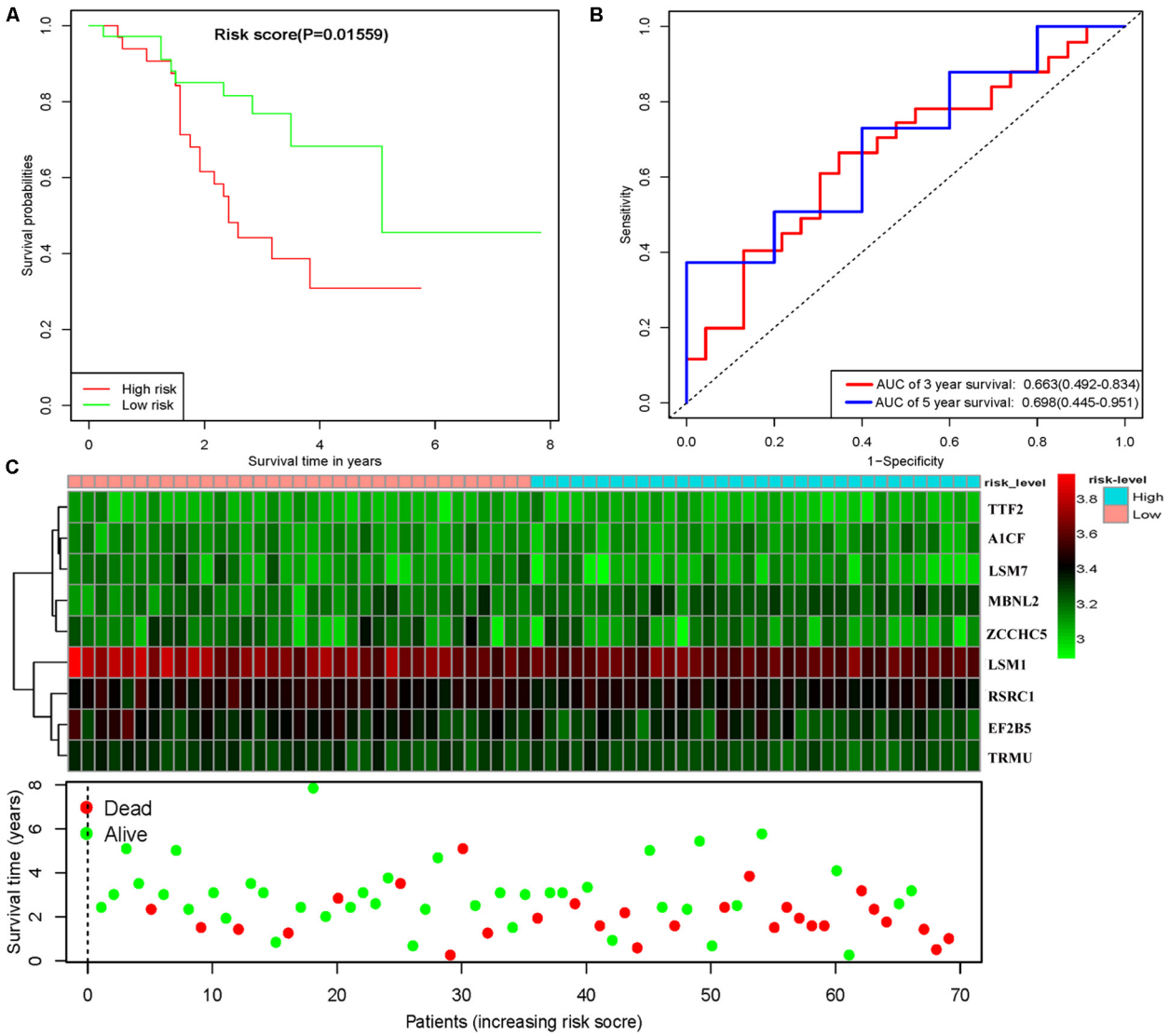


FIGURE 8 | Risk score analysis of nine-gene prognostic model in GSE73403 LUSC cohort. **(A)** Survival analysis according to risk score; **(B)** ROC analysis; **(C)** heat map and survival status of patients.

is currently known about the expression patterns and roles of RBPs in LUSC. In the present study, we integrated TCGA RNA sequencing data for LUSC and identified differentially expressed RBPs between cancer and normal tissue. We systematically investigated relevant biological pathways and constructed PPIs for these RBPs. Then, we performed survival analyses, ROC analyses, and copy-number alterations analyses to explore the potential biological functions and clinical values of the hub RBPs. We also screened key prognosis-related RBPs and constructed a risk model to predict LUSC prognosis based on a nine-RBP gene signature.

The biological functions and pathway enrichment analysis of these differentially expressed RBPs showed that the upregulated RBPs were significantly enriched in the cellular amide metabolic process, RNA processing, RNA metabolic process, RNA modification, RNA degradation, ribosome biogenesis, and mRNA surveillance pathway. The downregulated RBPs were mainly enriched in the mRNA metabolic process, RNA processing, regulation of translation, TGF-beta signaling pathway, and Toll-like receptor signaling pathway. In recent years, a large number of studies, have reported the role of aberrant RNA metabolism and RNA processing in various diseases (Li et al., 2017, 2018; Li S. et al., 2019; Li Y. et al., 2019). RNA processing factors were shown to have increased expression in poorly differentiated non-small-cell lung cancer cells (Geles et al., 2016). The TGF-beta signaling pathway is a classical tumorigenesis-related pathway; it exerts dual and opposing roles in oncogenesis, inhibiting cell proliferation in early tumors and inducing tumor progression and metastasis in advanced cancer (Seoane and Gomis, 2017; Batlle and Massague, 2019).

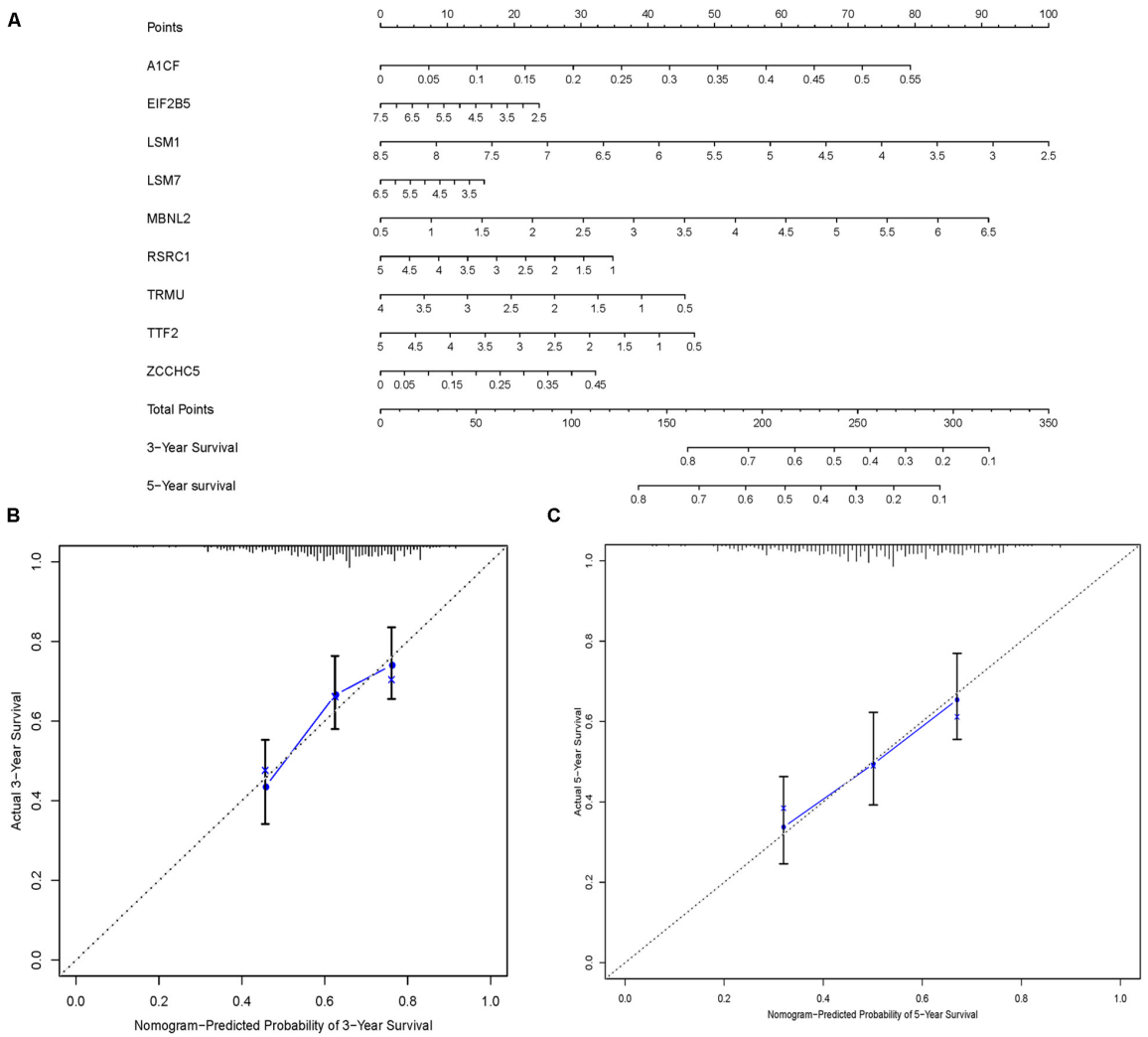


FIGURE 9 | Nomogram and calibration plots of nine RBPs. **(A)** Nomogram to predict 3- and 5-year OS in the TCGA cohort. **(B,C)** Calibration plots of the nomogram to predict OS at 3 and 5 years.

Previous studies have shown that RBPs can interact with the TGF-beta signaling pathway to regulate lung carcinogenesis (Kim et al., 2016; Bai et al., 2019). These results suggest that RBPs can affect the growth of tumor cells by regulating multiple biological processes, such as the TGF-beta signaling pathway, RNA metabolism, and RNA processing.

Subsequently, we obtained 10 hub RBPs by constructing a PPI network: MRPL15, MRPL13, MRPL4, MRPL3, MRPL24, MRPS12, MRPL11, MRPL21, MRPL36, and MRPL47. These hub RBPs are mitochondrial ribosomal proteins that are essential for maintaining mitochondrial functions. Impaired mitochondrial functions such as apoptosis and oxidative phosphorylation are found in most cancers, however, their mechanisms are unclear (Koc et al., 2015; Lee et al., 2017; Lin et al., 2019). Lee et al. (2017) found that suppressed MRPL13 expression increased hepatoma cell invasiveness. Koc et al. (2015) proposed that defects in mitochondrial function in head and neck squamous cell carcinoma might be caused by a decrease in MRPL11 expression. Shi et al. (2015) revealed that MRPL21 was significantly overexpressed in esophageal squamous cell carcinoma (ESCC) and could be used as a candidate prognostic

TABLE 3 | The prognostic effect of different clinical parameters.

	Univariate analysis			Multivariate analysis		
	HR	**95% CI**	***P*-value**	**HR**	**95%CI**	***P*-value**
Age	1.03	1.01–1.08	0.003	1.04	1.02–1.08	<0.001
Gender	1.19	0.81–1.75	0.380	1.21	0.82–1.80	0.333
Smoking	0.80	0.67–0.94	0.009	0.73	0.61–0.87	<0.001
Stage	1.24	1.01–1.51	0.036	1.50	0.93–2.41	0.0956
T	1.22	0.99–1.54	0.095	1.02	0.74–1.41	0.910
M	2.74	1.01–7.44	0.049	1.01	0.26–3.99	0.985
N	1.08	0.85–1.37	0.542	0.83	0.53–1.29	0.400
Risk score	1.93	1.55–2.40	<0.001	2.06	1.64–2.59	<0.001

biomarker. Although little is known about the relationship between mitochondrial ribosomal proteins and LUSC, our results indicate that impaired mitochondrial function is an important cause of LUSC, and further evaluation of potential roles of the 10 differentially expressed hub mitochondrial ribosomal proteins in LUSC may be worthwhile.

In addition, the prognosis-related hub RBPs were screened using univariate Cox regression analysis, LASSO regression analysis, Kaplan–Meier test, and multiple Cox regression analysis. We finally determined nine RBP-coding genes: *A1CF, EIF2B5, LSM1, LSM7, MBNL2, RSRC1, TRMU, TTF2*, and *ZCCHC5*. High expression of *LSM1, EIF2B5, TTF2, TRMU, LSM7,* and *RSRC1* was associated with a good prognosis in patients with LUSC, whereas that of *A1CF, MBNL2,* and *ZCCHC5* were related to poor prognosis. Next, the nine RBPs were used to construct a risk model by multiple stepwise Cox regression analysis to predict prognosis in LUSC patients. The ROC curve of the prognostic model showed that the nine-RBP genes signature had moderate performance for predicting OS at 3 years (AUC = 0.712) and 5 years (AUC = 0.696). A nomogram was constructed to enable practitioners to predict 3-, and 5-year OS of LUSC patients. According to the outcomes predicted by our model, patients with high risk scores have a poor prognosis, suggesting that they may need an adjusted treatment plan and individualized treatment.

Overall, our study provides novel insights into the role of RBPs in the tumorigenesis and progression of LUSC. Furthermore, our prognostic model showed good predictive performance with regard to survival, which may contribute to the development of new prognostic indicators for LUSC. Furthermore, the RBP-related gene marker showed a pivotal biological background, which demonstrates that these RBPs could be used in clinical adjuvant treatments. Nevertheless, our study had several limitations. First, our results were only based on single-omics (RNA sequencing); patients may exhibit heterogeneity owing to the different features of other omics data platforms. Moreover, our prognostic model was built on the TCGA LUSC dataset and was not validated with a clinical patient cohort; a prospective study should be performed to verify the results. Finally, the lack of some clinical characteristics in the datasets from TCGA may have decreased the statistical effectiveness and reliability of the multivariate stepwise Cox regression analysis.

CONCLUSION

We investigated the expression, potential functions, and prognostic values of aberrantly expressed RBPs via a series of bioinformatics analysis in LUSC. These RBPs were associated with oncogenesis, development, invasion, and metastasis. A nine-RBP coding gene prognostic model was developed that could act as an independent prognostic signature for LUSC. To the best of our knowledge, this is the first report of the establishment of an RBP-associated prognostic model for LUSC. These findings provide important insight into the pathogenesis of LUSC, which may contribute to clinical decision-making and individualized treatment.

DATA AVAILABILITY STATEMENT

The RNA-sequencing data of 501 LUSC samples and 49 normal lung tissue samples with corresponding clinical information were downloaded from the Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/).

AUTHOR CONTRIBUTIONS

WL and CG-Y conceived and designed the study and wrote the manuscript. WL, XL, and L-NG analyzed the data. All authors reviewed and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (81560343) and the Cuiying scientific and technological program of Lanzhou University Second Hospital (CY2018-MS10).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00185/full#supplementary-material

REFERENCES

Bai, H., Wang, C., Qi, Y., Xu, J., Li, N., Chen, L., et al. (2019). Major vault protein suppresses lung cancer cell proliferation by inhibiting STAT3 signaling pathway. *BMC Cancer* 19:454. doi: 10.1186/s12885-019-5665-6

Batlle, E., and Massague, J. (2019). Transforming growth factor-beta signaling in immunity and cancer. *Immunity* 2019, 924–940. doi: 10.1016/j.immuni.2019.03.024

Chen, H., Liu, J., Wang, H., Cheng, Q., Zhou, C., Chen, X., et al. (2019). Inhibition of RNA-binding protein Musashi-1 suppresses malignant properties and reverses paclitaxel resistance in ovarian carcinoma. *J. Cancer* 10, 1580–1592. doi: 10.7150/jca.27352

Cooke, A., Schwarzl, T., Huppertz, I., Kramer, G., Mantas, P., Alleaume, A. M., et al. (2019). The RNA-binding protein YBX3 controls amino acid levels by regulating SLC mRNA abundance. *Cell Rep.* 27, 3097e5–3106e5. doi: 10.1016/j.celrep.2019.05.039

Dong, W., Dai, Z. H., Liu, F. C., Guo, X. G., Ge, C. M., Ding, J., et al. (2019). The RNA-binding protein RBM3 promotes cell proliferation in hepatocellular carcinoma by regulating circular RNA SCD-circRNA 2 production. *eBio Medi.* 45, 155–167. doi: 10.1016/j.ebiom.2019.06.030

Du, F., Qiao, C., Li, X., Chen, Z., Liu, H., Wu, S., et al. (2019). Forkhead box K2 promotes human colorectal cancer metastasis by upregulating ZEB1 and EGFR. *Theranostics* 9, 3879–3902. doi: 10.7150/thno.31716

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6:l1. doi: 10.1126/scisignal.2004088

Geles, K. G., Zhong, W., O'Brien, S. K., Baxter, M., Loreth, C., Pallares, D., et al. (2016). Upregulation of RNA processing factors in poorly differentiated

lung cancer cells. *Trans.l Oncol.* 9, 89–98. doi: 10.1016/j.tranon.2016.01.006

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. *Nat. Rev. Genet.* 15, 829–845. doi: 10.1038/nrg3813

Jiang, Y., Li, T., Liang, X., Hu, Y., Huang, L., Liao, Z., et al. (2017). Association of adjuvant chemotherapy with survival in patients with stage II or III gastric cancer. *JAMA Surg.* 152:e171087. doi: 10.1001/jamasurg.2017.1087

Jiang, Y., Zhang, Q., Hu, Y., Li, T., Yu, J., Zhao, L., et al. (2018). ImmunoScore signature: a prognostic and predictive tool in gastric cancer. *Ann. Surg.* 267, 504–513. doi: 10.1097/SLA.0000000000002116

Kim, Y. E., Kim, J. O., Park, K. S., Won, M., Kim, K. E., and Kim, K. K. (2016). Transforming growth factor-beta-induced RBFOX3 inhibition promotes epithelial-mesenchymal transition of lung cancer cells. *Mol. Cells* 39, 625–630. doi: 10.14348/molcells.2016.0150

Koc, E. C., Haciosmanoglu, E., Claudio, P. P., Wolf, A., Califano, L., Friscia, M., et al. (2015). Impaired mitochondrial protein synthesis in head and neck squamous cell carcinoma. *Mitochondrion* 24, 113–121. doi: 10.1016/j.mito.2015.07.123

Lee, Y. K., Lim, J. J., Jeoun, U. W., Min, S., Lee, E. B., Kwon, S. M., et al. (2017). Lactate-mediated mitoribosomal defects impair mitochondrial oxidative phosphorylation and promote hepatoma cell invasiveness. *J. Biol. Chem.* 292, 20208–20217. doi: 10.1074/jbc.M117.809012

Li, S., Hu, Z., Zhao, Y., Huang, S., and He, X. (2019). Transcriptome-wide analysis reveals the landscape of aberrant alternative splicing events in liver cancer. *Hepatology* 69, 359–375. doi: 10.1002/hep.30158

Li, Y., McGrail, D. J., Xu, J., Li, J., Liu, N. N., Sun, M., et al. (2019). MERIT: systematic analysis and characterization of mutational effect on RNA interactome topology. *Hepatology* 70, 532–546. doi: 10.1002/hep.30242

Li, Y., McGrail, D. J., Xu, J., Mills, G. B., Sahni, N., and Yi, S. (2018). Gene regulatory network perturbation by genetic and epigenetic variation. *Trends Biochem. Sci.* 43, 576–592. doi: 10.1016/j.tibs.2018.05.002

Li, Y., Sahni, N., Pancsa, R., McGrail, D. J., Xu, J., Hua, X., et al. (2017). Revealing the determinants of widespread alternative splicing perturbation in cancer. *Cell Rep.* 21, 798–812. doi: 10.1016/j.celrep.2017.09.071

Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* 47, W199–W205. doi: 10.1093/nar/gkz401

Lin, Y., Jiang, M., Chen, W., Zhao, T., and Wei, Y. (2019). Cancer and ER stress: mutual crosstalk between autophagy, oxidative stress and inflammatory response. *Biomed. Pharmacother.* 118:109249. doi: 10.1016/j.biopha.2019.109249

Masuda, K., and Kuwano, Y. (2019). Diverse roles of RNA–binding proteins in cancer traits and their implications in gastrointestinal cancers. *Wiley Interdiscip. Rev. RNA* 10:e1520. doi: 10.1002/wrna.1520

Nahalka, J. (2019). The role of the protein-RNA recognition code in neurodegeneration. *Cell Mol. Life. Sci.* 76, 2043–2058. doi: 10.1007/s00018-019-03096-3

New, J., Subramaniam, D., Ramalingam, S., Enders, J., Sayed, A. A. A., Ponnurangam, S., et al. (2019). Pleotropic role of RNA binding protein CELF2 in autophagy induction. *Mol. Carcinog.* 58, 1400–1409. doi: 10.1002/mc.23023

Otsuka, H., Fukao, A., Funakami, Y., Duncan, K. E., and Fujiwara, T. (2019). Emerging evidence of translational control by AU-rich element-binding proteins. *Front. Genet.* 10:332. doi: 10.3389/fgene.2019.00332

Seoane, J., and Gomis, R. R. (2017). TGF-beta family signaling in tumor suppression and cancer progression. *Cold Spring Harb. Perspect. Biol.* 9:a022277. doi: 10.1101/cshperspect.a022277

Shi, Z. Z., Shang, L., Jiang, Y. Y., Shi, F., Xu, X., Wang, M. R., et al. (2015). Identification of genomic biomarkers associated with the clinicopathological parameters and prognosis of esophageal squamous cell carcinoma. *Cancer Biomark* 15, 755–761. doi: 10.3233/CBM-150517

Siegel, R., Naishadham, D., and Jemal, A. (2013). Cancer statistics, 2013. *CA Cancer J. Clin.* 63, 11–30. doi: 10.3322/caac.21166

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistics, 2019. *CA Cancer J. Clin.* 69, 7–34. doi: 10.3322/caac.21551

Soni, S., Anand, P., and Padwad, Y. S. (2019). MAPKAPK2: the master regulator of RNA-binding proteins modulates transcript stability and tumor progression. *J. Exp. Clin. Cancer Res.* 38, 121. doi: 10.1186/s13046-019-1115-1

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613.

Uhlen, M., Zhang, C., Lee, S., Sjostedt, E., Fagerberg, L., Bidkhori, G., et al. (2017). A pathology atlas of the human cancer transcriptome. *Science* 357:6352.

Velasco, M. X., Kosti, A., Penalva, L. O. F., and Hernandez, G. (2019). The diverse roles of RNA-binding proteins in glioma development. *Adv. Exp. Med. Biol.* 1157, 29–39. doi: 10.1007/978-3-030-19966-1_2

Wang, Q., Yang, S., Wang, K., and Sun, S. Y. (2019). MET inhibitors for targeted therapy of EGFR TKI-resistant lung cancer. *J. Hematol. Oncol.* 12:63. doi: 10.1186/s13045-019-0759-9

Wu, J., Zhou, X. J., Sun, X., Xia, T. S., Li, X. X., Shi, L., et al. (2017). RBM38 is involved in TGF-beta-induced epithelial-to-mesenchymal transition by stabilising zonula occludens-1 mRNA in breast cancer. *Br. J. Cancer* 117, 675–684. doi: 10.1038/bjc.2017.204

Wu, J. I., Lin, Y. P., Tseng, C. W., Chen, H. J., and Wang, L. H. (2019). Crabp2 promotes metastasis of lung cancer cells via HuR and integrin beta1/FAK/ERK signaling. *Sci. Rep.* 9:845. doi: 10.1038/s41598-018-37443-4

Zhang, J., Zheng, Z., Wu, M., Zhang, L., Wang, J., Fu, W., et al. (2019). The natural compound neobractatin inhibits tumor metastasis by upregulating the RNA-binding-protein MBNL2. *Cell Death Dis.* 10:554. doi: 10.1038/s41419-019-1789-5

Zhou, R., Shi, C., Tao, W., Li, J., Wu, J., Han, Y., et al. (2019). Analysis of mucosal melanoma whole-genome landscapes reveals clinically relevant genomic aberrations. *Clin. Cancer Res.* 25, 3548–3560. doi: 10.1158/1078-0432.CCR-18-3442

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 18 March 2020
doi: 10.3389/fgene.2020.00245

Identification of Autophagy-Associated Biomarkers and Corresponding Regulatory Factors in the Progression of Colorectal Cancer

Chunrui Zhang[1,2†], Jing Jiang[3†], Liqiang Wang[1†], Liyu Zheng[1], Jiankai Xu[1], Xiaolin Qi[4], Huiying Huang[1], Jianping Lu[1], Kongning Li[1,4]* and Hong Wang[4]**

[1] College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China, [2] Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China, [3] Obstetrics and Gynecology Department, The Second Affiliated Hospital of Harbin Medical University, Harbin, China, [4] Key Laboratory of Tropical Translational Medicine of Ministry of Education and College of Biomedical Information and Engineering, Hainan Medical University, Haikou, China

OPEN ACCESS

Edited by:
Li Guo,
Nanjing University of Posts and Telecommunications, China

Reviewed by:
Dapeng Hao,
University of Texas MD Anderson Cancer Center, United States
Yan Zhang,
Harbin Institute of Technology, China

***Correspondence:**
Hong Wang
wanghong@hainmc.edu.cn
Kongning Li
likongning@hainmc.edu.cn
Jianping Lu
lujianping1992@163.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: 26 December 2019
Accepted: 28 February 2020
Published: 18 March 2020

Citation:
Zhang C, Jiang J, Wang L, Zheng L, Xu J, Qi X, Huang H, Lu J, Li K and Wang H (2020) Identification of Autophagy-Associated Biomarkers and Corresponding Regulatory Factors in the Progression of Colorectal Cancer. Front. Genet. 11:245. doi: 10.3389/fgene.2020.00245

Autophagy is a self-degradation process that maintains homeostasis against stress in cells. Autophagy dysfunction plays a central role in the development of tumors, such as colorectal cancer (CRC). In this study, autophagy-related differentially expressed genes, their downstream functions, and upstream regulatory factors including RNA-binding proteins (RBP) involved in programmed cell death in the CRC were investigated. Transcription factors (TFs) and miRNAs have been shown to mainly regulate autophagy genes. Interestingly, we found that some of the RBP in the CRC, such as DDX17, SETDB1, and POLR3A, play an important regulatory role in maintaining autophagy at a basal level during growth by acting as TFs that regulate autophagy. Promoter methylations showed negative regulations on differentially expressed autophagy gene (DEAG), while copy number variations revealed a positive role in them. A proportional hazards regression analysis indicated that using autophagy-related prognostic signature can divide patients into high-risk and low-risk groups. Autophagy associated FDA-approved drugs were studied by a prognostic network. This would contribute to the identifications of new potential molecular therapeutic targets for CRC.

Keywords: autophagy, colorectal cancer, regulatory network, RNA-binding proteins, biomarkers

INTRODUCTION

Colorectal cancer (CRC) is a common digestive tract tumor (Chisanga et al., 2016). Among all cancer types, it is the third leading cause of death in the world. The overall 5-year survival rate of CRC patients is less than 40%, and the occurrence of CRC is consistently rising (Yang et al., 2015). However, the prognosis and therapy for CRC have not been significantly improved. Therefore, a proper selection of patients for aggressive treatment is necessary, new therapeutic strategies and prediction of prognosis of CRC is urgently needed.

Autophagy has been found to be associated with a variety of clinically relevant diseases, such as CRC. In the past ten years, autophagy has received extensive attention as a new treatment method.

Several studies indicate that the autophagy function plays a critical role in the development, maintenance, and progression of CRC (Yang et al., 2015; Katheder et al., 2017). The dysregulation of autophagy function disrupts the physiological processes and has been implicated in the pathogenesis of multiple diseases (Thorburn et al., 2014). Early efforts reported that there are relationships between multidimensional factors and autophagy function. BECN1 plays a key role in the autophagic process, its expression is found to be regulated by transcription factors (TFs), miRNAs, the abnormal methylation of the promoter region, and copy number variation (CNV) of the associated chromatin regions (Mei et al., 2016). In addition, RNA-binding proteins (RBP) play a key role in many processes as TF, including cellular differentiation, autophagy, apoptosis, and DNA repair (Gerstberger et al., 2014; Williams et al., 2019). For instance, some researches have shown that CELF2 RNA-binding protein regulates autophagy-mediated CRC cell death (New et al., 2019). Furthermore, Kudinov AE et al. found that MSI2 RNA-binding protein as a regulator of progenitor cell is elevated in colorectal adenocarcinomas and that its loss of function inhibits the growth of CRC cells (New et al., 2019). In the past decade, autophagy as a new therapeutics has attracted extensive attention. Increasing evidence indicates that autophagy function is crucial to tumor cell survival in CRC patients undergoing anticancer treatment (Roy and Debnath, 2010). Despite this, the potential values of some novel prognostic biomarkers related to autophagy function have not been thoroughly investigated. This study will focus on the potential prognostic roles of autophagy-related genes in CRC and will offer new targets for the treatment of CRC. Further understanding of the functional role of autophagy in CRC pathogenesis will allow us to improve the disease management.

In this study, the function of autophagy genes in four stages of CRC was investigated through the performance of functional enrichment analysis of the downstream RNAs. The upstream regulatory factors of autophagy genes were also identified in each stage by integrating multi-omics data in TCGA. Some key autophagy-related differentially expressed genes associated with the prognosis of CRC were identified through univariate Cox proportional hazards regression model. Then the mappings were drawn between FDA-approved drugs and their related autophagy gene. These findings not only shed light on the central functional role of autophagy-related genes in CRC, but may also contribute to the identification of molecular biomarkers in CRC and the development of clinical therapeutic modality.

MATERIALS AND METHODS

Colorectal Cancer Patient Cohorts

Gene and miRNA expression data, methylation data, and the clinical data of CRC patients were downloaded from TCGA[1] (Cancer Genome Atlas Research Network, 2008). There are 328 colon carcinoma (COAD) samples and 105 rectal carcinomas (READ) samples. Combined with clinical information, there are 41 normal patients, 45 stage I patients, 111 stage II patients, 83 stage III patients, and 39 stage IV patients in COAD samples. And there are 10 normal patients, 12 stage I patients, 28 stage II patients, 33 stage III patients, and 15 stage IV patients in READ samples. The corresponding CNV data were obtained from the Cancer Cell Line Encyclopedia[2] (Barretina et al., 2012). Additionally, a cohort of 177 COAD patients and 196 READ patients from the GEO database (GSE17536 and GSE87211) (Smith et al., 2010; Hu et al., 2018) was used as an independent external test set.

Autophagy Genes, Interaction Data, RNA-Binding Proteins, and Transcription Start Sites

Autophagy genes were collected from the cell death proteomics database[3] (Arntzen et al., 2013). A total of 1776 experimentally confirmed genes were used for the subsequent analysis. Protein-protein interactions were retrieved from the Human Protein Reference Database (HPPD)[4] (Keshava Prasad et al., 2009). The TF that targeted the autophagy genes were acquired from ChIPBase[5] (Zhou et al., 2017). The 2949 RBP were downloaded from the EuRBPDB[6] (New et al., 2019). The miRNA-gene targeted interaction was formed through the integration of miRecords[7] (Xiao et al., 2009), DIANA-TarBase[8] (Vlachos et al., 2015), and miRTarBase[9] (Chou et al., 2018) databases.

The transcription start sites (TSS) of autophagy genes were downloaded from GENCODE (Harrow et al., 2012). The mean of the methylation level for CG sites in autophagy gene transcription promoter regions was used as the methylation level of the autophagy genes.

Construction of Regulatory Networks and the Influence of Regulators on the Differentially Expressed Autophagy Gene

The significant differentially expressed autophagy gene (DEAG) and regulated gene pairs were obtained through the calculation of their linear correlation based on the expression data ($P < 0.05$). Linear regression was then used to calculate the significant TFs and miRNAs that targeted the DEAG based on known TF/miRNA-gene interaction ($P < 0.05$). The significant influence of CNVs or gene promoter methylations was denoted by the linear correlation between gene expression and their own CNVs or gene promoter methylations level ($P < 0.05$).

Statistical Analysis

The relationship between DEAGs expression level and patient survival was evaluated by the Cox regression analysis. Multivariate Cox regression analysis was used to fit the selected

[1] http://cancergenome.nih.gov/

[2] www.broadinstitute.org/ccle

[3] http://celldeathproteomics.uio.no./

[4] http://www.hprd.org/

[5] http://rna.sysu.edu.cn/chipbase/

[6] http://eurbpdb.syshospital.org/

[7] http://miRecords.umn.edu/miRecords

[8] http://www.microrna.gr/tarbase

[9] http://miRTarBase.mbc.nctu.edu.tw/

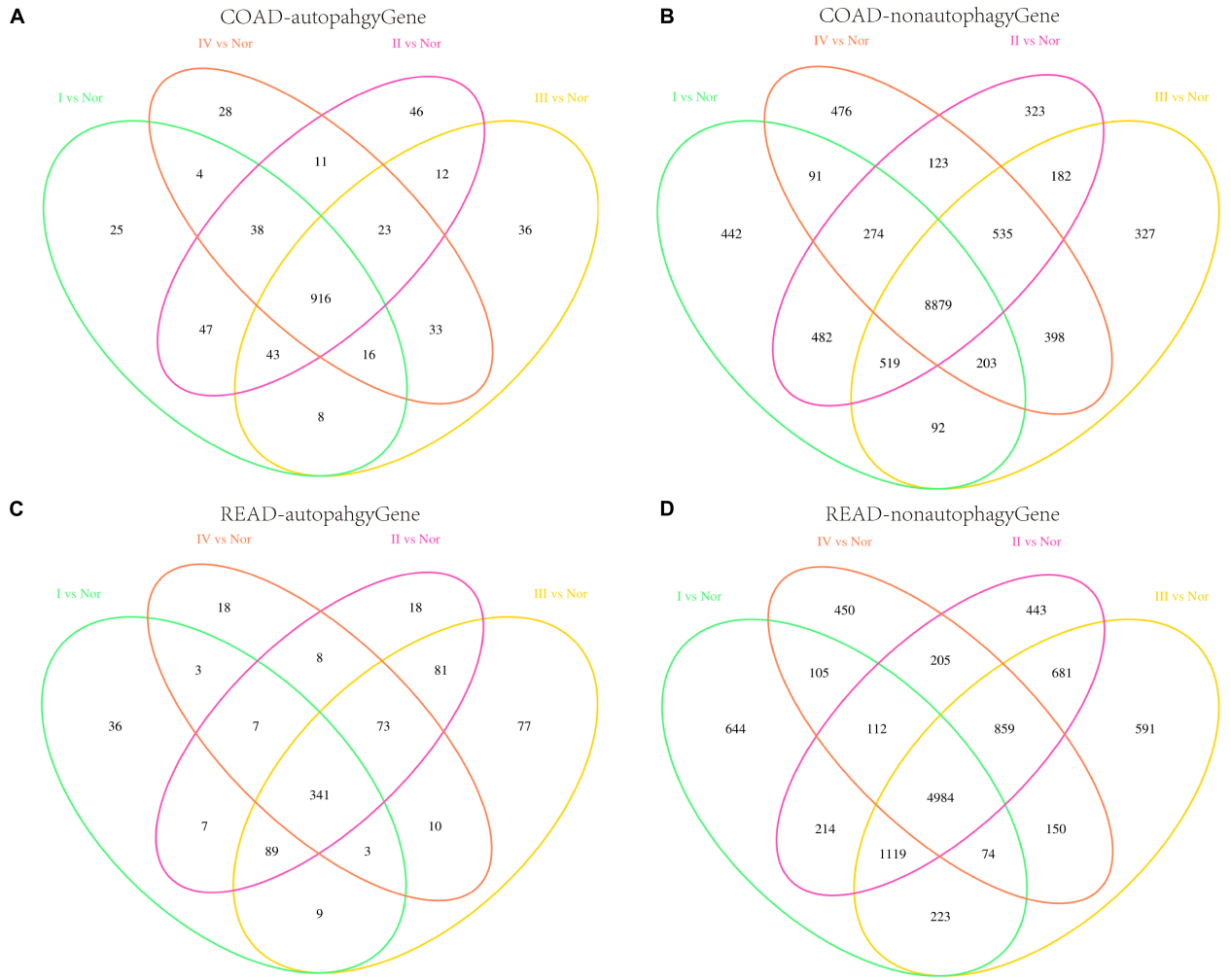


FIGURE 1 | Differentially expressed genes in the four stages. **(A,B)** Represent the Venn diagrams of differentially expressed autophagy genes and non-autophagy genes respectively in four stages of COAD. **(C,D)** Represent the Venn diagrams of differentially expressed autophagy genes and non-autophagy genes respectively in four stages of READ.

DEAGs (Lossos et al., 2004). The risk score of each patient was calculated with the estimated regression coefficient as the weight (Zhou et al., 2016). It was calculated as follows:

$$\text{Risk\_score} = \sum_{i=1}^{n} \beta_i \times \text{EXP}_{\text{gene}(i)}$$

where β_i is the Cox regression coefficient of gene i in the training set, and n is the number of survival related genes. The sensitivity and specificity of survival gene risk prediction were compared using the time-dependent receiver operating characteristic (ROC) curves, and the optimal patient stratified cutoff value was determined in the discovery cohort. Patients were divided into high risk group and low risk group in accordance with the above stratification cutoff. Kaplan–Meier survival analysis and log-rank test were performed to compare survival differences. Cox proportional risk regression was used for multivariate analysis to test whether the autophagy gene signature was independent of other clinic-pathological factors. The Cox proportional risk regression model was used to estimate the hazard ratio (HR) and the 95% confidence intervals (CI).

RESULTS

Construction of DEAG Regulatory Network

Genome-wide analysis of mRNA expression was performed to identify differentially expressed mRNAs, and the autophagy genes were extracted. There were 1097, 1136, 1087, and 1069 DEAGs between the normal and each stage (I, II, III, and IV) cancer samples in COAD. There were 495, 624, 683, and 463 DEAGs in READ ($P < 0.05$). It was found that a large proportion of the DEAGs were shared among the four stages (**Figures 1A,C**). Used Chi-square Test, we found the four stages significantly shared the majority of the DEAGs. The *P*-values

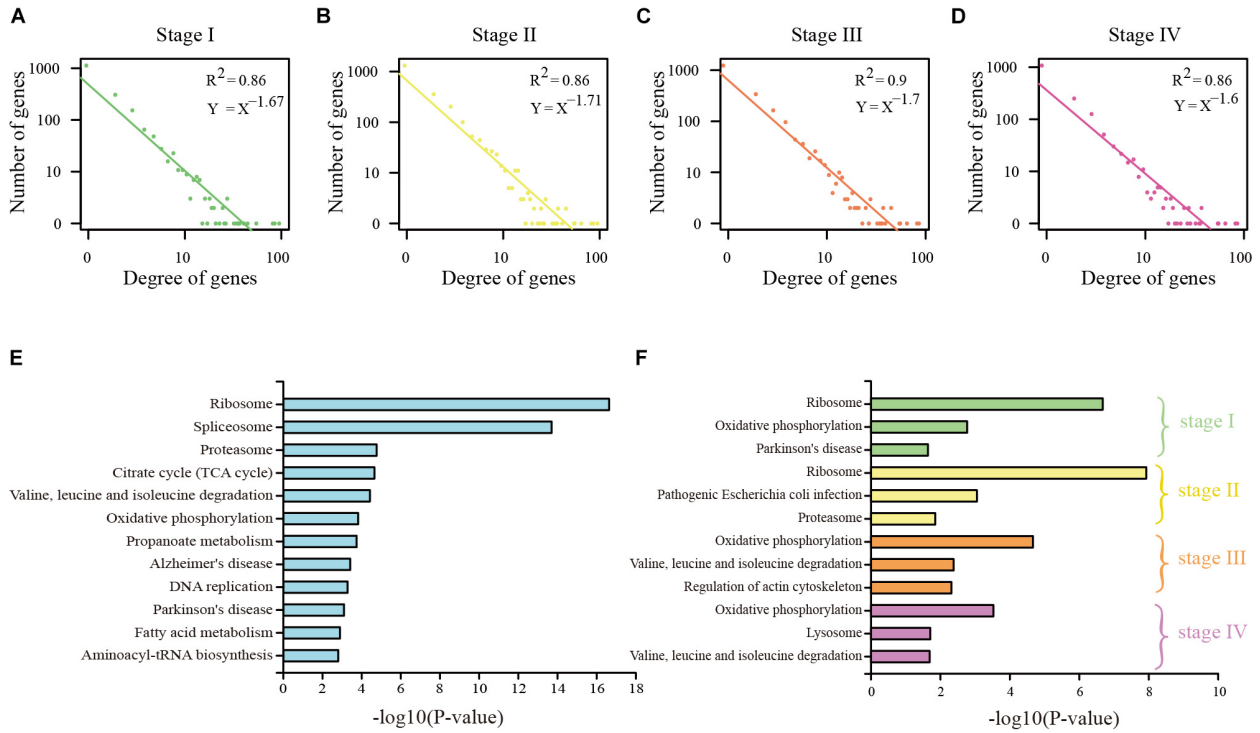


FIGURE 2 | Analysis of DEAG regulatory network in COAD. **(A)** The distribution of the degree of DEAG regulatory network in stage I. **(B)** The distribution of the degree of DEAG regulatory network in stage II. **(C)** The distribution of the degree of DEAG regulatory network in stage III. **(D)** The distribution of the degree of DEAG regulatory network in stage IV. **(E)** Significantly enriched KEGG pathways of common DEAGs. **(F)** Significantly enriched KEGG pathways of the DEAGs in every stage.

of each stage of COAD are 0.03601, 0.09514, 0.00041, and 0.00014, respectively. Genes which appeared in three, two and single stages were infrequent. By contrast differentially expressed non-autophagy genes that only appeared in a single stage were most common (**Figures 1B,D**). We identified 12079, 12453, 12222, and 12048 differentially expressed genes between the normal and each stage (I, II, III, and IV) cancer samples in COAD. And there were 7970, 9241, 9364, and 7402 differentially expressed genes in READ. The autophagy genes were extracted, there were 1097, 1136, 1087, and 1069 DEAGs in COAD and 495, 624, 683, and 463 DEAGs in READ. So differentially expressed non-autophagy genes were 10982, 11317, 11135, and 10979 in COAD and 7475, 8617, 8681, and 6939 in READ. These results imply that autophagy genes play an important role during the development and progression of CRC. Whereas a few autophagy genes, which were expressed differently in specific stages of cancer, reflect that these genes play a different role in different stages.

To study the regulation ability of the DEAGs, the regulatory network was constructed by the linear regression method (**Supplementary Figures S1, S2**). Through network topology analysis, it was discovered that the networks exhibit power law degree distribution. This illustrates the scale-free and small-world nature of these networks, which makes them similar to the general biological network (**Figures 2A–D** and **Supplementary Figures S3A–D**). In conclusion, many pieces of evidence indicate that specific DEAGs and their regulatory subnetwork in each of the cancer stages can better represent the function of autophagy genes in its own stage.

Functional enrichment analysis for KEGG pathway was performed on the common DEAGs and the specific DEAGs through the use of DAVID 6.8 bioinformatics tool (Huang et al., 2009). Common DEAGs in COAD were enriched in 22 KEGG pathways ($P < 0.05$), including ribosome, spliceosome, proteasome, propanoate metabolism, and fatty acid metabolism (**Figure 2E**). In READ, there were spliceosome, methane metabolism, fatty acid metabolism, and propanoate metabolism (**Supplementary Figure S3E**). This is consistent with the findings of Y. Boglev et al. Genetic mutations associated with ribosomal production provide a powerful stimulus to autophagy in affected tissues, allowing them to escape cell death. Autophagy is a specific response to damage in ribosome organisms (Boglev et al., 2013). However, the influence of DEAGs in each distinct stage was found to be a little different (**Figure 2F** and **Supplementary Figure S3F**). Stage I and stage II were analogous, and stage III and stage IV were analogous. This reflects that the genes in stage I and stage II play a similar role, and the genes in stage III and stage IV play a similar role.

Analysis of DEAG Regulatory Mechanism With Multi-Omics Data

Along with the development and maturation of the new generation sequencing technology, more and more multi-omics

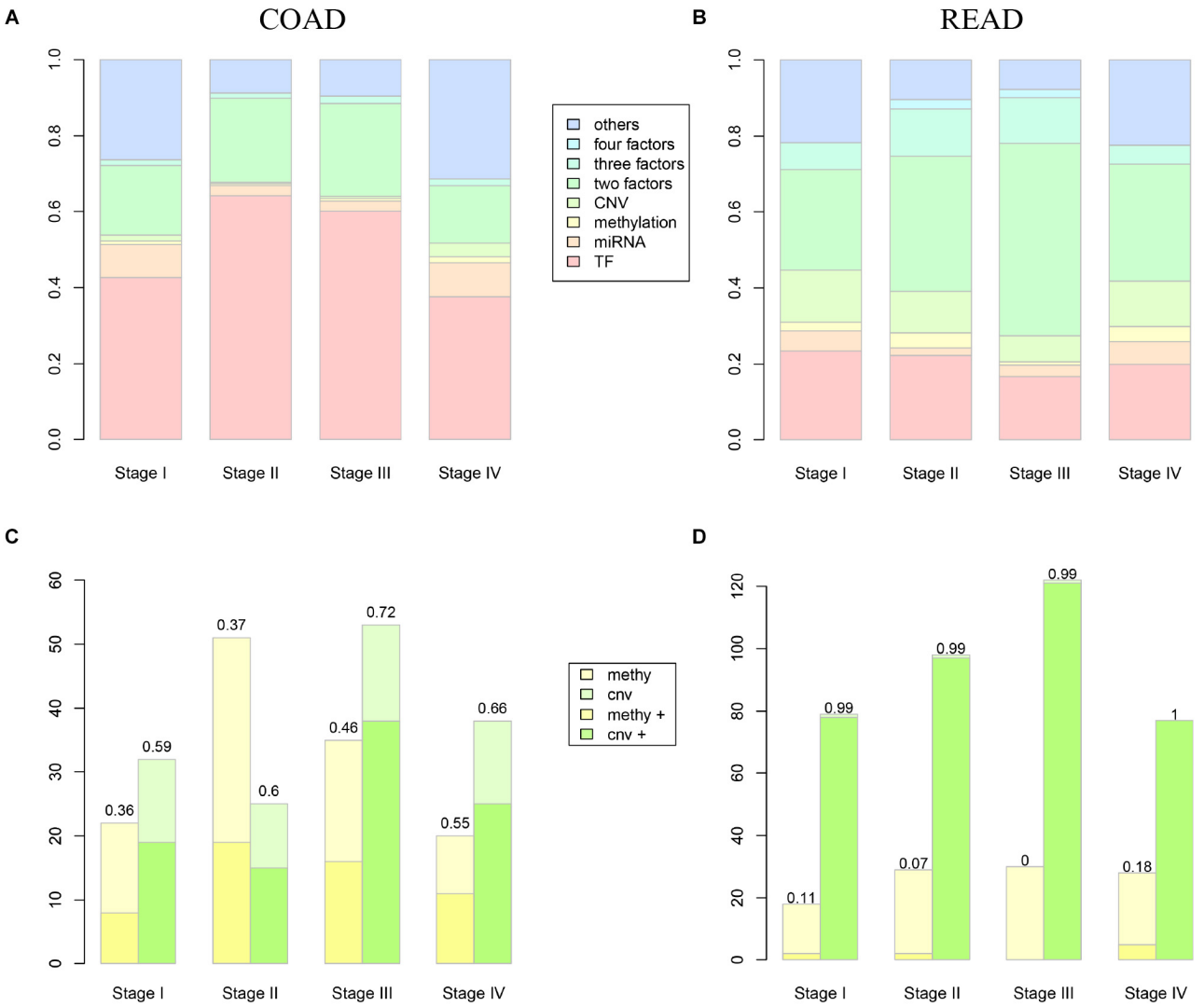


FIGURE 3 | The influence of different level factors. **(A,B)** Show the proportion of DEAGs regulated by CNVs, methylation, miRNAs, TFs, and composite factors in COAD and READ respectively. **(C,D)** Show the influence of methylation and CNV to DEAGs in COAD and READ respectively. Yellow boxes represent the number of autophagy genes influenced by methylation, while green boxes represent the number of autophagy genes influenced by CNVs. The lighter color represents negative regulation, and the deeper color represents positive regulation.

data could be obtained. This study primarily analyzed the impacts of CNVs, gene promoter methylations, miRNAs, and TFs on the expression of DEAG. To investigate the extent of influence, the percentage of DEAG regulated by each factor and the combination of multiple factors was calculated (**Figures 3A,B**, the detailed percentage of different factors for each stage was added to **Supplementary Table S3**). The majority of DEAGs are regulated by TFs. This is possibly due to a large amount of TFs present in the cells. The next major factor is miRNA, which negatively influenced these genes. A small number of DEAGs were affected by their own promoter CNVs and gene promoter methylations. Furthermore, a certain proportion of the DEAGs was subjected to a comprehensive regulation of multiple factors. Around 20% are jointly regulated by two factors. So, the DEAGs regulated by any two factors were thoroughly investigated (**Figure 4** and **Supplementary Figure S4**). There are no doubts that TF and miRNA synergistically influenced a large portion of the DEAGs, which may be a result of their relatively large quantity. The influence of CNVs and gene promoter methylations cannot be ignored. Hypermethylation of gene promoter generally has a negative influence on genetic expression, and the CNVs generally has a positive influence (**Figures 3C,D**). This pattern is consistent with the pre-transcriptional regulation of the gene.

Prognostic Value of the Biomarker for Survival Prediction

To further validate the prognostic performance, a univariate Cox proportional hazards regression model was used to evaluate the

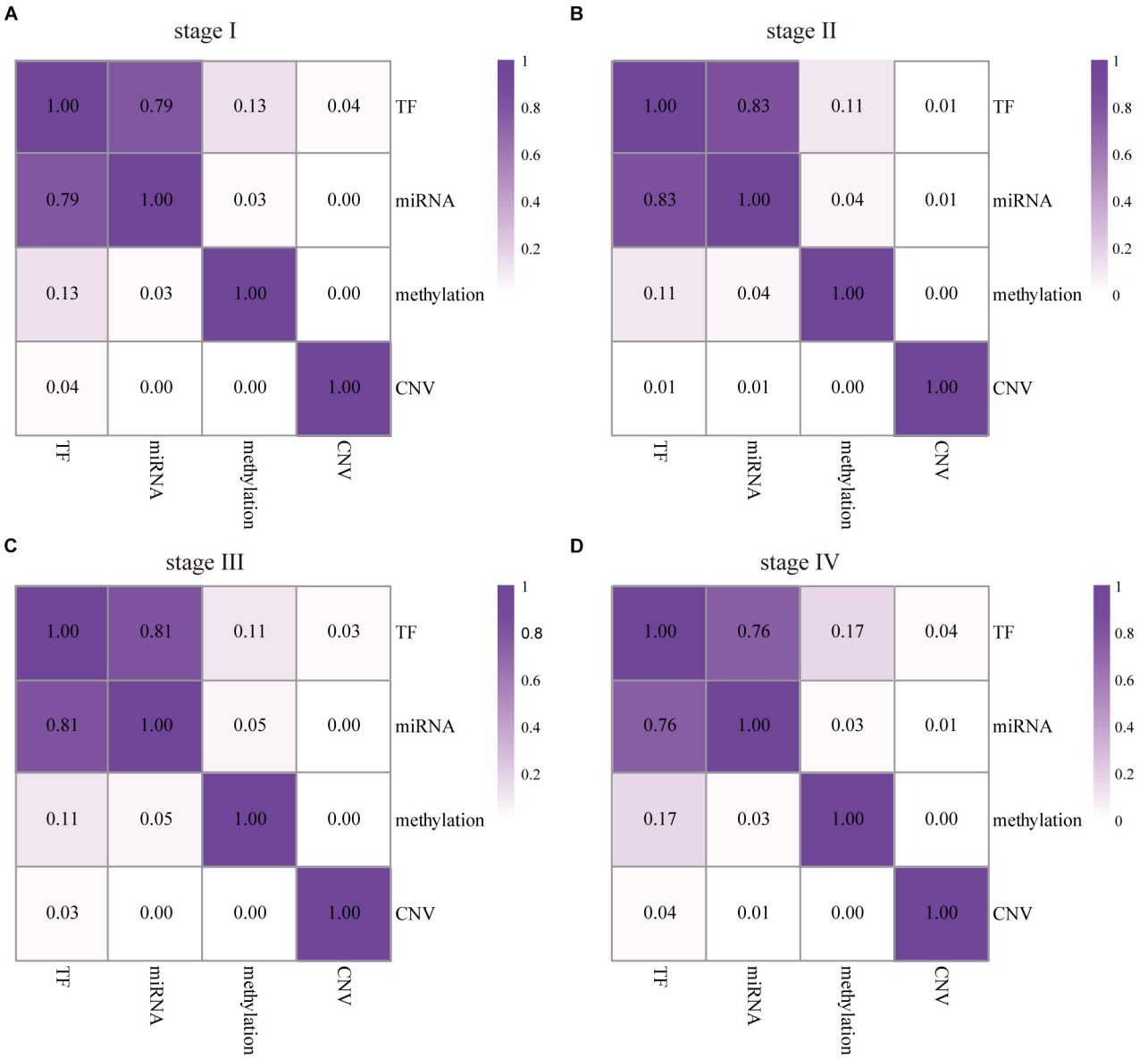


FIGURE 4 | The influence of any two factors to DEAGs in the four stages of COAD. The numbers represent the proportion of the DEAGs regulated by each pair of upstream regulators in stage I **(A)**, II **(B)**, III **(C)**, and IV **(D)**. The darker color represents a larger effect.

association between the DEAGs expression levels and overall survival (OS). The 281 COAD patients were divided randomly into a train dataset (n = 140) and an internal test dataset (141). It was found that nine genes were significantly associated with OS in the train dataset ($P < 0.01$). Using the regression coefficients estimated in the multivariate Cox regression analysis as weights, the risk score for each patient in the train dataset was calculated by a linear combination of the expression levels of the nine-gene. These scores were classified into high-risk group (n = 70) and low-risk group (n = 70) with the median risk score as the cutoff point. The result showed that patients in the high-risk group exhibited poor OS compared with those in the low-risk group (log rank $P < 0.05$) (**Figures 5A,D**). A time-dependent ROC curves analysis performed on the nine-gene and the area under curve (AUC) was achieved at 0.924 (**Figure 5G**). These genes can effectively stratify patients into different risk groups, which suggests that they may play essential roles in COAD. Internal test datasets were used to evaluate the prognostic value of the nine-gene signatures in predicting survival (**Figures 5B,E,H**) and a GEO dataset (**Figures 5C,F,I**). Patients of the internal test dataset and GEO dataset were divided into high-risk group and low-risk group with accordance to the same nine-gene signature score model derived from the train dataset. As in the train dataset, OS of high-risk group was significantly worse than that of the low-risk group (log rank $P < 0.01$). These results demonstrated that the nine

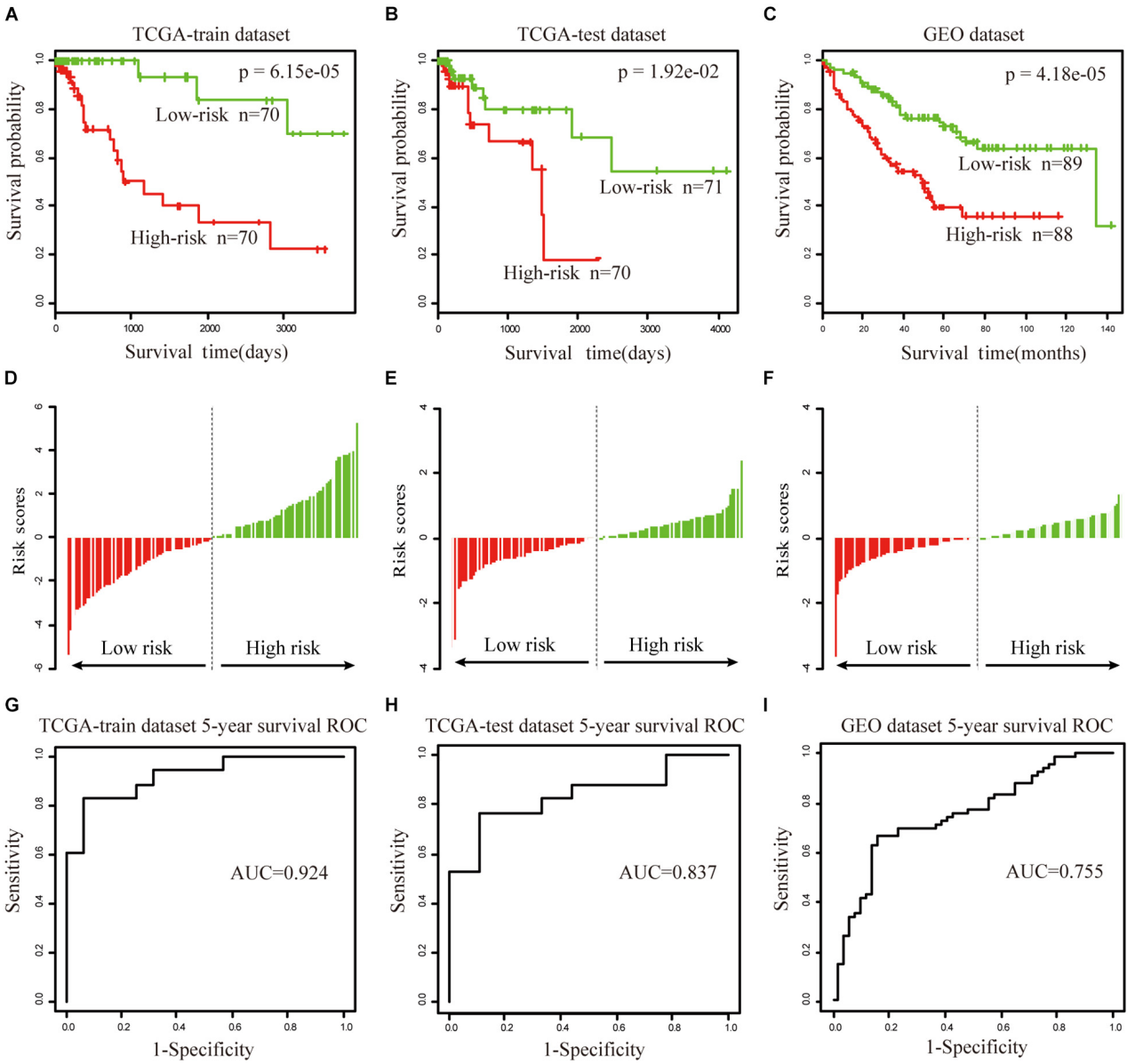


FIGURE 5 | The prognostic value of the nine genes signature in COAD. **(A–C)** Show the Kaplan–Meier survival curves for the train **(A)**, test **(B)**, and GEO datasets **(C)**. The red and green lines represent the high-risk and low-risk patients respectively. *P*-value means log rank *P*. **(D–F)** Show the detailed risk score distribution of patients in the train **(D)**, test **(E)**, and GEO datasets **(F)**. **(G–I)** Show the ROC curves and AUCs of the nine gene signature predicting patients' five-year survival in the train **(G)**, test **(H)**, and GEO datasets **(I)**.

genes were potential prognostic biomarkers for the prediction of tumor risk in COAD.

The univariate and multivariate analysis indicated that the nine-gene module biomarker was significantly associated with the OS of the COAD patients in the train and internal test dataset (**Table 1**). Additionally, the multivariate analysis also demonstrated that the designation of high-risk and low-risk groups remained statistically significant in the independent GEO dataset. In conclusion, these analyses demonstrated the capacity of the nine-gene biomarkers for COAD, and its ability to add value in the prognostic setting. This process was then systematically executed on the study of READ (**Supplementary Figure S5**), there are fifteen-gene biomarkers for READ.

Molecular Signatures of Prognostic Biomarkers

To investigate the clinical implications of the molecular signatures, we focused on the nine genes of COAD. There are six therapeutic targets of FDA-approved drugs through their associated TFs. SLC25A1 maintains mitochondrial integrity and bioenergetics in tumor cells. It prevents mitochondrial

TABLE 1 | Univariate and multivariate Cox regression analysis in the COAD.

Variables	Univariate analysis			Multivariate analysis		
	HR	95% CI	*P*-value	HR	95% CI	*P*-value
Train dataset						
Nine genes	2.795	1.894–4.125	2.258e-07	2.394	1.617–3.546	1.32e-05
Stage						
I,II	1(reference)			1(reference)		
III,IV	2.307	0.880–6.052	0.0892	2.639	0.742–9.390	0.134
Age	1.009	0.974–1.045	0.624	1.023	0.985–1.062	0.235
Tumor weight	1.004	1.001–1.006	0.005	1.004	1.000–1.006	0.028
Test dataset						
Nine genes	2.718	1.389–5.321	0.004	2.532	1.179–5.437	0.0172
Stage						
I,II	1(reference)			1(reference)		
III,IV	2.753	1.094–6.928	0.032	4.002	1.435–11.16	0.008
Age	1.015	0.981–1.050	0.396	1.021	0.984–1.06	0.264
Tumor weight	1.002	1.000–1.003	0.022	1.001	0.999–1.003	0.1001
GEO dataset						
Nine genes	2.718	1.774–4.165	4.384e-06	2.578	1.65–4.03	3.19e-05
Stage						
I,II	1(reference)			1(reference)		
III,IV	4.2199	2.387–7.459		4.065	2.285–7.234	1.85e-06
Age	1.006	0.988–1.025	0.492	1.017	0.999–1.037	0.0612

damage and circumvents mitochondrial depletion via autophagy, hence promoting proliferation (Catalina-Rodriguez et al., 2012). Several evidences implicate that SLC25A1 plays a role in cancer progression. High levels of SLC25A1 expression are associated with poor prognosis in lung cancer and estrogen receptor negative breast cancer (Georgiades et al., 1988; Jiang et al., 2017). In ovarian cancer patients, SLC25A1 mRNA levels are also associated with resistance to platinum-based chemotherapy, and blocking CTP function enhances sensitivity of cultured ovarian carcinoma cells to platinum (Georgiades et al., 1988; Jiang et al., 2017).

There were 41 FDA-approved drugs related to the six therapeutic targets, and they were connected by four TFs (**Figure 6A**). Ethanolamine derivatives of eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) have recently been found to induce autophagy by activating PPARG in human breast cancer cells (Rovito et al., 2015; Garay-Lugo et al., 2016). The PPARG gene is related to malignancy, which plays a vital role in the pathogenesis of multiple cancers in some clinical studies and animal models (Wang et al., 2015). AR plays a negative role in regulating the autophagy induced by celastrol, and it inhibits autophagy by transactivating mir-101 in prostate cancer cells (Guo et al., 2015). ESR1 is essential for sexual development as well as reproductive function and is involved in inducing autophagy of toxins (Chen and Xia, 2014; Tan et al., 2016).

These drugs can perform three types of treatment, including targeted therapy, hormone therapy, and chemotherapy. The Current studies have shown that the exposure to PT and balsalazide effectively inhibited the proliferation of human colon cancer HCT116 cells via inhibiting NF-κB activity and inducing apoptotic cell death. These suggest that the simultaneous administration of PT and balsalazide may provide a novel option for the treatment of colon cancer (Kim et al., 2015). The published evidence indicates that sulfasalazine prevents the development of dysplasia and CRC in patients with IBD (Eaden, 2003).

The distribution of four regulators was investigated on the six therapeutic targets in distinct stages (**Figure 6A**). It was found that each factor had stronger effects at different stages of cancer. In principle, their effects are present in stage II, III and IV, but there was almost no effect in stage I. A comprehensive network was constructed by integrating the upstream regulators and downstream regulated genes of the nine genes of COAD (**Figure 6B** and **Supplementary Table S1**). In final, we found that RBP play very important regulatory factors regulated autophagy-mediated CRC cell death in DEAG regulatory network (**Supplementary Table S2**). For example, DDX17 RNA-binding protein that regulated autophagy genes SLC25A1 and TRMT112 in the COAD is also important for the autophagy regulatory network. Similarly, POLR3A RNA-binding protein that acted as TFs performed the task of regulating autophagy genes in the READ (**Supplementary Table S2**). Prognosis biomarkers show the different regulation modes. It was found that more than half of the genes were regulated by TF, which existed in several stages. Then this process was systematically executed on the study of READ (**Supplementary Figure S6** and **Supplementary Table S1**).

DISCUSSION

Autophagy is associated with both tumorigenic and tumor progression in CRC (Lin et al., 2014). However, the clinical

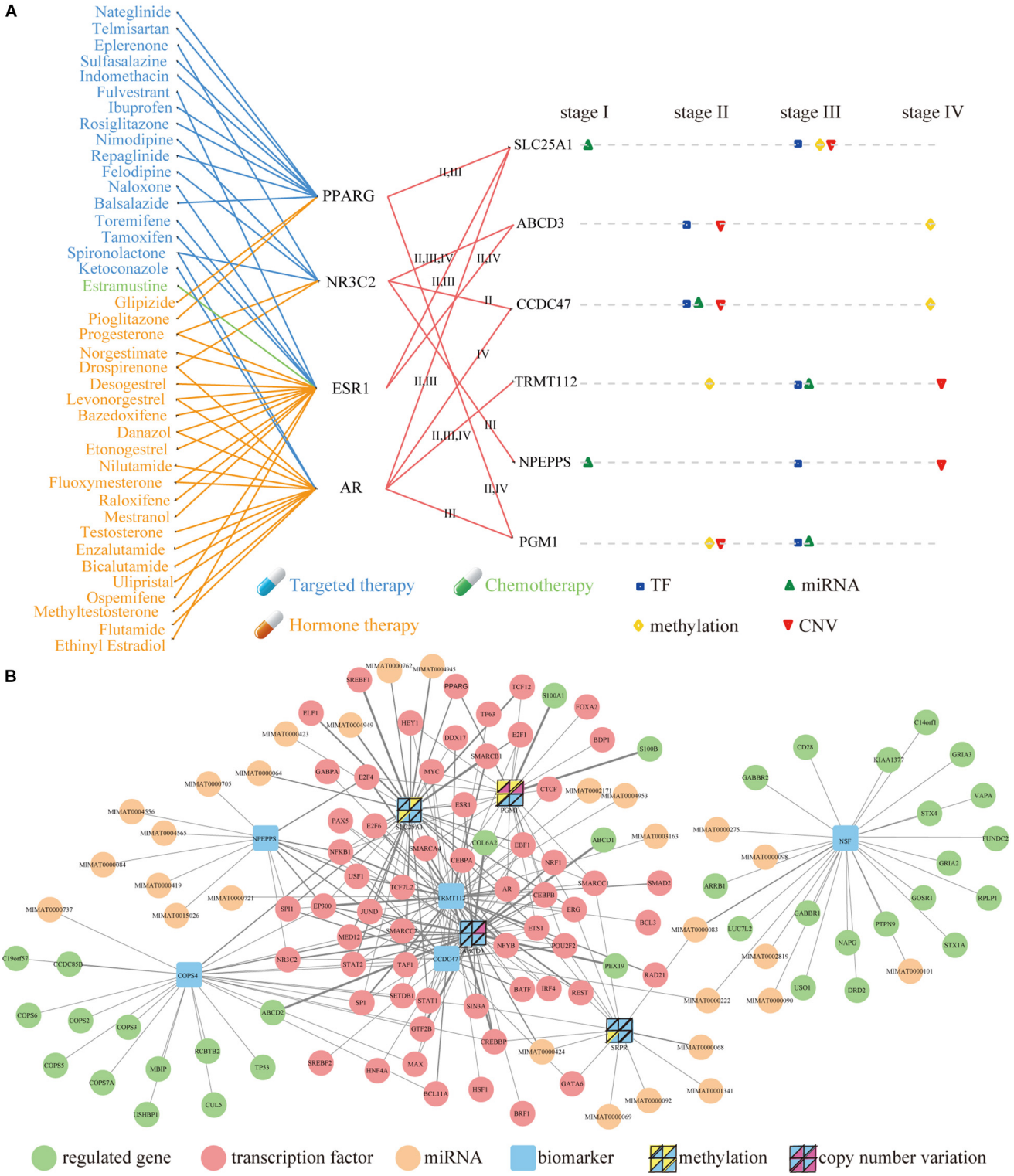


FIGURE 6 | Molecular Signatures and upstream and downstream network of nine biomarkers in COAD. **(A)** The mapping between FDA-approved drugs and their related genes (left) and the influence of four factors to the nine genes (right). The blue, green and orange lines represent targeted therapy, chemotherapy and hormone therapy respectively. Square, top triangle, diamond and bottom triangle represent the regulation of TF, miRNA, methylation and CNV on related genes in four stages. **(B)** Upstream and downstream network. The thickness of the line represented the quantity of interactions in these stages. The yellow and purple triangles (inside the four panel squares) represent that the nine genes are affected by their own methylation and CNV. Each panel in the four-panel squares represents each of the four cancer stages (top left is stage I, top right is stage II, bottom left is stage III, and bottom right is stage IV).

significance and autophagy function in CRC remains unclear. In this study, we have revealed the expression signatures of autophagy genes regulated by multiple factors, which include TF, miRNA, promoter methylation, and CNV. Some studies have demonstrated that RNA binding proteins as TFs play a key role in the development and function of CRC (New et al., 2019). RNA binding proteins regulate the expression of thousands of transcripts and are crucial for the regulation of CRC cellular processes, such as RNA splicing, modifications, transport, and translation (Kudinov et al., 2017; Chatterji and Rustgi, 2018). For instance, Zhou B et al. found that APOBEC3G, EEF1A2, EIF5AL1, and CELF3 as RNA binding proteins may provide a good prospect for the clinical diagnosis and treatment of patients with CRC metastasis (Zhou and Guo, 2018). As other examples, PTBP1 RNA binding protein that associated with tumor metastasis in CRC tissues directly interacts with autophagy gene ATG10 and regulates ATG10 expression level (Zhou and Guo, 2018). Therefore, research on the regulation of autophagy to improve clinical outcomes is becoming increasingly important. In conclusion, some novel prognostic biomarkers associated with autophagy in CRC should be further investigated in the future.

Autophagy genes are the key components of the autophagy-mediated regulatory network. They are implicated in the occurrence and development of CRC (Hao et al., 2017). We have systematically validated the autophagy genes of differential expression, through data comparison of diverse stages for CRC. Our findings were consistent with previous reports that the signature of autophagy genes changes with different expression variation in the progression of CRC. We further analyzed the potential functional implication of autophagy genes that were specifically expressed in various periods and found that the enriched biological processes and pathways of these genes play essential roles in diverse stages of CRC. Even more, our results showed that multiple factors that regulate DEAGs are significantly different. The TF and miRNAs that regulate the autophagy genes had a very low overlap in various stages of CRC. Therefore, the modulation of autophagy genes as potential prognostic biomarkers in CRC should be further researched.

To identify potential prognostic biomarkers in CRC, we evaluated the associations between expression levels of DEAGs and the survival of the patient by employing the Cox regression analysis (Lossos et al., 2004). Multiple evidences show that SLC25A1 overexpression is associated with poor prognosis of lung cancer and estrogen receptor-negative breast cancer (Georgiades et al., 1988). These genes have a strong prognostic ability and are independent of clinical factors. As significant prognostic factors in four stages of CRC, the signature of autophagy genes will have important effects on cancer-related biological processes. However, this observation should be interpreted with caution, because there are many uncertainties in the upstream regulatory factors of autophagy. Alterations in various molecular levels could cause expression dysregulation of autophagy genes (Hao et al., 2017). Therefore, further efforts are required to elucidate the corresponding contributions of various factors in the expression signatures of the autophagy gene of CRC (Wang et al., 2016). Also, it is essential that we continue to explore the biological functions of autophagy in the context of different interactions.

In summary, we identified the potential prognostic biomarkers in CRC and described their signatures in several stages of CRC. Along with the development of cancer clinical management approaches, this study will make a significant step toward transforming them from preclinical to clinical assessments.

DATA AVAILABILITY STATEMENT

Gene and miRNA expression data, methylation data, and the clinical data of CRC patients were downloaded from TCGA (http://cancergenome.nih.gov/) (Cancer Genome Atlas Research Network, 2008). Additionally, a cohort of 177 COAD patients and 196 READ patients from GEO database (GSE17536 and GSE87211) was used as an independent external test set.

AUTHOR CONTRIBUTIONS

HW, KL, and JL conceived and designed the experiments, and wrote the manuscript. CZ, JJ, LW, LZ, JX, XQ, and HH collected and analyzed the data. All authors read and approved the final version of the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (31501075, 31301094, and 31900493), the Natural Science Foundation of Heilongjiang Province (B201302), and the Postdoctoral Foundation of Heilongjiang Province (LBH-Z18167).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00245/full#supplementary-material

REFERENCES

Arntzen, M., Bull, V. H., and Thiede, B. (2013). Cell death proteomics database: consolidating proteomics data on cell death. *J. Proteome Res.* 12, 2206–2213. doi: 10.1021/pr4000703

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607. doi: 10.1038/nature11003

Boglev, Y., Badrock, A. P., Trotter, A. J., Du, Q., Richardson, E. J., Parslow, A. C., et al. (2013). Autophagy induction is a Tor- and Tp53-independent cell survival response in a zebrafish model of disrupted ribosome biogenesis. *PLoS Genet.* 9:e1003279. doi: 10.1371/journal.pgen.1003279

Cancer Genome Atlas Research Network, (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455, 1061–1068. doi: 10.1038/nature07385

Catalina-Rodriguez, O., Kolukula, V. K., Tomita, Y., Preet, A., Palmieri, F., Wellstein, A., et al. (2012). The mitochondrial citrate transporter, CIC, is essential for mitochondrial homeostasis. *Oncotarget* 3, 1220–1235. doi: 10.18632/oncotarget.714

Chatterji, P., and Rustgi, A. K. (2018). RNA binding proteins in intestinal epithelial biology and colorectal cancer. *Trends Mol. Med.* 24, 490–506. doi: 10.1016/j.molmed.2018.03.008

Chen, Y., and Xia, R. G. (2014). Screening and functional microarray analysis of differentially expressed genes related to osteoporosis. *Genet. Mol. Res.* 13, 3228–3236. doi: 10.4238/2014.April.25.8

Chisanga, D., Keerthikumar, S., Pathan, M., Ariyaratne, D., Kalra, H., Boukouris, S., et al. (2016). Colorectal cancer atlas: an integrative resource for genomic and proteomic annotations from colorectal cancer cell lines and tissues. *Nucleic Acids Res.* 44, D969–D974. doi: 10.1093/nar/gkv1097

Chou, C. H., Shrestha, S., Yang, C. D., Chang, N. W., Lin, Y. L., Liao, K. W., et al. (2018). MiRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* 46, D296–D302. doi: 10.1093/nar/gkx1067

Eaden, J. (2003). Review article: the data supporting a role for aminosalicylates in the chemoprevention of colorectal cancer in patients with inflammatory bowel disease. *Aliment. Pharmacol. Ther.* 18(Suppl. 2), 15–21. doi: 10.1046/j.1365-2036.18.s2.3.x

Garay-Lugo, N., Dominguez-Lopez, A., Miliar Garcia, A., Aguilar Barrera, E., Gomez Lopez, M., Gomez Alcala, A., et al. (2016). n-3 Fatty acids modulate the mRNA expression of the Nlrp3 inflammasome and Mtor in the liver of rats fed with high-fat or high-fat/fructose diets. *Immunopharmacol. Immunotoxicol.* 38, 353–363. doi: 10.1080/08923973.2016.1208221

Georgiades, Y., Chiron, M., and Joumard, R. (1988). Establishment of atmospheric pollution standards for motor vehicles. *Sci. Total Environ.* 77, 215–230. doi: 10.1016/0048-9697(88)90057-5

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. *Nat. Rev. Genet.* 15, 829–845. doi: 10.1038/nrg3813

Guo, J., Huang, X., Wang, H., and Yang, H. (2015). Celastrol induces autophagy by targeting AR/miR-101 in prostate cancer cells. *PLoS One* 10:e0140745. doi: 10.1371/journal.pone.0140745

Hao, H., Xia, G., Wang, C., Zhong, F., Liu, L., and Zhang, D. (2017). miR-106a suppresses tumor cells death in colorectal cancer through targeting ATG7. *Med. Mol. Morphol.* 50, 76–85. doi: 10.1007/s00795-016-0150-7

Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., et al. (2012). GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res.* 22, 1760–1774. doi: 10.1101/gr.135350.111

Hu, Y., Gaedcke, J., Emons, G., Beissbarth, T., Grade, M., Jo, P., et al. (2018). Colorectal cancer susceptibility loci as predictive markers of rectal cancer prognosis after surgery. *Genes. Chromosomes Cancer* 57, 140–149. doi: 10.1002/gcc.22512

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923

Jiang, L., Boufersaoui, A., Yang, C., Ko, B., Rakheja, D., Guevara, G., et al. (2017). Quantitative metabolic flux analysis reveals an unconventional pathway of fatty acid synthesis in cancer cells deficient for the mitochondrial citrate transport protein. *Metab. Eng.* 43, 198–207. doi: 10.1016/j.ymben.2016.11.004

Katheder, N. S., Khezri, R., O'Farrell, F., Schultz, S. W., Jain, A., Schink, M. K. O., et al. (2017). Microenvironmental autophagy promotes tumour growth. *Nature* 541, 417–420. doi: 10.1038/nature20815

Keshava Prasad, T. S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., et al. (2009). Human Protein Reference Database–2009 update. *Nucleic Acids Res.* 37, D767–D772. doi: 10.1093/nar/gkn892

Kim, H., Kim, S., Park, Y., Liu, Y., Seo, S. Y., Kim, S. H., et al. (2015). Balsalazide potentiates parthenolide-mediated inhibition of nuclear factor- k B signaling in HCT116 human colorectal cancer cells. *Intest. Res.* 13, 233–241.

Kudinov, A. E., Karanicolas, J., Golemis, E. A., and Boumber, Y. (2017). Musashi RNA-binding proteins as cancer drivers and novel therapeutic targets. *Clin. Cancer Res.* 23, 2143–2153. doi: 10.1158/1078-0432.CCR-16-2728

Lin, G., Hill, D. K., Andrejeva, G., Boult, J. K. R., Troy, H., Fong, A. C. L. F. W. T., et al. (2014). Dichloroacetate induces autophagy in colorectal cancer cells and tumours. *Br. J. Cancer* 111, 375–385. doi: 10.1038/bjc.2014.281

Lossos, I. S., Czerwinski, D. K., Alizadeh, A. A., Wechser, M. A., Tibshirani, R., Botstein, D., et al. (2004). Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *N. Engl. J. Med.* 350, 1828–1837. doi: 10.1056/NEJMoa032520

Mei, Y., Ramanathan, A., Glover, K., Stanley, C., Sanishvili, R., Chakravarthy, S., et al. (2016). Conformational flexibility enables the function of a BECN1 region essential for starvation-mediated autophagy. *Biochemistry* 55, 1945–1958. doi: 10.1021/acs.biochem.5b01264

New, J., Subramaniam, D., Ramalingam, S., Enders, J., Sayed, A. A. A., Ponnurangam, S., et al. (2019). Pleotropic role of RNA binding protein CELF2 in autophagy induction. *Mol. Carcinog.* 58, 1400–1409. doi: 10.1002/mc.23023

Rovito, D., Giordano, C., Plastina, P., Barone, I., De Amicis, F., Mauro, L., et al. (2015). Omega-3 DHA- and EPA-dopamine conjugates induce PPARgamma-dependent breast cancer cell death through autophagy and apoptosis. *Biochim. Biophys. Acta* 1850, 2185–2195. doi: 10.1016/j.bbagen.2015.08.004

Roy, S., and Debnath, J. (2010). Autophagy and tumorigenesis. *Semin. Immunopathol.* 32, 383–396. doi: 10.1007/s00281-010-0213-0

Smith, J. J., Deane, N. G., Wu, F., Merchant, N. B., Zhang, B., Jiang, A., et al. (2010). Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology* 138, 958–968. doi: 10.1053/j.gastro.2009.11.005

Tan, Y. Q., Zhang, J., Du, G. F., Lu, R., Chen, G. Y., and Zhou, G. (2016). Altered autophagy-associated genes expression in T cells of oral lichen planus correlated with clinical features. *Mediators Inflamm.* 2016:4867368.

Thorburn, A., Thamm, D. H., and Gustafson, D. L. (2014). Autophagy and cancer therapy. *Mol. Pharmacol.* 85, 830–838. doi: 10.1124/mol.114.091850

Vlachos, I. S., Paraskevopoulou, M. D., Karagkouni, D., Georgakilas, G., Vergoulis, T., Kanellos, I., et al. (2015). DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Res.* 43, D153–D159. doi: 10.1093/nar/gku1215

Wang, H., Wang, Y., Qian, L., Wang, X., Gu, H., Dong, X., et al. (2016). RNF216 contributes to proliferation and migration of colorectal cancer via suppressing BECN1-dependent autophagy. *Oncotarget* 7, 51174–51183. doi: 10.18632/oncotarget.9433

Wang, Y., Chen, Y., Jiang, H., Tang, W., Kang, M., Liu, T., et al. (2015). Peroxisome proliferator-activated receptor gamma (PPARG) rs1801282 C>G polymorphism is associated with cancer susceptibility in asians: an updated meta-analysis. *Int. J. Clin. Exp. Med.* 8, 12661–12673.

Williams, F. P., Haubrich, K., Perez-Borrajero, C., and Hennig, J. (2019). Emerging RNA-binding roles in the TRIM family of ubiquitin ligases. *Biol. Chem.* 400, 1443–1464. doi: 10.1515/hsz-2019-0158

Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., and Li, T. (2009). miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res.* 37, 851. doi: 10.1093/nar/gkn851

Yang, M., Zhao, H., Guo, L., Zhang, Q., Zhao, L., Bai, S., et al. (2015). Autophagy-based survival prognosis in human colorectal carcinoma. *Oncotarget* 6, 7084–7103.

Zhou, B., and Guo, R. (2018). Integrative analysis of significant RNA-binding proteins in colorectal cancer metastasis. *J. Cell. Biochem.* 119, 9730–9741. doi: 10.1002/jcb.27290

Zhou, K. R., Liu, S., Sun, W. J., Zheng, L. L., Zhou, H., Yang, J. H., et al. (2017). ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. *Nucleic Acids Res.* 45, D43–D50. doi: 10.1093/nar/gkw965

Zhou, M., Zhong, L., Xu, W., Sun, Y., Zhang, Z., Zhao, H., et al. (2016). Discovery of potential prognostic long non-coding RNA biomarkers for predicting the risk of tumor recurrence of breast cancer patients. *Sci. Rep.* 6:31038. doi: 10.1038/srep31038

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 26 March 2020
doi: 10.3389/fgene.2020.00265

Modulator-Dependent RBPs Changes Alternative Splicing Outcomes in Kidney Cancer

*Yang Wang[1,2], Steven X. Chen[1], Xi Rao[1] and Yunlong Liu[1]**

[1] Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, United States,
[2] State Key Laboratory of Biocatalysts and Enzyme Engineering, School of Life Sciences, Hubei University, Wuhan, China

Alternative splicing alterations can contribute to human disease. The ability of an RNA-binding protein to regulate alternative splicing outcomes can be modulated by a variety of genetic and epigenetic mechanisms. In this study, we use a computational framework to investigate the roles of certain genes, termed modulators, on changing RBPs' effect on splicing regulation. A total of 1,040,254 modulator-mediated RBP-splicing interactions were identified, including 137 RBPs, 4,309 splicing events and 2,905 modulator candidates from TCGA-KIRC RNA sequencing data. Modulators function categories were defined according to the correlation changes between RBPs expression and their targets splicing outcomes. QKI, as one of the RBPs influencing the most splicing events, attracted our attention in this study: 2,014 changing triplets were identified, including 1,101 modulators and 187 splicing events. Pathway enrichment analysis showed that QKI splicing targets were enriched in tight junction pathway, endocytosis and MAPK signaling pathways, all of which are highly associated with cancer development and progression. This is the first instance of a comprehensive study on how alternative splicing outcomes changes are associated with different expression level of certain proteins, even though they were regulated by the same RBP. Our work may provide a novel view on understanding alternative splicing mechanisms in kidney cancer.

Keywords: alternative splicing, RNA-binding protein, modulation, cancer, dysregulation

OPEN ACCESS

Edited by:
Yongsheng Kevin Li,
Harbin Medical University, China

Reviewed by:
Xiaowen Chen,
Harbin Medical University, China
Dario Balestra,
University of Ferrara, Italy
Weixin Xie,
The Ohio State University, United States

***Correspondence:**
Yunlong Liu
yunliu@iu.edu

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *20 January 2020*
Accepted: *05 March 2020*
Published: *26 March 2020*

Citation:
Wang Y, Chen SX, Rao X and Liu Y (2020) Modulator-Dependent RBPs Changes Alternative Splicing Outcomes in Kidney Cancer. Front. Genet. 11:265. doi: 10.3389/fgene.2020.00265

INTRODUCTION

Renal cell carcinoma (RCC) is a common malignancy, representing 4.2% of all new cancer cases, with about 73,820 new cases and 14,770 deaths estimated for 2019 in the United States (Siegel et al., 2019). RCC is radiotherapy- and chemotherapy-resistant, and surgery remains first-line therapy (Hsieh et al., 2017; Yin et al., 2019). Despite early surgical treatment, 30% of patients with a localized tumor eventually develop metastases, and 2 years survival rate of patients with metastatic kidney renal clear cell carcinoma (KIRC) is less than 20% (Mickisch, 2002; Janzen et al., 2003). Therefore, identification of underlying molecular mechanisms and metastatic potential of KIRC are essential for improvements in early diagnosis and treatment.

Dysregulation of alternative splicing (AS) is widely considered a new hallmark of cancer and its products are acknowledged as potentially useful biomarkers (Ladomery, 2013). Recent estimates indicate that nine out of every 10 human genes undergo AS in a cell type- or condition-specific manner to create distinct RNA transcripts from the same pre-mRNA molecule (Wang et al., 2008). The key role of AS is further confirmed by the linkage of splicing regulation to numerous human diseases, including neurological disorders and many types of cancer (Scotti and Swanson, 2016).

Regulation of AS is a complicated process in which numerous interacting components are at work, including *cis*-acting elements and *trans*-acting factors, complicated by the functional coupling between transcription and splicing (Wang et al., 2015). Corruption of the process may lead to disruption of normal cellular function and eventually disease. Thus, understanding the regulatory patterns that control AS events has the potential not only to give valuable molecular insights but also to provide solutions for various diseases.

AS events are largely controlled by RNA-binding proteins (RBPs) that recognize specific regulatory sequences embedded in the pre-mRNA transcripts (Gerstberger et al., 2014). However, splicing complexes are intricate molecular machines that process tens to hundreds of RNA target genes. At any given time, depending on the context and cellular stimuli, an RBP will affect only a subset of its RNA target genes. This specificity is often provided by a certain factors we named as "modulators," such as signaling proteins, microRNAs, lncRNAs that control RBPs activity through several different mechanisms, including: expression level (Payne et al., 2018), protein stability and turnover (Garcia-Maurino et al., 2017), nuclear/cytoplasmic localization (Di Liegro et al., 2014), altered protein interactions (Jankowsky and Harris, 2015), and co-transcriptional regulation (Shukla and Oberdoerffer, 2012). Modulators help a cell combine different external signals and make complex downstream decisions. Elucidating their function is necessary for understanding and controlling cell's response to external stimuli at transcriptional level.

With the increased availability of large data sets derived from high-throughput experiments and computer algorithms, investigating complex transcriptional dysregulation between RBPs and AS events in complex diseases is now possible. Recently, the ENCODE project published eCLIP data sets for 150 RBPs across K562 and HepG2 cell types (Van Nostrand et al., 2016; Yee et al., 2019). Technological advances have made it possible to define the comprehensive target networks of individual RBPs with high accuracy by integrating global splicing profiles upon depletion of each RBP and genome-wide maps of *in vivo*, direct protein-RNA interactions (Zhang et al., 2010; Weyn-Vanhentenryck et al., 2014).

In this study, we established a computational method for dissecting the relationship among RBPs, alternatives splicing events and a kind of proteins that may influenced the splicing regulation effect of RBPs. Our method is unique in its ability to discover how alternative splicing outcomes is changing when modulator expression is different, even though they were regulated by the same RBP. It is the first time in which a triplet describing the relationships among modulators, RBPs and the outcomes of their alternative splicing targets is reported. The triplet contains three objects: a specific RBP, a splicing target regulated by RBP, and a modulator candidate that may change splicing regulation of the RBP. This method was applied to RCC using TCGA-KIRC dataset to identify modulator-dependent RBPs and their targets splicing outcomes in kidney cancer. QKI, as one of the key RBPs in this study, has the greatest number of influenced splicing events. Functional enrichment analysis showed that the inferred QKI modulators were highly associated with regulation progress of some hallmark cancer genes, including ARMH4, LINC01268, PDP2, LAPTM4B, and CD7. The results showed that different expression of modulators was associated with the changing roles of RBPs on regulating their targets alternative splicing outcomes. We expect that such integrated analysis could reveal the roles of RBPs and provide novel insights into understanding alternative splicing mechanisms in kidney cancer.

MATERIALS AND METHODS

Identify Alternative Splicing Events and Gene Expression

Paired-end RNA sequencing data from 480 RCC patients was downloaded from The Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma (TCGA-KIRC). The percentage of inclusion (PSI) of spliced events were derived using Mixture of Isoforms (MISO) (Lee et al., 2013). A PSI value was computed for every identified event in each sample, and the original AS events were further processed to generate high-confidence events by retaining events with a PSI value greater than 0.1 in at least 100 samples from 480 (∼21% samples in total). Then, events that occurred in both the curated and TCGA datasets were retained to form the final set of AS events. In this study, we only focused on skipped exon (SE) alternative splicing events. We defined an altered skipped exon as any exon of a transcript excluding the first and the last exons. Finally, we only kept the events that at least 100 patients' have PSI value and coefficient of variation (CV) of PSI was larger than 0.1. Gene differential expression analysis was performed using edgeR (Robinson et al., 2010), and CPM was used to estimated gene expression.

Identify RBPs Targets Using eCLIP Data

We used crosslinking immunoprecipitation (CLIP) data for 150 RBPs profiled in eCLIP peaks (Van Nostrand et al., 2016) downloaded from ENCODE in bed format (Consortium Encode Project, 2012). The peaks in two immortalized human cell types, K562 and HepG2, were filtered by peak enrichment larger than 8 ($\log_2$FC $\geq$ 3) and $p < 10^{-5}$ as recommended (Van Nostrand et al., 2016). Since the agreement between peaks in two replicates was moderate (the median Jaccard distance 25 and 28% in K562 and HepG2, respectively), we took the union of peaks between the two replicates in both cell lines and then pooled the resulting peaks. We defined an RBP-binding splicing event as the region upstream 300 base pairs of the exon to downstream 300 base pairs of the exon.

Genomes and Transcript Annotations

February 2009 assembly of the human genome (hg19, GRCh37) was download from Genome Reference Consortium (Church et al., 2011). The respective transcript annotation v19 was download from GENCODE website (Harrow et al., 2012). Transcript annotations were parsed to extract positions of introns and exons.

Gene Function and Categories Analysis

Functional enrichment analysis was carried out via the hypergeometric test using the clusterProfiler R package. We used Human Gene Ontology annotation provided by Gene Ontology (GO) Consortium (Ashburner et al., 2000; The Gene Ontology Consortium, 2017). GO terms enrichment with adjusted $p < 0.01$ and KEGG pathway enrichment with adjusted $p < 0.05$. The gene types we discussed in this study including immune related genes, which were download from IMMPORT database[1] and TRRUST v2 database.[2] The lncRNA gene type annotation was based on biomaRt software suite in R.

Construction Modulator-RBP-Splicing Triplets in KIRC

The probabilistic model is similar to Li et al. (2017) as follows:

$$Y_{target} = \beta_0 + \beta_1 X_{rbp} + \beta_2 X_m + \beta_3 X_{rbp} X_m + \varepsilon \quad (1)$$

where, the X_{rbp}, X_m, and Y_{target} are the gene expression of RBP and its modulator, and the splicing outcomes of the affected target gene, respectively. X_{rbp} and X_m represent the effect of RBP and modulator, respectively, on target by themselves alone (main effects), while β_3 represents the effect of their interaction. If an RBP and modulator interaction influences target splicing outcomes, we expect β_3 to be non-zero.

We divide rank-ordered expression values of a gene by tertiles and further discretize the triplets using:

$$x' = \begin{cases} 1 & \text{if } x \text{ is in upper tertile} \\ NULL & \text{if } x \text{ is in middle tertile} \\ 0 & \text{if } x \text{ is in lower tertile} \end{cases} \quad (2)$$

Values are ranked and transformed by tertials as follows, and coefficients are estimated from the differences in observed proportions of frequencies:

(1) Splice events are ranked according to PSI.
(2) RBP is ranked by its expression.
(3) Each modulator is ranked by their expression.

After discretization, we only consider the eight bins, where none of the genes has "NULL" value, covering ∼33% of the samples. This simple strategy has been shown to maximize entropy among groups and the selection of significant triplets' method can be found in Babur et al. (2010).

Statistical Analysis and Software

The data were analyzed and visualized using R statistics software version 3.4.1 and ggplot2 package. Correlations were assessed using Pearson correlation test. Survival curves were generated by the Kaplan Meier method using the median H-score as the cutoff, and differences were analyzed with the log-rank test.

[1]https://www.immport.org
[2]https://www.grnpedia.org

RESULTS

Category of Modulator Action

We developed a framework to infer the modulators of RBPs whose expression strongly correlates with changes of a RBP's effect on regulating targets splicing outcomes. Here, the transcriptional activity of a RBP was evaluated by the Pearson correlation between the expression level of RBPs and its target splicing outcomes. A schematic diagram of work-flow is provided in **Figure 1**.

The proposed method takes five inputs: gene expression profiles, an RBP of interest, a list of modulator candidates, splicing profiles, and RBPs' binding information. Candidate modulators may include all genes satisfying the criteria. In addition, the expression of the modulator candidates and RBPs were required to be statistically independent. Each possible triplet was then independently tested using the PCCs (Pearson correlation coefficients) estimator, and by comparing $\Delta PCCs$ we defined the subtype of modulation categories. False positives were controlled using appropriate statistical thresholds. Three possible modes of modulator action were identified, depending on whether RBP-splicing correlation increased or decreased as a function of the modulator abundance.

Category of Modulator Action

For each triplet, three possible modes of modulator action were identified depending on whether RBP-splicing correlation increased or decreased as a function of the modulator abundance. The three models are "attenuates splicing," "enhances splicing," and "'inverts splicing." Among them, attenuates/enhances splicing modes including two sub-types: attenuates/enhances exon exclusion and attenuates/enhances exon inclusion; inverts splicing mode means that the mode of modulator switches from exon inclusion to exon exclusion or from exon exclusion to exon inclusion. These cases and details interpretations are listed in **Table 1**.

Identify Modulators of QKI in Kidney Cancer

We applied the proposed method for identifying modulators for 150 RBPs. 14,707 exons were selected from 42,485 annotated skipped exons that are derived using the gene structures of ENSEMBL database. We identified 1,040,254 significant modulator-mediated triplets from 40,623,520 potential modulator-RBP-splicing interactions at FDR $\leq$ 0.01 using TCGA-KIRC data. The potential interactions consisted of 137 RBPs, 4,309 splicing events, and 2,905 modulator candidates. Among these RBPs, 13 RBPs were filtered out as the PSI distribution among 480 patients did not meet our criteria (the PSI coefficient of variation should be larger than 0.1). RNA-binding protein Quaking (QKI) had the greatest number (68.9%, 199 out of 289) of modulated spliced exons.

We identified 2,014 Modulator-QKI-Splicing triplets. The triplets include 1,101 modulators and 187 splicing events corresponding to 130 genes. According to the correlation between QKI expression and its target splicing

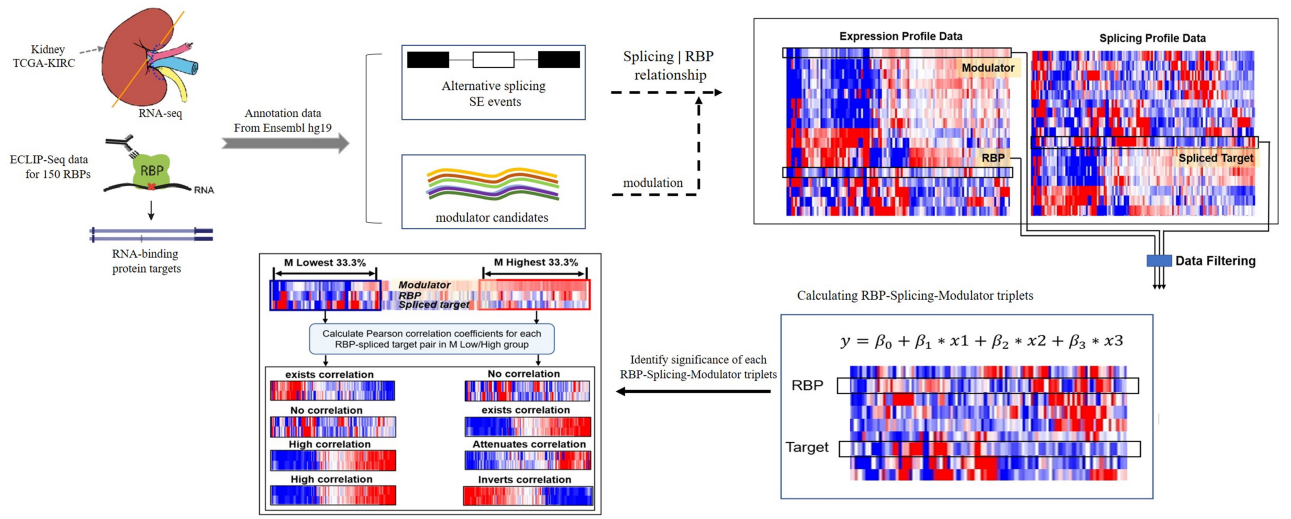


FIGURE 1 | A schematic diagram of workflow. Briefly, the dataset in this study derived from TCGA-KIRC and ENCODE ECLIP-Seq, hg19 was used as reference gene annotation from Ensemble. Each triplet contains three object: RBP, target and a modulator. Gene expression level is the input of RBP and modulator candidates, splicing outcomes (PSI value) is used to estimate the splicing level of target. Data filtering criteria as follows: (1) log2 (CPM) $\geq$ 1 (2) remove events with "NA" samples > 100 (3) CV(PSI) > 0.1. Then using the linear regression model to predict triplets. Only the triplets with significant β_3 *p*-value will be considered and selected to the following analysis. Finally, for each triplet, we group the samples into "low" and "high" groups based on the expression level of modulator (bottom/top 33% samples) in the specific triplet, and we compare the Pearson correlation coefficient values of RBP expression and target PSI value in two groups, identify the modulator function categories.

TABLE 1 | Categories of modulator mediated RBP regulations on splicing targets.

Modulation category	PCC_{low}	PCC_{high}	$\Delta PCCs$	Subtype mode
Attenuates splicing	− (EE)	− (EE)	$\lvert PCC_{low}\rvert > \lvert PCC_{high}\rvert$ or *p-value.high* > 0.05	Attenuates exon exclusion (AEE)
Enhances splicing	− (EE)	− (EE)	$\lvert PCC_{low}\rvert < \lvert PCC_{high}\rvert$ or *p-value.low* > 0.05	Enhances Exon exclusion (EEE)
Inverts splicing	+ (EI)	− (EE)		Exon inclusion to exclusion (ExonIE)
Inverts splicing	− (EE)	+ (EI)		Exon exclusion to inclusion (ExonEI)
Enhances splicing	+ (EI)	++(EI)	$\lvert PCC_{low}\rvert < \lvert PCC_{high}\rvert$ or *p-value. Low* > 0.05	Enhances EI (EEI)
Attenuates splicing	+ +(EI)	+(EI)	$\lvert PCC_{low}\rvert > \lvert PCC_{high}\rvert$ or *p-value.high* > 0.05	Attenuates EI (AEI)

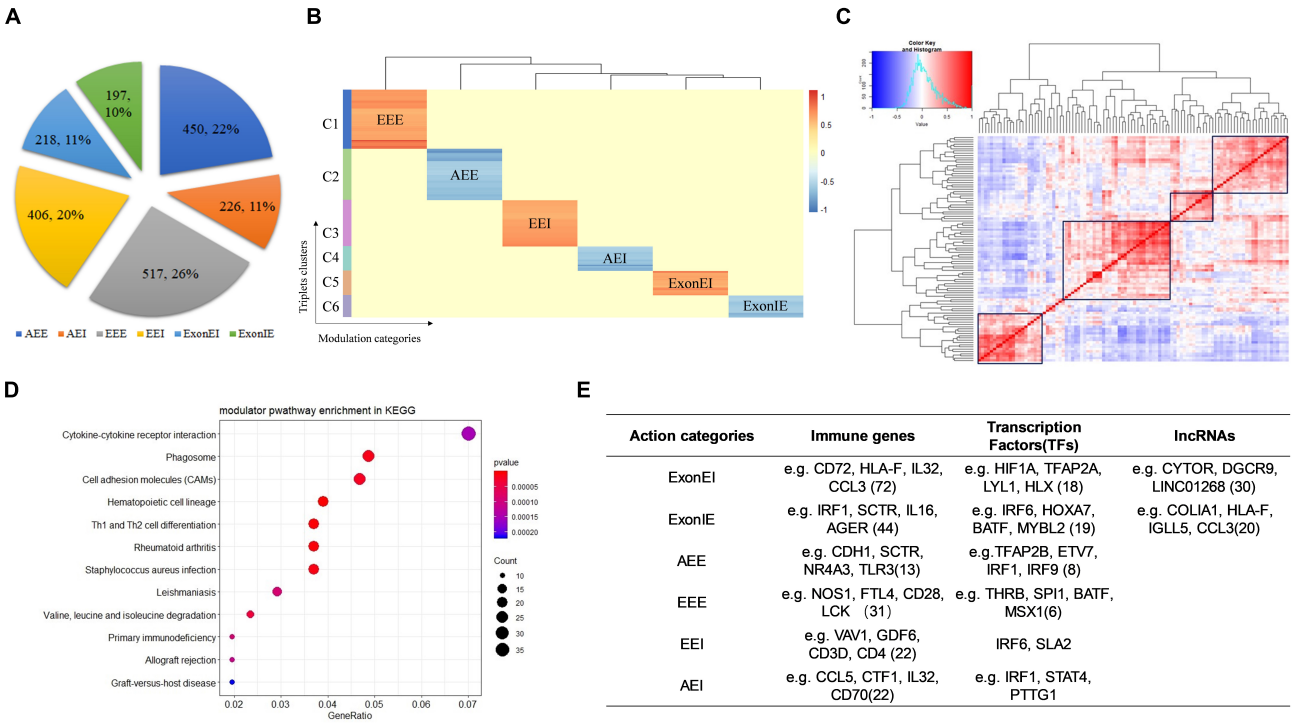
"+" and "−" signs in the columns indicate positive and negative values of Pearson correlation coefficient. "EE" represents exon exclusion, and "EI" means exon inclusion. The modulation categories of "attenuates splicing." "enhances splicing" or "inverts splicing" only refer to the roles of RBP on specific alternative spliced exon.

outcomes, six modulator sub-categories were identified, including 450 triplets in "Attenuates_Exon_Exclusion," 226 triplets in "Attenuates_Exon_Inclusion," 517 triplets in "Enhances_Exon_Exclusion," 406 triplets in "Enhances_Exon_Inclusion," 218 triplets in "Exon_Exclusion_to_Inclusion," and 197 triplets in "Exon_Inclusion_to_Exclusion" (**Figure 2A**).

Furthermore, we observed that most modulators affected multiple splicing targets were multimodal, and the same modulator may play opposite roles on different QKI targets. For example, ARMH4 inverts the splicing activity of QKI on its target CLTC: the inclusion of CLTC's spliced exon was correlated with increasing expression of QKI when ARMH4 is lowly expressed, while such association was inversed when ARMH4 is highly expressed. However, ARMH4 played enhanced exon inclusion role on QKI-STIM1 pair when its expression level changed from low to high. In another case, while the modulator KRT17 influenced 11 splicing targets of QKI, the role of KRT17 on these splicing outcomes changed among EEE, ExonIE, and ExonEI. The observation indicated that the distinct modulation patterns were triplets dependent rather than depending on specific RBPs or target genes. Our findings support this complexity in modulators typically had many target-specific effects. These findings suggested that more complex models are needed to better elucidate that how splicing is regulated.

Hence, we clustered modulators by their regulation patterns, yielding distinct groups of modulators that mediated splicing dysregulation in specific patterns (**Figure 2B**). For instance, the modulators in cluster 1–2 tend to reverse the QKI activity on regulating splicing targets' outcomes, whereas those modulators in cluster 4 and cluster 6 tend to enhance QKI splicing activity. In conclusion, these modulators may work as antagonistic or coactivators to mediate QKI splicing activity.

Among these modulator-mediated triplets, we noticed that many modulators regulate the same QKI splicing targets, this may be because some of the modulators co-express or play similar functions in related pathways. As the result showed in **Figure 2C**, modulators were grouped into several clusters



E

Action categories	Immune genes	Transcription Factors(TFs)	lncRNAs
ExonEI	e.g. CD72, HLA-F, IL32, CCL3 (72)	e.g. HIF1A, TFAP2A, LYL1, HLX (18)	e.g. CYTOR, DGCR9, LINC01268 (30)
ExonIE	e.g. IRF1, SCTR, IL16, AGER (44)	e.g. IRF6, HOXA7, BATF, MYBL2 (19)	e.g. COLIA1, HLA-F, IGLL5, CCL3(20)
AEE	e.g. CDH1, SCTR, NR4A3, TLR3(13)	e.g.TFAP2B, ETV7, IRF1, IRF9 (8)	
EEE	e.g. NOS1, FTL4, CD28, LCK (31)	e.g. THRB, SPI1, BATF, MSX1(6)	
EEI	e.g. VAV1, GDF6, CD3D, CD4 (22)	IRF6, SLA2	
AEI	e.g. CCL5, CTF1, IL32, CD70(22)	e.g. IRF1, STAT4, PTTG1	

FIGURE 2 | Identify modulators of QKI in KIRC. **(A)** Six mode sub-categories according to the correlation between QKI expression and its target splicing outcomes. The number in the pie chart means the percentage of each sub-categories. AEI indicates attenuates the exon inclusion, AEE indicates attenuates exon exclusion, EEI indicates enhances exon inclusion, EEE indicates enhances exon exclusion, ExonIE indicates reverses exon inclusion to exon exclusion, ExonEI indicates reverse exon exclusion to inclusion. **(B)** Modulators clusters by their regulation patterns. Six clusters were grouped according to modulator sub-categories. **(C)** Correlation heatmap of QKI modulators. The redder the color the higher correlation between two modulators. The values in the matrix were the normalized gene expression of modulators. **(D)** KEGG enrichment analysis of QKI modulators. **(E)** Modulators categories according to gene biotypes and features, including immune genes, transcription factors and lncRNAs.

according to their expression's correlation. This may be a potential reason why the spliced outcome of same target could be influenced by many modulators. In addition, the pathway enrichment results shown that these modulators were highly enriched in categories which were known to be associated with cancer development and progression, including cytokine-cytokine receptor interaction, Th1 and Th2 cell differentiation, and cell adhesion molecules (**Figure 2D**).

Furthermore, we classified the modulators of QKI according to gene biotypes and features, including immune related genes, transcription factors and lncRNAs (**Figure 2E**). The types of categories provide a framework for understanding many types of dysregulation on splicing.

Functional Analysis of QKI Modulators

To confirm the QKI-splicing-modulator triplet signatures as independent predictors, we selected six inferred modulator-influenced triplets to compare the association among QKI-splicing-modulators. The modulators we focused on were obtained from the analysis result 3.3, including immune genes (CCL3, HLA-F, AGER), transcription factors (ARMH4, STAT4), and lncRNAs (LINC01268).

As an example, immune gene AGER as a modulator of QKI, who shown differentially expressed level between cohort and normal samples in KIRC, played inverts exon exclusion role on regulating the splicing outcomes of GABRE. Comparing the two patterns in different groups, when the expression of modulator AGER is low, the PCCs between QKI expression and GABRE splicing level (PSI) is −0.1, while such correlation inverts to 0.32 in another group whose AGER expression high. Similar pattern we found that the correlation between QKI and its splicing target STIM1 was lost from 0.45 to no significant correlation when the modulator immune gene HLA-F expression differentially in two groups. Meanwhile, LINC01268 as modulator played attenuated effect on regulating QKI splicing activity. The correlation between QKI and target CTNND1 was −0.54 in LINC01268 expressed low group, while such correlation gone when LINC01268 expression becomes high (**Figure 3**).

To investigate the association between dysregulated target splicing outcomes and kidney cancer, we performed survival analysis and compared the expression level of these modulators in kidney tumor samples and normal samples based on TCGA-KIRC dataset. Results shown that most of the modulators were differentially expressed between tumor and normal samples and overall survival associated. Although KRT17 as one of the modulators we inferred did not show too much differentially expressed, the clinical information obtained from

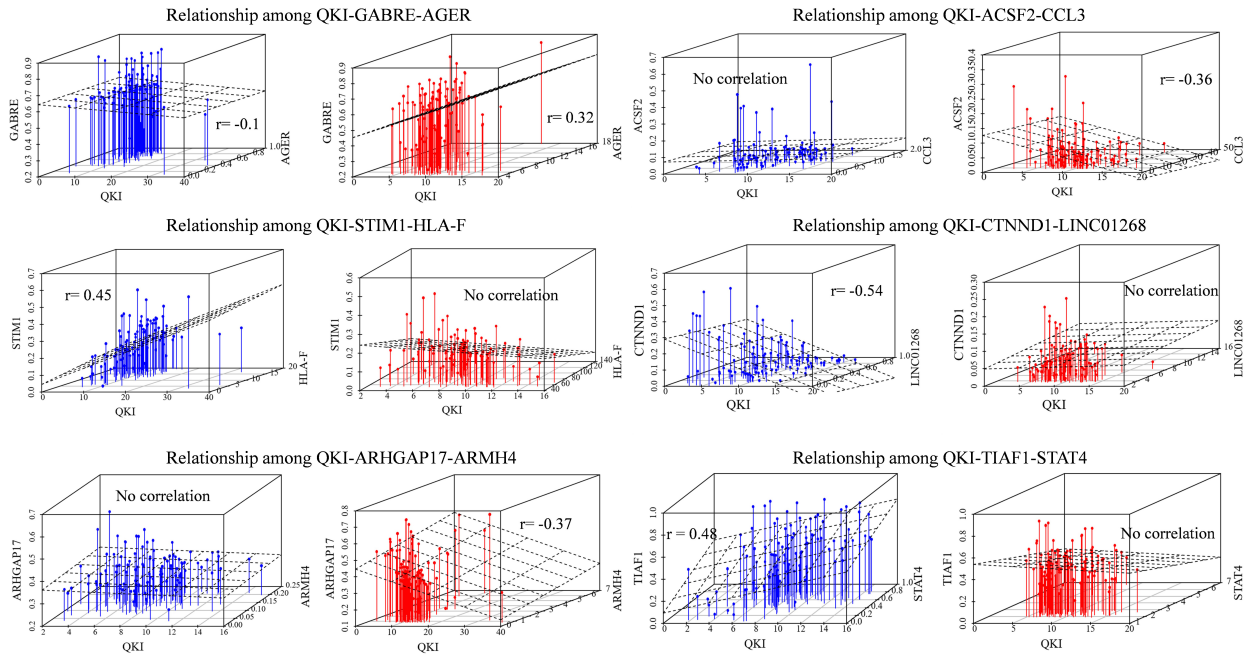


FIGURE 3 | Relationships among QKI-modulator-target. The blue means the samples in modulator expression low group (expression in bottom 33%), the red means the samples in modulator expression high group (expression in top 33%). The correlation value is the Pearson correlation between expression level of QKI and the splicing outcomes of its targets. No correlation means the statistical *p*-value of correlation is not significant (*p*-value cutoff setup as 0.01).

TCGA indicated that gene expression in kidney cancer was significantly associated with overall survival outcomes. The gene expression levels and survival analysis of top 10 modulators who has the most influenced targets of QKI were compared (**Figures 4A,B**).

KEGG enrichment analysis revealed that these target genes were enriched in categories known to be related to cancer development and progression (**Figure 4C**), such as tight junction pathway, transcriptional mis-regulation in cancer, endocytosis, and MAPK signaling pathway. The top enriched GO terms of these influenced target genes were associated with transcriptional regulation progress, including RNA splicing, cell growth and protein binding (**Figures 4D–F**). The results were reasonable as QKI regulated its target mainly on splicing level, once the expression of QKI was perturbed by the modulators, the roles of QKI on its targets, including binding, splicing, cellular development and transcriptional regulation would be influenced consequently.

In addition, cancer-relevant modulators were identified though tumor associated gene list from the Network of Cancer Genes (NCG, v6.0) database (Repana et al., 2019) and Tumor suppressor gene database (TSGene v2.0) (Zhao et al., 2016), separately (**Figure 4G**). The 2,372 tumor diver genes obtained from NCG including 711 known cancer genes and 1,661 candidate cancer genes. Among them, 149 genes overlapped with tumor diver genes, almost reaching 13% (149/1,179) of total numbers of modulators we inferred in this study. Meanwhile, approximately 7% (77/1,179) modulators were tumor suppressor genes, and the gene type was protein coding gene.

Analysis the Splicing Outcomes of CTNND1 Influenced by Modulators in Kidney Cancer

In this study, we found that the spliced outcome of the 20th exon of CTNND1 has the most inferred modulators, including 30 lncRNAs and 80 protein coding RNAs. The corresponding AS event is "chr11:57582866: 57582972: + @ chr11: 57583387: 57583473: + @ chr11:57583769: 57586652:+." Previously study reported that CTNND1 encodes a member of the Armadillo protein family, which function in adhesion between cells and signal transduction (Zhu et al., 2012), multiple CTNND1 isoforms are expressed in cells via alternative splicing, only full-length CTNND1 promotes invasiveness (Yanagisawa et al., 2008). Two modulation categories were identified including 107 attenuates exon exclusion (AEE), 7 enhances exon exclusion (EEE). The alternative spliced exon of CTNND1 and some of it's in each category are shown in **Figure 5A**.

For example, TNFRSF14 as one of modulators of QKI attenuates the splicing regulation on the 20th exon exclusion of CTNND1. We found that, in TNFRSF14 expression low group, the correlation between QKI expression and CTNND1 PSI is −0.38 (p = 1.7e−05), while such correlation is lost when in TNFRSF14 expression high group (correlation = 0.001, p = 0.98). This indicated that high expression of modulator TNFRSF14 may play negative effect on changing the splicing activity of QKI. In addition, we found that LAPTM4B, as another modulator of QKI, played enhanced exon exclusion role on regulating the 20th exon splicing outcome when it expression is high (**Figure 5B**).

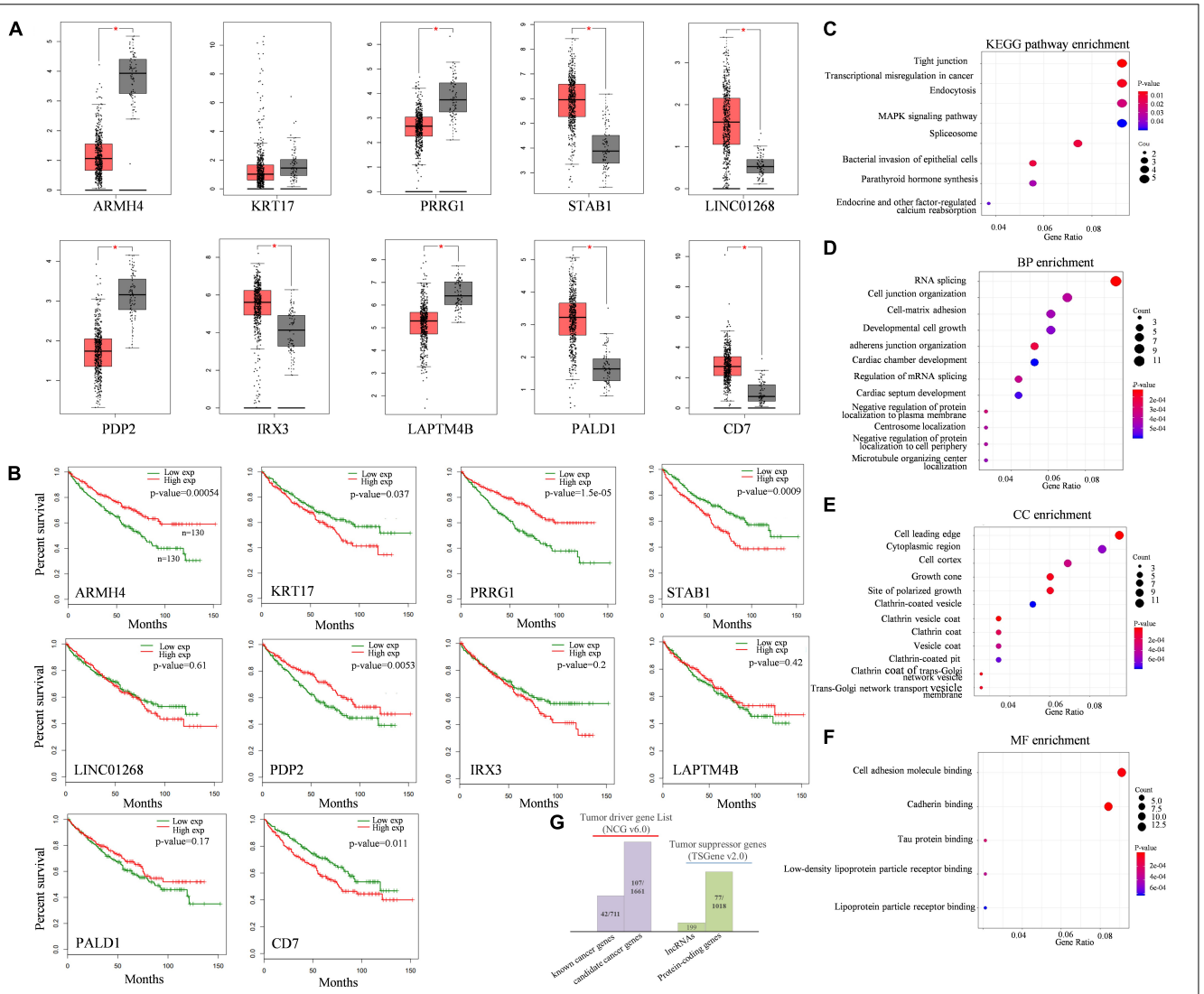


FIGURE 4 | Functional analysis of QKI modulators. **(A)** Expression levels of top10 modulators in KIRC tumor tissues (red boxes) compared with normal tissues (gray boxes**). (B)** survival analysis of top10 modulators. The red line means the modulator expression is high, the green line means the modulator is expression low. Modulator is grouped according to the median value of its expression level. **(C–F)** KEGG pathway and GO enrichment analysis, including biological process (BP), cellular component (CC) and molecular function (MF). **(G)** Cancer-relevant modulators identification according to Network of Cancer genes (NCG) and Tumor suppressor gene (TSGene) database. (*The top10 modulators were selected based on the number of their influenced splicing targets).

Thus, the results showed that differentially expressed modulators indeed changed the role of QKI on regulating CTNND1's splicing outcomes, and we believed that this kind of regulation may provide important insights for study dysregulation of splicing outcomes associated with various diseases.

DISCUSSION

Alternative splicing alterations may confer a selective advantage to the tumor, such as angiogenesis (Amin et al., 2011), proliferation (Bechara et al., 2013), cell invasion (Venables et al., 2013), and avoidance of apoptosis (Izquierdo et al., 2005). Some splicing mRNA isoforms could change the reading frame, resulting in the generation of different protein isoforms with diverse functions and/or localizations (Sutandy et al., 2018). One of the traditional methods to estimate the functions of mRNAs or protein is comparing the difference of gene expression level (Kim et al., 2014; Lorthongpanich et al., 2019; Xu et al., 2019).

However, not all detected alternative splicing events might necessarily result in mRNAs or proteins expression level changing. In addition, global description of alternative splicing networks and demonstration of their functional consequences have now emerged as one of the biggest challenges of the field (Baralle and Giudice, 2017). By integrating gene expression profile with splicing outcomes of alternative splicing events may be one of the possible ways to study the functional consequences for most of the identified splicing events.

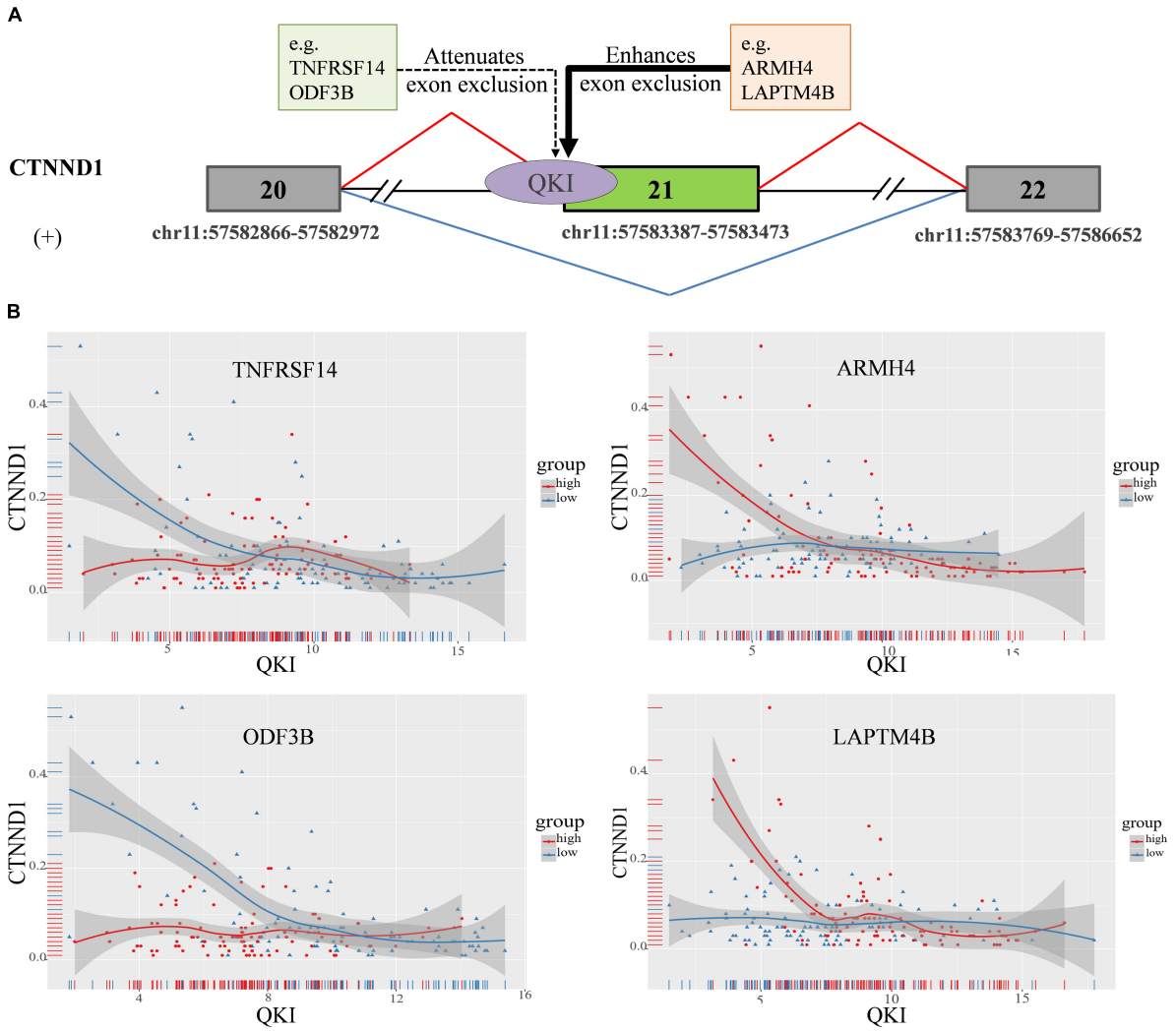


FIGURE 5 | CTNND1 splicing outcomes influenced by modulator expression in KIRC. **(A)** Examples of modulators in each mode sub-categories, including exon exclusion to inclusion (ExonEI), exon inclusion to exclusion (ExonIE), enhances exon exclusion (EEE) and attenuates exon exclusion (AEE). **(B)** Four modulators influence splicing outcomes of CTNND1, including TNFRSF14, ARMH4, LAPTM4B, and ODF3B. The red means the samples in modulator expression high group (top 33%), the blue means the samples in modulator expression low group (bottom 33%). The x-axis is the expression level of QKI, y-axis is splicing outcome (PSI value) of CTNND1.

In this study, we established a computational method for identification the modulators whose expression is associated with changing the targets splicing outcomes of RBPs in KIRC. Previously, several computational methods have been developed to identify modulators associated with transcription factors (TFs) regulation activity on expression level (Wang et al., 2009; Babur et al., 2010; Li et al., 2016, 2017), these studies discussed the transcriptional activities of TFs can be influenced by the expression level of modulators. Our method is unique in its ability to discover how alternative splicing outcomes is changing when modulator expression is different, even though they were regulated by the same RBP. And the method aimed at dissecting the effects of disruption in RBPs and hopefully it could provide insight into studying alternative splicing networks during development, cell differentiation, and in disease.

During tissue development and cell differentiation specific RBPs are finely regulated at their expression levels, localization, their own splicing, mRNA stability, and translation efficiency (Baralle and Giudice, 2017). RBPs bind to *cis*-elements promoting or inhibiting splice site recognition, hence RBP expression coordinates alternative splicing networks during development. We focused on 150 RBPs in this study, and only 137 RBPs remained in the final analysis due to there was no splicing targets of them within our criteria. Three possible modes of modulator actions were defined in this study, depending on the correlation changes between RBP and its target splicing outcomes when modulator expression is different. Among these RBPs, we found that QKI had the greatest number of influenced spliced exons (68.9%, 199 out of 289), and 2,014 Modulator-QKI-Splicing triplets were finally identified focused on QKI. Results showed

that most modulators affected multiple splicing targets were multimodal, and the same modulator may play opposite roles on different QKI targets.

For example, high expression level of modulator ADAM8 enhanced QKI role on regulating ACSF2 exon exclusion, while it enhanced target FMNL2 exon inclusion in regulating splicing outcomes. Another example, low expression level of AJM1 modulated QKI attenuated exon exclusion on regulating CTNND1 splicing outcomes, while such modulation role changed into enhanced exon inclusion when the target became ATF2. The detailed information about the modulators roles in the triples could be found in **Supplementary Table S1**.

Pathway enrichment results showed that all these influenced splicing target events of QKI were enrolled in cancer development and progression related pathways, including tight junction pathway, transcriptional mis-regulation in cancer, endocytosis, and MAPK signaling pathways. This evidence indicated that these alternative spliced events played crucial roles in kidney cancer, and changes the splicing outcomes of them may result in dysregulation in alternative splicing networks.

Among all these influenced splicing events, CTNND1 attracted more attention as the splicing outcomes of the 20th exon has the most inferred modulators. Previously study reported that CTNND1 was a tumor-driver gene, whose alternative splicing was related to cell invasion and metastasis (Yanagisawa et al., 2008). In addition, CTNND1 (p120) consists of central ARM domain flanked by the N-terminal regulatory (NTR) and C-terminal tail region (CTR) (Ishiyama et al., 2010), and the 20th exon of CTNND1 is in CTR region. Thus, different splicing outcomes of CTNND1 may influence the domain function, resulting in the generation of different protein isoforms with diverse functions.

We identified 114 inferred modulators of QKI-CTNND1 pair, including 30 lncRNAs and 80 protein-coding RNAs. Han et al. (2016) reported that MALAT1 may play as a tumor-suppressor gene in gliomas, and high MALAT1 expression linked to cell proliferation and metastasis. In our results, we noticed that high expression of modulator MALAT1 tended to attenuate QKI regulation role on splicing the 20th exon exclusion in CTNND1. The correlation between QKI expression and CTNND1 PSI was -0.51 (p = 4.3e$-$09) in MALAT1 expression low group, and such correlation changed into -0.20 (p = 0.02) in MALAT1 expression high group. LINC00174 as another inferred modulator had been reported that it exerted a tumorigenesis role in glioma. LINC00174 knockdown inhibited cell proliferation, migration, invasion and glycolysis (Shi et al., 2019). We found that, when LINC00174 expression is low, the correlation between QKI expression and splicing outcomes of CTNND1 is -0.37 (p = 2.6e$-$05), while such correlation was lost (r = -0.07, p = 0.44) when LINC00174 expression became high. These evidences showed that regulation of alternative splicing outcomes is a complex progress, it different splicing consequence not only associated with RBP but also associated with other proteins expression.

Although the established model in this study and the corresponding results appear helpful for understanding the alternative splicing regulation, there are some limitations. First, many proteins tended to show similar expression pattern, certain RBP-target pairs may have more than two inferred modulators, these results may contain certain false positive modulators. In addition, the function and mechanism of how modulator changes the RBPs regulation on their target splicing outcomes need to be further studied by experiments, for example, modulator co-expressed or physically interacted with RBPs, and this is a long way for us to go. Finally, we expect that our study could provide novel insights for understanding the dysregulation of alternative splicing in cancer.

DATA AVAILABILITY STATEMENT

Paired-end RNA sequencing data from 480 RCC patients was downloaded from the Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma (TCGA-KIRC).

AUTHOR CONTRIBUTIONS

YL and YW designed the study. YW and SC wrote the code, analyzed data, and wrote the manuscript. XR calculated the alternative splicing events in KIRC. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We would appreciate detailed suggestions from reviewers who significantly helped us improved the early version of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00265/full#supplementary-material

TABLE S1 | Inferred QKI modulators and triplets' information.

REFERENCES

Amin, E. M., Oltean, S., Hua, J., Gammons, M. V., Hamdollah-Zadeh, M., Welsh, G. I., et al. (2011). WT1 mutants reveal SRPK1 to be a downstream angiogenesis target by altering VEGF splicing. *Cancer Cell* 20, 768–780. doi: 10.1016/j.ccr.2011.10.016

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Babur, O., Demir, E., Gonen, M., Sander, C., and Dogrusoz, U. (2010). Discovering modulators of gene expression. *Nucleic Acids Res.* 38, 5648–5656. doi: 10.1093/nar/gkq287

Baralle, F. E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* 18, 437–451. doi: 10.1038/nrm.2017.27

Bechara, E. G., Sebestyen, E., Bernardis, I., Eyras, E., and Valcarcel, J. (2013). RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. *Mol. Cell.* 52, 720–733. doi: 10.1016/j.molcel.2013.11.010

Church, D. M., Schneider, V. A., Graves, T., Auger, K., Cunningham, F., Bouk, N., et al. (2011). Modernizing reference genome assemblies. *PLoS Biol.* 9:e1001091. doi: 10.1371/journal.pbio.1001091

Consortium Encode Project, (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi: 10.1038/nature11247

Di Liegro, C. M., Schiera, G., and Di Liegro, I. (2014). Regulation of mRNA transport, localization and translation in the nervous system of mammals (review). *Int. J. Mol. Med.* 33, 747–762. doi: 10.3892/ijmm.2014.1629

Garcia-Maurino, S. M., Rivero-Rodriguez, F., Velazquez-Cruz, A., Hernandez-Vellisca, M., Diaz-Quintana, A., De la Rosa, M. A., et al. (2017). RNA binding protein regulation and cross-talk in the control of AU-rich mRNA fate. *Front. Mol. Biosci.* 4:71. doi: 10.3389/fmolb.2017.00071

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. *Nat. Rev. Genet.* 15, 829–845. doi: 10.1038/nrg3813

Han, Y., Wu, Z., Wu, T., Huang, Y., Cheng, Z., Li, X., et al. (2016). Tumor-suppressive function of long noncoding RNA MALAT1 in glioma cells by downregulation of MMP2 and inactivation of ERK/MAPK signaling. *Cell Death Dis.* 7:e2123. doi: 10.1038/cddis.2015.407

Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 22, 1760–1774. doi: 10.1101/gr.135350.111

Hsieh, J. J., Purdue, M. P., Signoretti, S., Swanton, C., Albiges, L., Schmidinger, M., et al. (2017). Renal cell carcinoma. *Nat. Rev. Dis. Primers* 3:17009. doi: 10.1038/nrdp.2017.9

Ishiyama, N., Lee, S. H., Liu, S., Li, G. Y., Smith, M. J., Reichardt, L. F., et al. (2010). Dynamic and static interactions between p120 catenin and E-cadherin regulate the stability of cell-cell adhesion. *Cell* 141, 117–128. doi: 10.1016/j.cell.2010.01.017

Izquierdo, J. M., Majos, N., Bonnal, S., Martinez, C., Castelo, R., Guigo, R., et al. (2005). Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Mol. Cell.* 19, 475–484. doi: 10.1016/j.molcel.2005.06.015

Jankowsky, E., and Harris, M. E. (2015). Specificity and nonspecificity in RNA-protein interactions. *Nat. Rev. Mol. Cell Biol.* 16, 533–544. doi: 10.1038/nrm4032

Janzen, N. K., Kim, H. L., Figlin, R. A., and Belldegrun, A. S. (2003). Surveillance after radical or partial nephrectomy for localized renal cell carcinoma and management of recurrent disease. *Urol. Clin. North Am.* 30, 843–852. doi: 10.1016/s0094-0143(03)00056-9

Kim, Y., Choi, J. W., Lee, J. H., and Kim, Y. S. (2014). Loss of CDC14B expression in clear cell renal cell carcinoma: meta-analysis of microarray data sets. *Am. J. Clin. Pathol.* 141, 551–558. doi: 10.1309/AJCP4PE4JPSRGBQS

Ladomery, M. (2013). Aberrant alternative splicing is another hallmark of cancer. *Int. J. Cell Biol.* 2013:463786. doi: 10.1155/2013/463786

Lee, J. H., Ang, J. K., and Xiao, X. (2013). Analysis and design of RNA sequencing experiments for identifying RNA editing and other single-nucleotide variants. *RNA* 19, 725–732. doi: 10.1261/rna.037903.112

Li, J., Wang, Y., Rao, X., Wang, Y., Feng, W., Liang, H., et al. (2017). Roles of alternative splicing in modulating transcriptional regulation. *BMC Syst. Biol.* 11(Suppl. 5):89. doi: 10.1186/s12918-017-0465-6

Li, Y., Wang, Z., Wang, Y., Zhao, Z., Zhang, J., Lu, J., et al. (2016). Identification and characterization of lncRNA mediated transcriptional dysregulation dictates lncRNA roles in glioblastoma. *Oncotarget* 7, 45027–45041. doi: 10.18632/oncotarget.7801

Lorthongpanich, C., Thumanu, K., Tangkiettrakul, K., Jiamvoraphong, N., Laowtammathron, C., Damkham, N., et al. (2019). YAP as a key regulator of adipo-osteogenic differentiation in human MSCs. *Stem Cell Res. Ther.* 10:402. doi: 10.1186/s13287-019-1494-4

Mickisch, G. H. (2002). Principles of nephrectomy for malignant disease. *BJU Int.* 89, 488–495. doi: 10.1046/j.1464-410x.2002.02654.x

Payne, J. L., Khalid, F., and Wagner, A. (2018). RNA-mediated gene regulation is less evolvable than transcriptional regulation. *Proc. Natl. Acad. Sci. U.S.A.* 115, E3481–E3490. doi: 10.1073/pnas.1719138115

Repana, D., Nulsen, J., Dressler, L., Bortolomeazzi, M., Venkata, S. K., Tourna, A., et al. (2019). The Network of Cancer Genes (NCG): a comprehensive catalogue of known and candidate cancer genes from cancer sequencing screens. *Genome Biol.* 20:1. doi: 10.1186/s13059-018-1612-0

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Scotti, M. M., and Swanson, M. S. (2016). RNA mis-splicing in disease. *Nat. Rev. Genet.* 17, 19–32. doi: 10.1038/nrg.2015.3

Shi, J., Zhang, Y., Qin, B., Wang, Y., and Zhu, X. (2019). Long non-coding RNA LINC00174 promotes glycolysis and tumor progression by regulating miR-152-3p/SLC2A1 axis in glioma. *J. Exp. Clin. Cancer Res.* 38:395. doi: 10.1186/s13046-019-1390-x

Shukla, S., and Oberdoerffer, S. (2012). Co-transcriptional regulation of alternative pre-mRNA splicing. *Biochim. Biophys. Acta* 1819, 673–683. doi: 10.1016/j.bbagrm.2012.01.014

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistics, 2019. *CA Cancer J. Clin.* 69, 7–34. doi: 10.3322/caac.21551

Sutandy, F. X. R., Ebersberger, S., Huang, L., Busch, A., Bach, M., Kang, H. S., et al. (2018). In vitro iCLIP-based modeling uncovers how the splicing factor U2AF2 relies on regulation by cofactors. *Genome Res.* 28, 699–713. doi: 10.1101/gr.229757.117

The Gene Ontology Consortium, (2017). Expansion of the gene ontology knowledgebase and resources. *Nucleic Acids Res.* 45, D331–D338. doi: 10.1093/nar/gkw1108

Van Nostrand, E. L., Pratt, G. A., Shishkin, A. A., Gelboin-Burkhart, C., Fang, M. Y., Sundararaman, B., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* 13, 508–514. doi: 10.1038/nmeth.3810

Venables, J. P., Brosseau, J. P., Gadea, G., Klinck, R., Prinos, P., Beaulieu, J. F., et al. (2013). RBFOX2 is an important regulator of mesenchymal tissue-specific splicing in both normal and cancer tissues. *Mol. Cell. Biol.* 33, 396–405. doi: 10.1128/MCB.01174-12

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476. doi: 10.1038/nature07509

Wang, K., Saito, M., Bisikirska, B. C., Alvarez, M. J., Lim, W. K., Rajbhandari, P., et al. (2009). Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nat. Biotechnol.* 27, 829–839. doi: 10.1038/nbt.1563

Wang, Y., Liu, J., Huang, B. O., Xu, Y. M., Li, J., Huang, L. F., et al. (2015). Mechanism of alternative splicing and its regulation. *Biomed. Rep.* 3, 152–158. doi: 10.3892/br.2014.407

Weyn-Vanhentenryck, S. M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., et al. (2014). HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep.* 6, 1139–1152. doi: 10.1016/j.celrep.2014.02.005

Xu, X., Yu, Y., Zong, K., Lv, P., and Gu, Y. (2019). Up-regulation of IGF2BP2 by multiple mechanisms in pancreatic cancer promotes cancer proliferation by activating the PI3K/Akt signaling pathway. *J. Exp. Clin. Cancer Res.* 38:497. doi: 10.1186/s13046-019-1470-y

Yanagisawa, M., Huveldt, D., Kreinest, P., Lohse, C. M., Cheville, J. C., Parker, A. S., et al. (2008). A p120 catenin isoform switch affects Rho activity, induces tumor cell invasion, and predicts metastatic disease. *J. Biol. Chem.* 283, 18344–18354. doi: 10.1074/jbc.M801192200

Yee, B. A., Pratt, G. A., Graveley, B. R., Van Nostrand, E. L., and Yeo, G. W. (2019). RBP-Maps enables robust generation of splicing regulatory maps. *RNA* 25, 193–204. doi: 10.1261/rna.069237.118

Yin, L., Li, W., Wang, G., Shi, H., Wang, K., Yang, H., et al. (2019). NR1B2 suppress kidney renal clear cell carcinoma (KIRC) progression by regulation of LATS 1/2-YAP signaling. *J. Exp. Clin. Cancer Res.* 38:343. doi: 10.1186/s13046-019-1344-3

Zhang, C., Frias, M. A., Mele, A., Ruggiu, M., Eom, T., Marney, C. B., et al. (2010). Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. *Science* 329, 439–443. doi: 10.1126/science.1191150

Zhao, M., Kim, P., Mitra, R., Zhao, J., and Zhao, Z. (2016). TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. *Nucleic Acids Res.* 44, D1023–D1031. doi: 10.1093/nar/gkv1268

Zhu, Y. T., Chen, H. C., Chen, S. Y., and Tseng, S. C. (2012). Nuclear p120 catenin unlocks mitotic block of contact-inhibited human corneal endothelial monolayers without disrupting adherent junctions. *J. Cell Sci.* 125(Pt 15), 3636–3648. doi: 10.1242/jcs.103267

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 14 August 2020
doi: 10.3389/fgene.2020.00930

Poly C Binding Protein 1 Regulates p62/SQSTM1 mRNA Stability and Autophagic Degradation to Repress Tumor Progression

Wenliang Zhang[1,2†], Shaoyang Zhang[2†], Wen Guan[2†], Zhicong Huang[2], Jianqiu Kong[2], Chunlong Huang[3], Haihe Wang[2,4] and Shulan Yang[1,5*]*

[1] Translational Medicine Centre, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [2] Department of Biochemistry, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China, [3] Department of Hepatobiliary Surgery, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [4] Center for Stem Cell Biology and Tissue Engineering, Sun Yat-sen University, Guangzhou, China, [5] Guangdong Engineering and Technology Research Center for Disease-Model Animals, Sun Yat-sen University, Guangzhou, China

Accumulating evidence show that Poly C Binding Protein 1 (PCBP1) is deleted in distinct types of tumors as a novel tumor suppressor, but its tumor suppression mechanism remains elusive. Here, we firstly describe that downregulation of PCBP1 is significantly associated with clinical ovarian tumor progression. Mechanistically, PCBP1 overexpression affects various autophagy-related genes expression at various expression levels to attenuate the intrinsic cell autophagy, including the autophagy-initiating ULK, ATG12, ATG7 as well as the bona fide marker of autophagosome, LC3B. Accordingly, knockdown of the endogenous PCBP1 in turn enhances autophagy and less cell death. Meanwhile, PCBP1 upregulates p62/SQSTM1 via inhibition p62/SQSTM1 autophagolysome and proteasome degradation as well as its mRNA stability, consequently accompanying with the caspase 3 or 8 activation for tumor cell apoptosis. Importantly, clinical ovary cancer sample analysis consistently validates the relevance of PCBP1 expression to both p62/SQSTM1 and caspase-8 to overall survival, and indicates PCBP1 may be a master player to repress tumor initiation. Taken together, our results uncover the tumorigenic mechanism of PCBP1 depletion and suggest that inhibition of tumor cell autophagy with autophagic inhibitors could be an effective therapeutical strategy for PCBP1-deficient tumor.

Keywords: PCBP1, p62/SQSTM1, autophagy, apoptosis, ovary cancer, colon cancer, caspase-8

OPEN ACCESS

Edited by:
Juan Xu,
Harbin Medical University, China

Reviewed by:
Jing Su,
Jilin University, China
Verica Paunovic,
University of Belgrade, Serbia

****Correspondence:***
Haihe Wang
wanghaih@mail.sysu.edu.cn
Shulan Yang
yangshl3@mail.sysu.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *26 May 2020*
Accepted: *27 July 2020*
Published: *14 August 2020*

Citation:
Zhang W, Zhang S, Guan W, Huang Z, Kong J, Huang C, Wang H and Yang S (2020) Poly C Binding Protein 1 Regulates p62/SQSTM1 mRNA Stability and Autophagic Degradation to Repress Tumor Progression. Front. Genet. 11:930. doi: 10.3389/fgene.2020.00930

Abbreviations: 3-MA, 3-Methyladenine; Act D, Actinomycin D; ANOVA: Analysis of Variance; A 5TG, Autophagy-related gene; CQ, Chloroquine diphosphate salts; IHC, I b mmunohistochemistry; KD: knock down; LC3B, Microtubule-associated protein Light Chain 3B; OCs, Ovarian cancer specimens; p62, p62/SQSTM1; PARP, Poly (ADP-ribose) polymerase; PCBP1, Poly C Binding Protein 1; PFA, Paraformaldehyde; RT-PCR, Reverse transcriptase-polymerase chain reaction; SDS-PAGE, SDS-polyacrylamide gel electrophoresis.

INTRODUCTION

Poly C binding protein 1 (PCBP1) as an RNA binding protein is widely involved in different gene regulatory levels, which include gene transcription, translation, RNA transportation splicing and posttranscription (Giles et al., 2003; Nishinakamura et al., 2007; Ravikumar et al., 2010; Wang et al., 2010, 2018, 2019; Tripathi et al., 2016; Zhang et al., 2017a,b; Ishii et al., 2018; Shi et al., 2018; Zhao et al., 2019). Recently, PCBP1 as a novel tumor suppressor is characterized to be downregulated in many cancer types on inhibition of tumor formation and metastasis (Guo and Jia, 2018), including gastric cancer (Ji et al., 2017) and thyroid cancer (Ji et al., 2017). We have uncovered that PCBP1 delays the translation of metastatic PRL-3 which is broadly downregulated in variety of tumors, indicating that PCBP1 could be a potential tumor suppressor (Wang et al., 2010). In 2017, Jiani Guo et al. (2017) reported that PCBP1 mediates drug resistance in colorectal cancer. It was reported that PRL-3 enhances autophagy and promotes cell proliferation under nutrient-efficient and nutrient-poor conditions (Huang et al., 2014). Following this information, we then disclosed that PCBP1, as the suppressor of PRL-3, can really inhibit the starvation-induced autophagy of tumor cells to block tumorigenic initiation, independent of PRL-3 (Zhang et al., 2016), but we still do not know whether PCBP1 participates in the normal basal autophagy process in the nutrition-efficient situation. We recently showed that PCBP1 also increases cell cycle inhibitor, $p27^{Kip1}$ expression via its RNA binding capability to repress tumor cell cycle progression (Shi et al., 2018). Together with the previously mentioned, PCBP1 seems to work on multiple facets to inhibit tumor initiation and progression, but the underlying detailed mechanism remains elusive.

Autophagy is an evolutionally conserved process, in which the double-membrane vesicles (autophagosomes) initiated from multiple autophagy-related gene (ATG) products swallow and digest the damaged cytoplasmic organelles or proteins through lysosome-dependent degradation. Autophagy is putatively known as a cytoprotective response to cell stress for cell survival from cell death under stresses (Degenhardt et al., 2006; Villa et al., 2017). In contrast, this process could contribute to cell demise (autophagic cell death) when last too long time or works on overdose (Mizushima, 2007; Maiuri et al., 2010; Bialik et al., 2018). Many stresses can induce autophagy, which are composed of starvation, hypoxia and rapamycin inductions (Klionsky et al., 2012). Microtubule-associated protein Light Chain 3B (LC3B) is well known as the mammalian homolog of yeast *Atg8*. During autophagy, the LC3B protein undergoes modification from LC3B I to LC3B II served as a hallmark of formation (Klionsky et al., 2012). The p62/SQSTM1 (p62) protein is a link formed between LC3B II and autophagic substrates. p62 usually incorporates into the integrated autophagosome and can be subsequently degraded in autolysosomes, when autophagy process fully accomplishes (Bjorkoy et al., 2005; Pankiv et al., 2007). Thus, initiative autophagic flux can be indicated by LC3B-II amount and the accomplishment by p62 degradation status, respectively (Klionsky et al., 2012). The current results indicated that autophagy has dual roles in either promoting tumor initiation or inhibiting tumor progression (Levine, 2007; Mizushima, 2007; Galluzzi et al., 2015; Singh et al., 2018). Generally, tumor cells in tumor mass center are lack of nutrition, thus have higher autophagic flux than those in tumor margin regions, to prevent their death (Degenhardt et al., 2006). From another way, autophagy also exists in dying cells to result in the eventual cell death through the excessive consumption of cellular components (Janku et al., 2011; Young et al., 2012; Huang et al., 2013). Thus, roles of autophagy in tumorigenesis are highly dependent on pathological and physiological conditions of cell context and microenvironment. So far, it remains elusive whether PCBP1 modulates and participates in tumor cell autophagy in the nutrition-efficient condition.

Apoptosis is a form of programmed cell death and characterized by the cascade activation of caspases (Fulda and Debatin, 2006; Li and Yuan, 2008; D'Arcy, 2019). Caspase-8 is an initiator caspase in apoptosis. The auto-activation of caspase-8 starts from its oligomerization and self-cleavage. Subsequently, the activated caspase-8 facilitates the activation of pro-caspase-3, which is an executioner caspase, and promotes the apoptotic cleavage of poly (ADP-ribose) polymerase (PARP) for apoptosis (Kruidering and Evan, 2000). Recent reports suggested that cross talking between autophagy and apoptosis can coordinately regulate cell fate (Wu et al., 2014). As mentioned above, PCBP1 can suppress tumorigenesis, but we still also do not understand if it is related to tumor cell death.

Our results suggest that PCBP1 not only downregulates autophagic flux in the starvation conditions by suppressing LC3B expression as previously reported (Zhang et al., 2016), but also coordinately represses multiple autophagic genes expression, including ULK1, ATG7, ATG12 and p62 to suppress tumor cell autophagy initiation and commitment, and eventually to enhance tumor cell apoptosis. Thus, the expression states of PCBP1, p62 and Caspase-8 could be predictive biomarkers, and the anti-autophagic approaches would be potential therapeutical strategy for patients with silence of PCBP1 and high autophagic flux.

MATERIALS AND METHODS

Cell Culture and Treatment. Human Ovarian Cancer Cell Lines

SK-OV3 and A2780, Chinese hamster ovarian cell line (CHO), human colorectal cancer cell lines DLD-1 and HCT-116 were purchased from the ATCC company and maintained as previously reported by our group (Shi et al., 2018). Unless otherwise specified, chloroquine diphosphate salts (CQ, 50 μM 24 h, Sigma-Aldrich), 3-Methyladenine (3-MA, 3 μM, 12 h, Selleck, China), actinomycin D (Act D, 0.5 μg/ml, Biosharp, 8 h or 12 h) and MG132 (20 μM, 12 h, Sigma-Aldrich) was, respectively, dissolved in PBS or DMSO, and used to inhibit autophagic degradation, terminate the novel transcription or inhibit protein degradation.

DNA Constructs and Transfection

The A2780-PCBP1, DLD-1-PCBP1, HCT-116-PCBP1, and SK-OV3-PCBP1 stable cell lines were established as previously described (Wang et al., 2010). Four specific shRNA constructs as previously described (Zhang et al., 2016) for knockdown of PCBP1. The cell transfections were performed out with the Lipofectamine 2000 reagent (Invitrogen) as the manufacturer's instructions.

Antibodies and Western Blots (WB)

Antibodies were used against the following: PCBP1 (Cat No. sc-137249, Santa Cruz, WB 1:500); p62/SQSTM 1 (Cat No. 5114, Cell Signaling Technology (CST), WB 1:1000); LC3B (Cat No. 3868, CST, WB 1:1000, IF 1:200); Caspase-8 (Cat No.13423-1-AP, Protein Tech, China, WB 1:500); Caspase-3 (Cat No. 9662, CST, WB 1:1000); PARP (Cat No. 9542, CST, WB 1:1000); c-caspase-3 (Cat No. 9661, CST, WB 1:1000); ULK1 (Cat No.4773, CST, WB 1:1000); c-PARP (Cat No. 9541, CST, WB 1:1000); ATG7 (Cat No.8558, CST, WB 1:2000); ATG12 (Cat No.4180, CST, WB 1:1000); ATG5 (Cat No.2630S, CST, WB 1:1000); β-Actin (Cat No.4967, CST, WB1:2000); anti-mouse HRP-labeled secondary antibody (Cat No. 7076S, CST, WB 1:2000), GAPDH (Cat No. CW0101A, CW Bio-tech, China, WB 1:2000) and anti-mouse HRP-labeled secondary antibody (Cat No. 7076S, CST, WB 1:2000). The western blots protocols were followed as previously reported by our group (Zhang et al., 2016). Protein band intensity was, respectively, quantified and analyzed with densitometry by using Image J software.

Calculating Densitometry of Immunoblots

The quantitative densitometry of immunoblots was carried out as previously described (Zhang et al., 2016). Protein band intensity was respectively quantified and analyzed by using Image J software. Relative protein levels were calculated using densitometry values for GAPDH or β-actin as calibrators and shown under the protein bands.

Immunofluorescence Microscopy

For immunofluorescence, cells were seeded and grown on cover slips. And then, cells were washed in 1 X PBS and fixed in 4% paraformaldehyde (PFA). After permeabilization with 0.2% Triton X-100 (Biosharp), cells were probed with LC3B primary antibodies (CST, Cat No. 3868, IF 1:200) followed by anti-rabbit Alexa Fluor 594. Subsequently, cells were incubated with DAPI and mounted. Cells were observed by LSM710 confocal microscopy (Carl Zeiss AG).

Autophagy Assays

Autophagic flux was detected based on the amount of endogenous LC3B puncta, and validated with p62 protein level for autophagy completion. Cells were pre-treated with 50 μM CQ and DMSO for indicated time, respectively. The autophagy assay protocols are followed as previously reported by our group (Zhang et al., 2016).

Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR)

5 μg of RNA was reverse transcribed to cDNA by using M-MLV Reverse Transcriptase (Promega, Cat No. M1705) following the manufacturer's instruction. Gene-specific primers used for RT-PCR amplification of p62, ULK1, ATG 12, ATG 7, PRL-3 and internal control GAPDH were listed in **Supplementary Table 1**. The RT-PCR amplification is performed by using GoTaq® DNA polymerase (Promega, Cat No. M3005) in the amplification system, according to the manufacturer's protocol. DNA band intensity was quantified with densitometry via Image J software.

Flow Cytometry Apoptosis Analysis

To investigate the impact of PCBP1 overexpression on the tumor cell apoptosis, A2780-PCBP1 and A2780-GFP cells were seeded and cultured overnight, followed with treatment of 3 μM Paclitaxel or DMSO for 27 h, respectively. Next, adherent cells were trypsinized with 0.25% EDTA-free trypsin and stained with an APC-conjugated Annexin V and 7-ADD kit (KeyGEN BioTECH, China) following the manufacturer's instructions. The percentage of apoptotic cells were quantified by flow cytometer (Gallios, Beckman, United States).

mRNA Stability Assay

A2780-PCBP1 or GFP control cells were treated with DMSO and 0.5 μg/ml Act D (Biosharp) for 8 and 12 h, respectively. Five microgram of RNA was used for cDNA synthesis, and the transcript abundances of p62 mRNAs and GAPDH mRNA controls were quantified via semi-quantitative RT-PCR. The primers of p62 and GAPDH for RT-PCR was provided (**Supplementary Table 1**).

Ribosome Profiling of Autophagy Genes

Ribosome profiling protocols were followed as previously reported by our group (Shi et al., 2018). ULK1, ATG12, ATG7, PRL3, GAPDH mRNA was detected by qRT-PCR with the primers (**Supplementary Table 1**) as above.

Cell Proliferation Assay

Cells were seeded in 96-well plates in triplicates and cultured overnight. Then, cell culture medium was replaced with 100 μl serum-containing media with or without CQ treatment at the indicated time points. After treatments, cell counting kit-8 (CCK8) reagent (BB-4202-1, Best Bio, China) was added in each well (10 μl per well) and the cells were incubated for another 2 h. The absorbance at 450 nm of each well was measured by enzyme-labeling instrument (Multiskan G0; Thermo Fisher Scientific, Inc.).

Immunochemistry (IHC) Analyses of PCBP1, p62, and Caspase 8 Expressions in Ovary Tumor Samples

Human freshly frozen colon tumor tissues were collected from Sun Yat-sen University Cancer Center, under their Standard Experimental Ethics Protocol and approved by Sun Yat-sen University Research Ethics Committee. Three ovary tissue arrays

were purchased from Xi 'an Ailina Biotechnology Co., Ltd., under their Standard Experimental Ethics Protocol. These tissue arrays were stained with anti-PCBP1 (1:200 dilution, Abcam, United States), anti-p62 (1:200 dilution, Sanying Biotechnology Co., Ltd., Wuhan, China) and anti-Caspase-8 (1:200 dilution, Protein Tech, China) antibodies, respectively. Immunochemistry protocol and the semi-quantitative score was followed as previously reported by our group (Shi et al., 2018). The semi-quantitative score is presented as Score = SI (staining intensity) x PP (percentage of positive cells), in which SI was determined as ten levels including 0 (negative); 1; 2; 3; 4; 5 (**Supplementary Figure 1**). Likewise, PP was defined as 0, < 5%; 0.2, 6%–30%; 0.5, 31%–70%; and 1.0, > 70% positive cells. As an example, the relative score in **Supplementary Figure 1** based on IRS are determined as 0, 1, 2, 3, 4, and 5.

Statistical Analysis

Statistical analyses were performed by GraphPad Prism software (Version 5.00). Unless otherwise specified, the unpaired *t*-tests and Analysis of Variance (ANOVA) analysis were used to compare two groups and multiple groups, respectively. $P < 0.05$ were considered as significant. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

RESULTS

PCBP1 Is Downregulated and Negatively Related to Tumor Malignancy in Ovarian Tumors

We previously reported that PCBP1 is broadly downregulated in lung and colon cancers (Wang et al., 2010). PCBP1 downregulation promotes cell proliferation and tumorigenicity of ovarian cancer cells both *in vitro* and *in vivo* (Zhang et al., 2016; Shi et al., 2018). To further confirm the clinical relevance of PCBP1 to ovarian cancer progression, we examined PCBP1 expression state by immunohistochemistry (IHC) in 90 cases of ovarian cancer, compared with the 10 corresponding cancer adjacent ovarian tissues among them. Our IHC results were scored (**Supplementary Figure 1**) and indicated that PCBP1 was more detectable in cancer adjacent tissues than ovarian cancer samples (**Figures 1A,B**). Among ovarian cancer cases, according to tumor node metastasis (TNM) classification, PCBP1 levels were especially lower in pT3 group (tumor with micro-metastasis of extra pelvic peritoneum confirmed by microscope) vs. pT1/2 group (tumor with or without pelvic spread) (**Figures 1C,D**), as well as the late clinical stage (stage III and IV) vs. early clinical stages (stage I and II) based on clinicopathologic features (**Figures 1E,F**). Statistically significant decreased PCBP1 expression was also found in ovarian cancer samples with positive lymph node and distant metastasis statues, compared with those with negative status (**Figures 1G,H**). Moreover, we summarized the correlation between PCBP1 expression and clinicopathologic variables of patients with ovarian cancer (**Table 1**). In addition, our immunohistochemical staining results also suggested that PCBP1 protein expression is significantly decreased in tumor regions of colorectal cancer compared with the paired fresh normal tissues (**Supplementary Figure 2**; Shi et al., 2018). The overall analysis showed that low PCBP1 levels were significantly associated with tumor stages ($p < 0.01$), clinical stages ($p < 0.05$), lymph node and distant metastasis ($p < 0.05$).

PCBP1 Inhibits the Basal Intrinsic Autophagy in Tumor Cells

Although we have shown that PCBP1 inhibits autophagy to repress cell proliferation in nutrition-deficient condition via downregulation of LC3B, a key gene of autophagy degradation (Zhang et al., 2016), but it is still not well understood that whether PCBP1 is also involved in the intrinsic basal autophagy regulation to impair tumorigenesis. To thoroughly answer this question, we conducted immunoblotting and showed that, compared with the control cells transfected with empty GFP vector, GFP-PCBP1 overexpression an obviously decreased the expression of LC3B I and LC3B II in A2780, DLD-1, HCT-116, and SK-OV3 cells from different tissue origin in nutrient-rich condition (**Figure 2A**). These results also indicate the existence of intrinsic autophagy event in various tumor cells, which could be regulated by PCBP1 overexpression. To investigate the fundamental function of PCBP1 in the basal autophagy regulation, immunofluorescence staining is performed and further validated the reduced aggregated LC3B puncta by PCBP1 overexpression upon an autophagy degradation inhibitor, chloroquine (CQ) treatment in both A2780 and DLD-1 cells under nutrition-rich conditions (**Figure 2B**). The LC3B puncta intensity in a cell is quantified and results demonstrate that PCBP1 overexpression cells have dramatically less LC3B puncta intensity than the parent control cells (**Figure 2C**). In addition, we noted that the quantified results of LC3B puncta intensity in **Figure 2C** did not significant decrease in the PCBP1 overexpression group compared with the control group, but the immunoblotting results in **Figure 2A** showed a significant decrease in LC3B expression. We think that the possible reason is due to the sensitivity of LC3B antibody that is not so sensitive for immunofluorescence detection, in comparison to immunoblotting detection. In line to this, immunoblotting results indicated that PCBP1 overexpression downregulates both types of LC3B I and II expression in full-nutrition situation, compared with its GFP control cells (**Figure 2D**, lane 1 vs. 2 and lane 5 vs. 6). Whereas, upon CQ treatment, LC3B levels were not completely recovered to the control GFP level (**Figure 2D**, lane 3 vs. lane 4 and lane 7 vs. lane 8), which is consistent with our previous results that PCBP1 downregulates LC3B expression independently of autophagic degradation (Zhang et al., 2016).

To verify if endogenous PCBP1 really participate in basal autophagy regulation, endogenous PCBP1 was knocked down with 4 specific shRNAs in A2780, CHO, and HCT-116 cells with relatively high PCBP1 expression (**Supplementary Figures 3A,B**). Immunofluorescence staining revealed the increased aggregation of LC3B puncta intensity by deletion of endogenous PCBP1 upon CQ treatment in both CHO and HCT-116 cells under nutrition-efficient conditions (**Figure 2E**). CHO cells are relatively normal cells compared with the ovarian

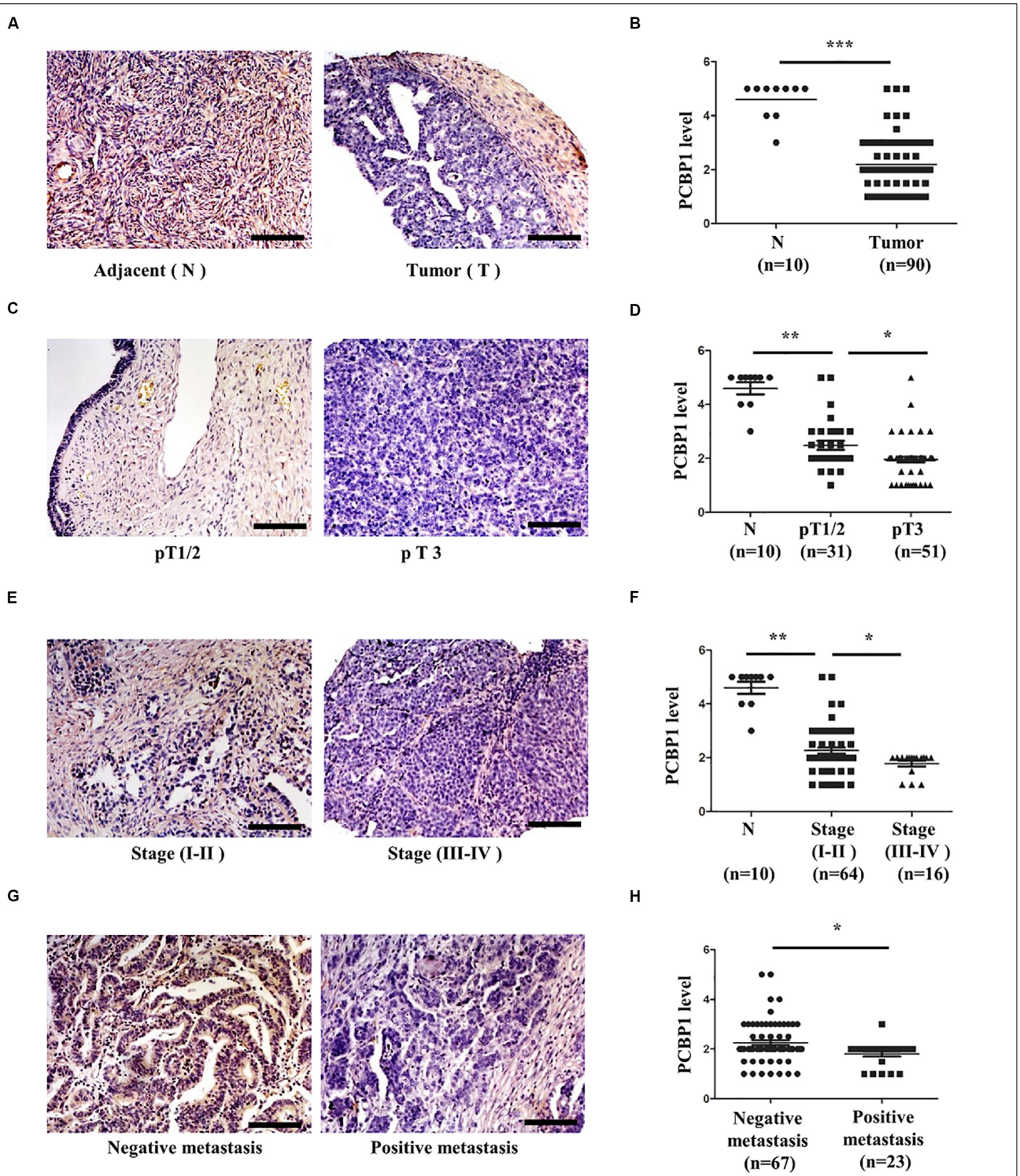


FIGURE 1 | PCBP1 expression is negatively related to ovary cancer progression. **(A)** Representative immunohistochemistry (IHC) analysis of PCBP1 expression in the paired adjacent normal tissues (N) and their ovarian tumor samples (T). **(B)** Statistical analysis of PCBP1 expression between normal and tumor tissues detected in **(A)**. **(C)** Representative PCBP1 expression level in pT1/2 tumors (with or without pelvic spread) vs. pT3 tumor (with micro-metastasis of extra pelvic peritoneum confirmed by microscope). **(D)** Statistical comparison of PCBP1 expression between the pT1/2 tumors and the pT3 tumors detected in **(C)**. **(E)** Representative PCBP1 expression in tumor samples with different tumor progression (clinical stages). **(F)** Statistical analysis of PCBP1 expression with tumor progression (clinical stages). **(G)** Representative PCBP1 expression in primary tumor samples and the metastatic tumor samples. **(H)** Statistical analysis of PCBP1 expression between the primary and metastatic ovary cancer samples. All IHC images are photographed with 200X amplification. Scale bars equal to 100 μm. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

TABLE 1 | PCBP1 expression and clinicopathologic characteristics of ovarian cancer patients.

	PCBP1			
Characteristics	**Low**	**Median**	**High**	***p*-value**
Pathology				
Adjacent	–	1	9	< 0.0001
Tumor	70	13	7	
Ages				
= 46	23	5	5	0.6748
>46	47	9	11	
Tumor stage				
T_{1+2}	19	8	4	0.0062
T_{3+4}	44	5	2	
Clinical stage				
I + II	46	13	5	0.0317
III + IV	16	–	–	
Metastasis				
Negative	49	13	5	0.0177
Positive	22	1	–	

t-test was used. PCBP1 low means IHC score as 1 and 2; Median, 3; High, 4 and 5. Refer to ***Supplementary Figure 1A****.*

cancerous cells. Therefore, the results from CHO cells can evidently indicate the intrinsic autophagy event in normal ovary cells (tissues) to further show the general information. Likewise, deletion of PCBP1 robustly boosted the modified LC3B-II accumulation, which indicated as the ratio of LC3B-II to GAPDH in A2780 cell line (**Supplementary Figure 3C**, lane 1 vs. lane 2) in nutrition-rich conditions. After CQ treatment, LC3B levels were still more than the control level, demonstrating downregulation effect of LC3B by PCBP1 deletion (**Supplementary Figure 3C**, lane 3 vs. lane 4, lane 5 vs. lane 6). In this study, although the ovary cancer here is focused, the consistent results observed in CHO cell line and two colon cancer cell lines could to suggest the general observation of PCBP1 in autophagy modulation. Taken together, our results verified that PCBP1 downregulates LC3B expression to result in basal autophagy inhibition.

PCBP1 Upregulates p62/SQSTM1 Through Inhibiting Its Autophagic Degradation

Considering that LC3B expression can be suppressed by PCBP1 (Zhang et al., 2016), and it may not suitable to evaluate the influence of basal autophagic flux by PCBP1 with LC3B accumulation. In addition, autophagy is a complex, dynamic process, and LC3B-II accumulation just reflects the autophagic initiation, not the performance outcome (Ravikumar et al., 2010). Thus, p62/SQSTM1 (p62) was used to examine PCBP1's influence in the basal autophagic flux, as p62 is the well-known substrate of autophagic degradation. According to this concept, we determined p62 level and western blotting results presented that, compared with the GFP control cells, p62 expression was robustly increased in various types of PCBP1-overexpressing cells under normal condition (**Figure 3A**). Quantification of p62 also confirmed a significantly higher p62 expression in PCBP1 overexpressing A2780 and DLD-1 cells compared with their GFP control cells, respectively (**Figure 3B**). On the contrary, endogenous PCBP1 depletion evidently decreased p62 protein level in A2780 cell line, HCT-116 cell line and CHO cell lines (**Figure 3C**). When inhibiting the autophagy completion with CQ, p62 level could be almost restored back to that of the parental control cells in DLD-1 cells, but not in A2780 (**Figure 3D**) exceptionally. Accordingly, CQ treatment similarly reversed p62 level in the PCBP1 depletion cells back to that in normal GFP control level (**Figure 3E**), suggestive of the p62 regulation by PCBP1 is at least an autophagy-dependent manner.

p62 Is Also Upregulated via Proteasome Degradation Inhibition and mRNA Stabilization by PCBP1

Comparing p62 protein level between the CQ inhibited parental control cells and PCBP1-overexpressing cells, it seems that p62 could be modulated not only by the autophagy-dependent degradation, since CQ treatment did not restore p62 expression back to the normal level. We further investigated which other mechanism may be involved in p62 regulation. We treated cells with Act D or MG132 to block the novel mRNA transcription or protein proteasome degradation, respectively. Immunoblots showed that, after MG132 (a known proteasome inhibitor) treatment, p62 expression was obviously increased in the GFP cells, but not in GFP-PCBP1 overexpressing cells, compared with the corresponding control groups (**Figure 4A**), indicating PCBP1 can somehow inhibit proteasome-mediated p62 degradation. As PCBP1 is an RNA-binding protein as we have shown, we further carried out a semi-quantitative RT-PCR detection of p62 mRNA level, and our results indicated that PCBP1 overexpression led to more p62 mRNA amount, whereas endogenous PCBP1 knock down resulted in less p62 mRNA, compared with control cells (**Figure 4B**). To further clarify whether p62 mRNA amount alteration is due to its mRNA stability or transcriptional activation, we first blocked the novel mRNA synthesis with Act D treatment and detected stability of the newly synthesized p62 mRNA and its protein expression. Results showed that compared with that of control cells, PCBP1 delayed the degradation efficiency of the nascent p62 mRNA (**Figure 4B**). In addition, we knocked down the ATG5 expression in both DLD-1 GFP and PCBP1-overexpressing cells to reduce autophagic degradation and demonstrated that p62 protein was still high in the PCBP1-overexpressing cells, presenting the similar ratio to p62 mRNA level (**Figure 4C**). Altogether, the above results showed the authentic cause of p62 upregulation by PCBP1 is resulted from the multiple events, including the inhibitions of p62 autophagic and proteasome-mediated degradations, as well as the stabilized p62 mRNA level.

PCBP1 Coordinately Regulates Multiple Autophagy Genes on Translational Level

Given PCBP1 as an RNA-binding protein, it is involved in various levels of gene expression regulations, including transcription,

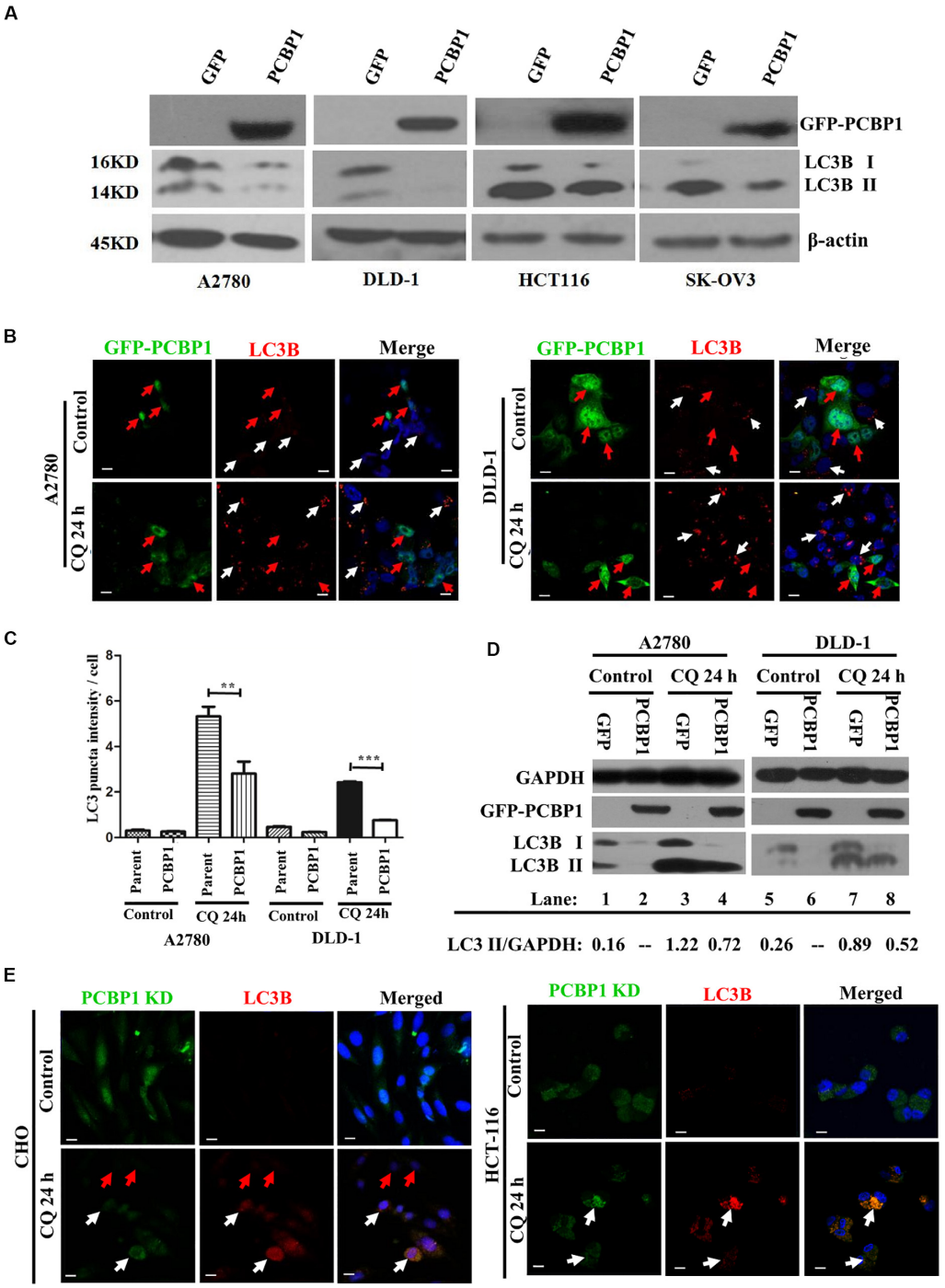


FIGURE 2 | PCBP1 Overexpression represses autophagy. **(A)** Immunoblots of PCBP1, LC3B I and II in PCBP1-overexpression (GFP-PCBP1) and the GFP control cells. β-actin is used as an internal loading control. **(B)** Immunofluorescence staining of LC3B puncta in both A2780 and DLD-1 cells with or without CQ treatment for 24 h. GFP-PCBP1 cells are in green, while parent control cell are not green. Cells counterstained with DAPI are shown in blue. White arrows point LC3B puncta (Red) in the parental control cells without GFP-PCBP1 transfection (not green), while the red arrows point LC3B puncta in GFP-PCBP1-transfected cells (green). Bars equal to 25 μm. **(C)** Quantifications of LC3B puncta intensity per cell in B are presented as histograms (mean ± SD). Parent group indicates the parent control cells without GFP-PCBP1 (not green), while the PCBP1 group indicates the cells with transfected GFP-PCBP1 (green). $^{**}P < 0.01$; $^{***}P < 0.001$, $n = 5$. **(D)** Immunoblots of LC3B I, II in A2780 and DLD-1 cells with PCBP1 and GFP overexpression. Upon CQ or DMSO (control) treatment for 24 h, cells were analyzed. Immunoblot intensity ratio of LC3B II to GAPDH were, respectively, quantified and normalized, and indicated in each lane. **(E)** Immunofluorescence staining of LC3B puncta in both A2780 and DLD-1 cells with or without CQ treatment for 24 h. PCBP1 knockdown (KD) cells are in green, while parent control cells are not green. Cells counterstained with DAPI are in blue. White arrows point LC3B puncta (Red) in the PCBP1 KD cells, while the red arrows point LC3B puncta in the parental control cells. Scale bars in **(B,E)** equal to 25 μm.

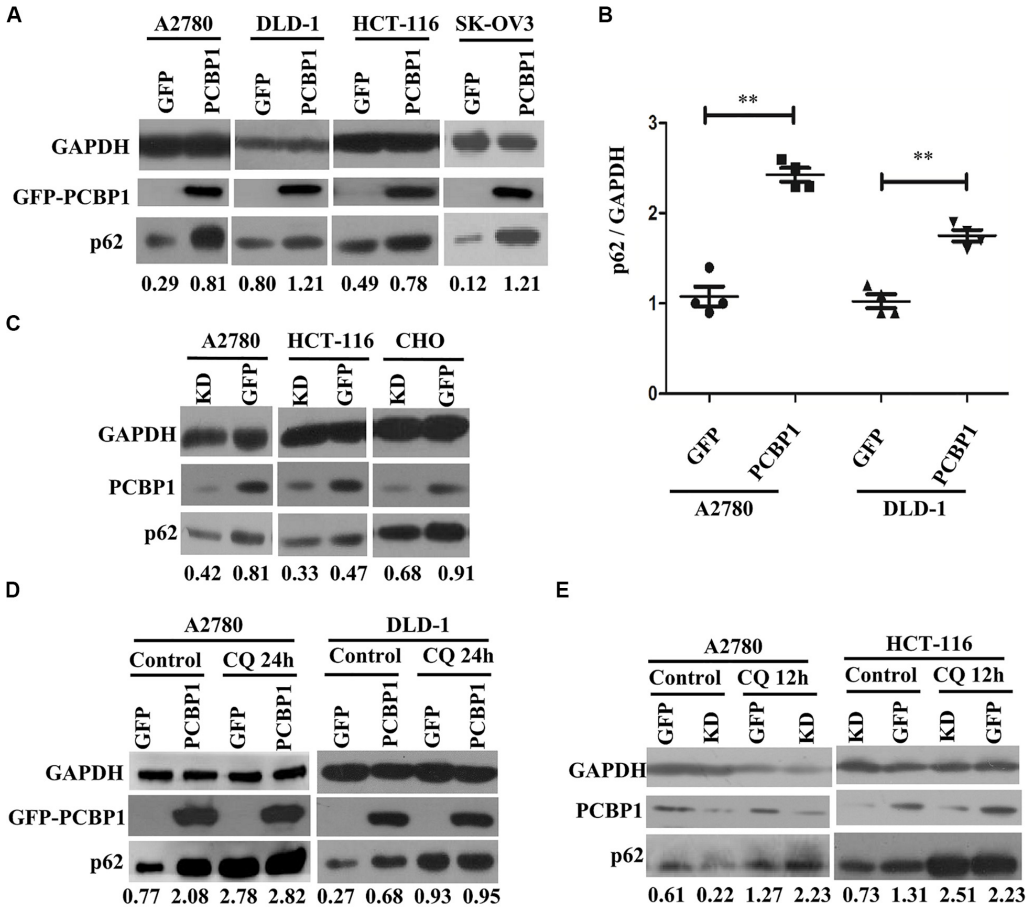


FIGURE 3 | PCBP1 inhibits autophagy completion. **(A)** Immunoblots of p62 and GAPDH in the PCBP1 overexpression (GFP-PCBP1) and the GFP control cells by at least two independents. **(B)** Statistical analysis of p62 expression in PCBP1 overexpressing cells, compared with the GFP control cells by at least three independent immunoblots. $^{**}p < 0.01$. **(C)** Western blots of p62 in the indicated endogenous PCBP1 knockdown cells. Four GFP-confused shRNAs against PCBP1 were transfected into cells to establish the PCBP1 knockdown cells, while the GFP empty vector-transfected cells are control cells. **(D)** Immunoblots of p62 in A2780 and DLD-1 cells with PCBP1 and GFP overexpression. Upon CQ or DMSO (control) treatment for 24 h, cells were analyzed by at least two independents. **(E)** Immunoblots of p62 in A2780 and HCT-116 cells with PCBP1 knock down (KD) by at least two independents. GAPDH is used as an internal loading control, and the expression ratio of p62 to GAPDH was quantified and normalized, and indicated under each lane.

splicing, translation and RNA stability (Giles et al., 2003; Wang et al., 2010; Zhang et al., 2010; Shi et al., 2018). To determine whether PCBP1 can regulate the autophagy-related genes at different levels, such as at the translational level, we conducted the gene expression analysis by PCBP1 overexpression as well as the ribosome profiling analyses of these genes' translational status. Ribosome profiling results showed that PCBP1 clearly repressed ULK, ATG12, ATG7 translational efficiency, compared the known PCBP1 translationally repressed metastatic phosphatase PRL-3 (Wang et al., 2010), whereas the housekeeping gene GAPDH as an internal control was not affected (**Figure 5A**). In addition, the ATG12 and ATG7 were observed in our PCBP1-bound mRNA pool by immunoprecipitation (**Supplementary Figure 4**). To confirm the results, we did immunoblots and showed that these autophagy-related protein levels were clearly decreased accordingly (**Figure 5B**), validating the translational influence of PCBP1 in autophagy-initiating genes expression. Taken together, all above results suggested that PCBP1 represses autophagy through multiples target genes at various regulation levels.

PCBP1 Coordinately Favors Tumor Cell Apoptosis

Given that autophagy plays different roles in tumorigenic outcome, to examine tumor cell fate upon autophagy retardation by PCBP1 in the normal culture conditions, we analyzed the apoptotic status of cells with PCBP1 aberrant overexpression. In line with our previous observation (Zhang et al., 2016), compared with the parent control cells, the PCBP1 aberrant expression obviously increased the cleaved-caspase-8 (c-caspase-8), cleaved-caspase-3 (c-caspase-3), and cleaved-PARP (c-PARP) levels in DLD-1 cells, as well as in A2780 cells with the exception of the caspase-8, which may be due to the cell type context specificity. Those results consistently suggested that PCBP1 could instinctively drive cell apoptosis, while suppressing autophagy

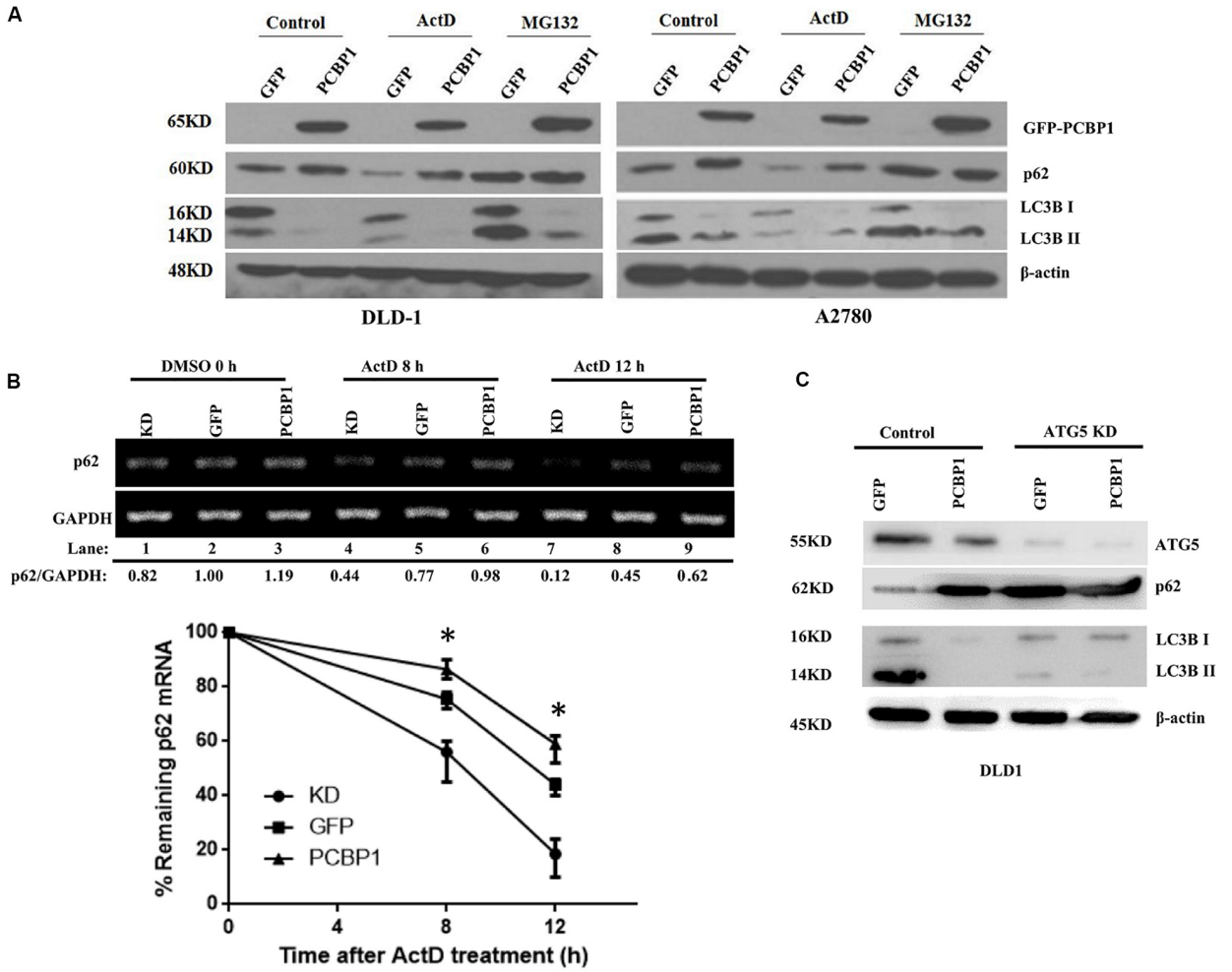


FIGURE 4 | PCBP1 enhance p62 expression in multiple levels. **(A)** Immunoblots of p62 and LC3B in the indicated GFP-PCBP1 cells as well as their parental cells treated without (Control) or with Actinomycin D (ActD) or MG132, respectively. **(B)** Semi-quantitative RT-PCR analysis of p62 mRNA stability (upper panel) in A2780 cells with endogenous PCBP1 knockdown (KD) or PCBP1 overexpression (PCBP1), compared with control cells (GFP) by at three independents. GAPDH is used as an internal control, and the ratio of p62 mRNA level to GAPDH was quantified, normalized and indicated under each lane. *$P < 0.05$. **(C)** Immunoblot of p62 protein level in DLD-1 cells with or without ATG5 knockdown by at least two independents. GAPDH was used as an internal loading control.

(**Figure 6A**). On the contrary, the instinct apoptotic signals attenuated in both PCBP1 depletion cells (**Figure 6B**). Upon autophagy blockade with CQ treatment in nutrient-sufficient condition, the apoptotic status was increased at the relative level in the parental cells which is similar to that in PCBP1-overexpressing cells (**Figure 6C**), demonstrating that the high level of p62 induced by PCBP1 can somehow enhance apoptosis. This observation is in line with the notion that p62 is indeed involved in cell death regulation (Zhang et al., 2016; Tao et al., 2020). To confirm the physiological outcome, we conducted cell proliferation assay and showed that PCBP1 overexpression delayed cell proliferation in normal condition, compared with the control cells; while inhibition of intrinsic autophagy with CQ treatment in both the parental control cell and PCBP1 overexpressing cells efficiently repressed the cell proliferation, further indicating this intrinsic autophagy benefits to cell proliferation (**Supplementary Figure 5**). Apoptosis analysis also showed overexpression of PCBP1 could even favor tumor cell apoptosis in normal culture condition (**Figure 6D**). Together, our results clearly demonstrated that PCBP1 coordinately inhibits intrinsic tumor cell autophagy and favors cell apoptosis for tumorigenesis inhibition.

Clinicopathological Correlation of PCBP1, p62, and Caspase-8 in Ovarian Cancer

To validate if there is a general clinicopathological correlation between PCBP1 and p62, p62 expression was examined with IHC in another set of the serial sections of ovarian tissues. IHC results revealed that p62 was similarly detectable in ovarian cancer adjacent tissues, and subdued in ovarian tumor samples (**Figure 7A** and **Supplementary Figure 6A**). Moreover, similar to PCBP1 expression pattern in ovarian sample, we

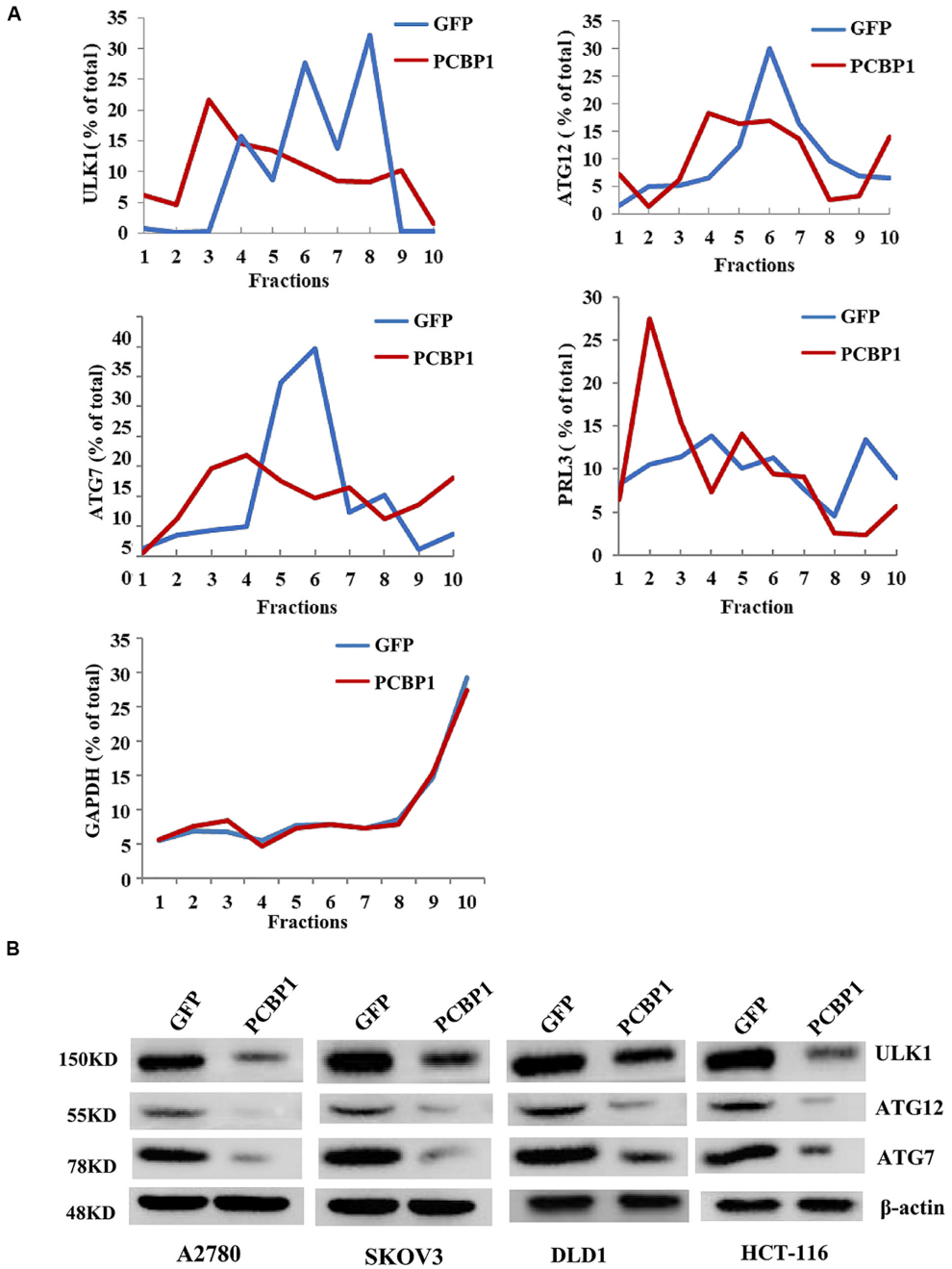


FIGURE 5 | PCBP1 translationally enhance autophagy genes expression. **(A)** Polyribosome profiling of the indicated autophagy-promoting genes in A2780 cells with GFP-PCBP1 overexpression as well as the control GFP cells. As PRL-3 is translationally repressed by PCBP1 (Wang et al., 2010), it is used as positive control, GAPDH as a negative control. **(B)** Immunoblots of the indicated autophagy-promoting genes in various cell types with GFP-PCBP1 expression, compared with their GFP cells by at least two independents.

observed statistically significant p62 expression decrease in ovarian cancer samples vs. the adjacent samples (**Figure 7B**). However, there was no difference between pT3 group (tumor with micro-metastasis of extra pelvic peritoneum confirmed by microscope) vs. pT1/2 group (tumor with or without pelvic spread) tumors (p = 0.5920) (**Figure 7C**), late clinical stage (stages III and IV) vs. early clinical stages (stages I and II) (p = 0.8129) based on clinical features (**Figure 7D**), positive lymph node and distant metastasis statues vs. those with negative status (p = 0.9531) (**Supplementary Figure 6B**). Furthermore, we also observed that p62 is consistently higher in PCBP1 positive-adjacent tissue, while consistently lower in pT1 (without pelvic spread) and pT3 tumor tissue with downtrend (**Figure 7A**). To thoroughly validate the relationship between PCBP1, p62 and

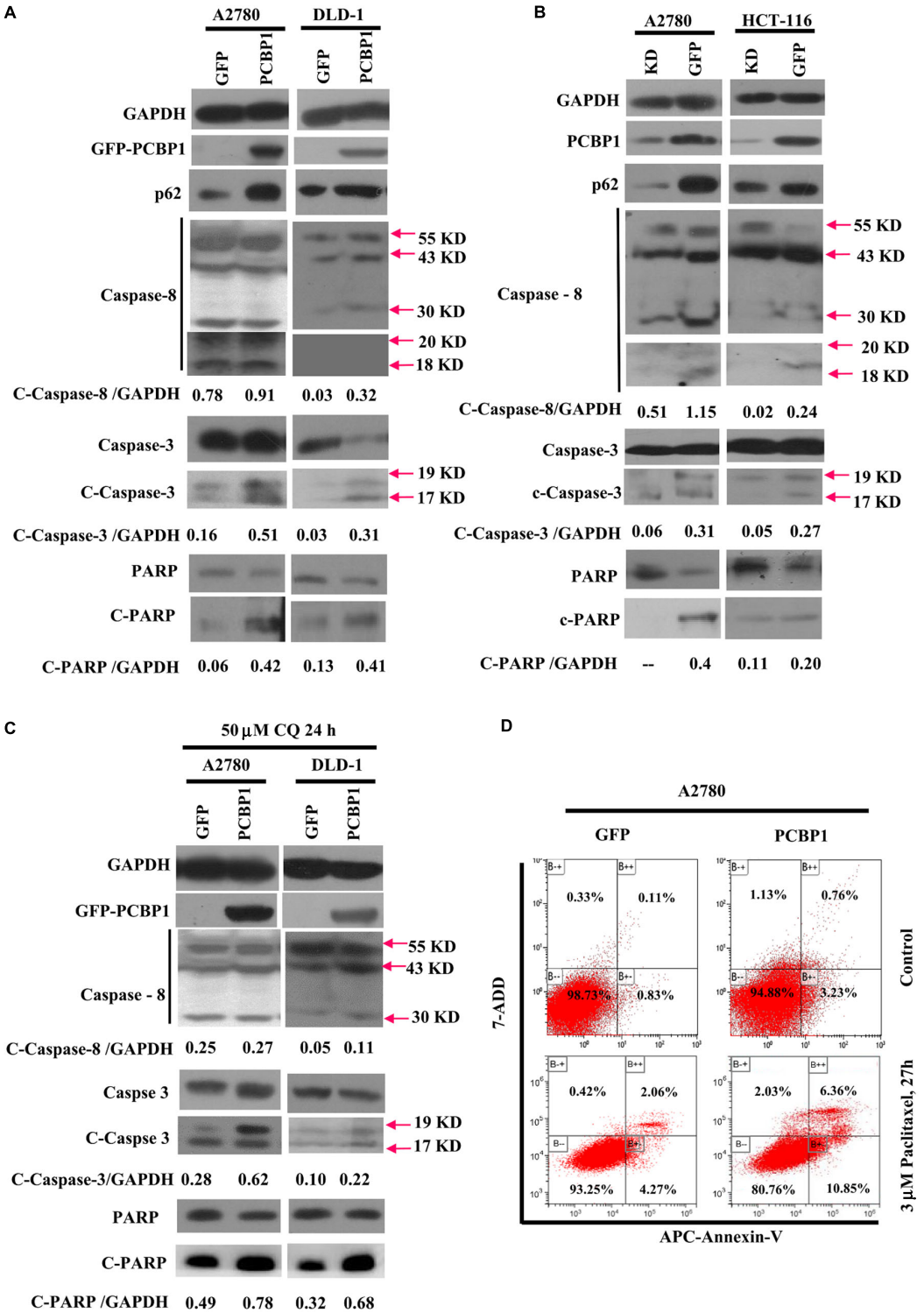


FIGURE 6 | PCBP1 overexpression enhances caspases activation for apoptosis. **(A)** Immunoblots of p62, PARP and caspases and their cleaved forms in A2780 and DLD-1 cells with PCBP1 and GFP overexpression at least two independents. **(B)** Immunoblots of p62, PARP and caspases and their cleaved forms in A2780 and HCT-116 cells with endogenous PCBP1 knockdown (KD), compared with A2780 and HCT-116 GFP controls by at least two independents. **(C)** Upon CQ treatments for 24 h, immunoblots of PARP, caspases and their cleaved forms in A2780 and DLD-1 cells with PCBP1 and GFP overexpression at least two independents. Arrows point and show the molecular weight of each protein. The cleaved forms of these proteins are shown as c-protein. **(D)** Flow cytometric analyses of the percentages of apoptotic cells in A2780 cells with PCBP1 overexpression, compared with the A2780 GFP control cells by at least two independents. Overall percentages of apoptotic cells are defined as the sum of Annexin-V positive cells.

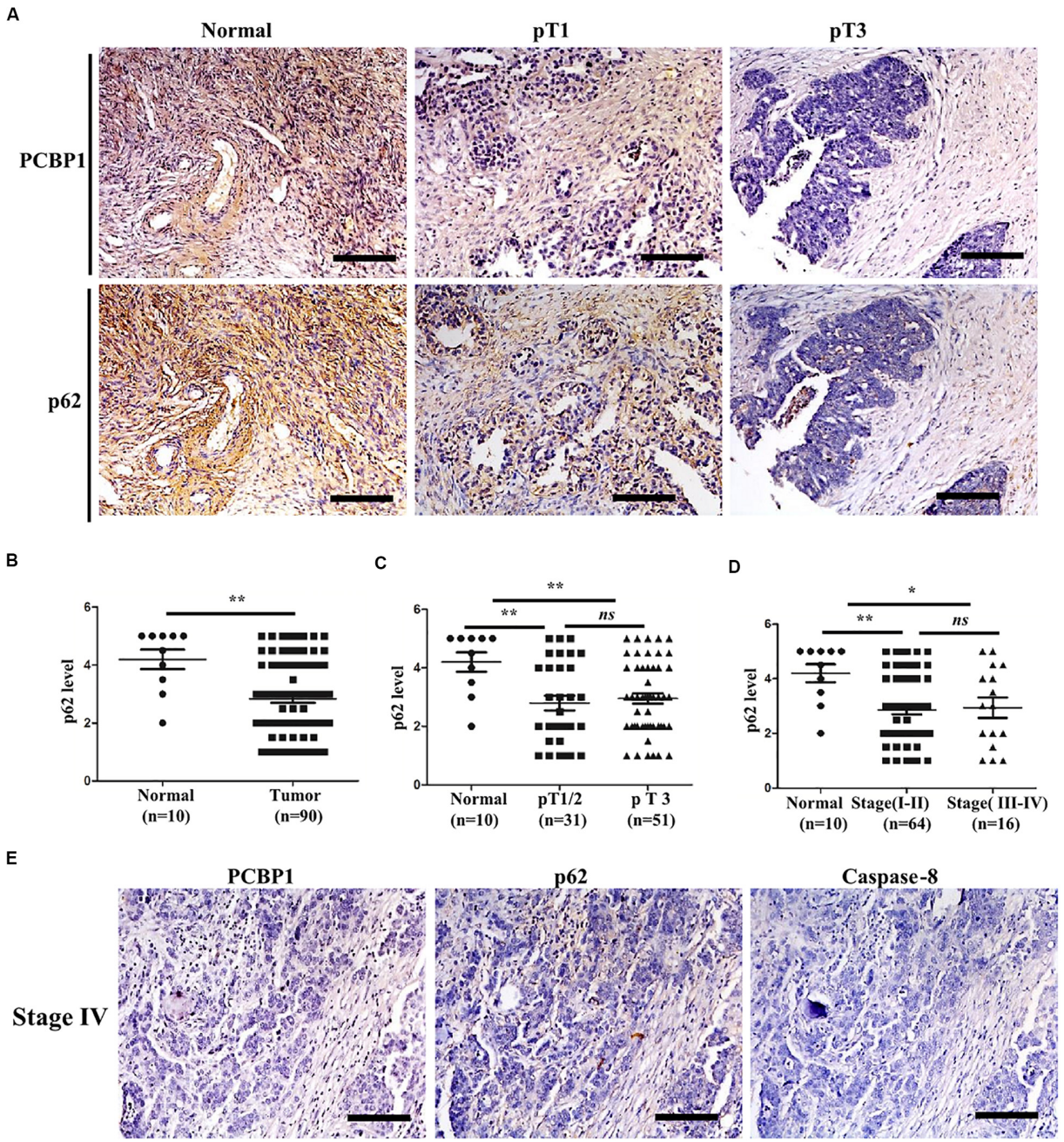


FIGURE 7 | Relevance of PCBP1 to p62 and caspase 8 in ovary cancers. **(A)** Representative immunohistochemistry (IHC) analysis of p62 expression in the paired adjacent normal tissues and their ovarian tumor samples. pT1, tumor without pelvic spread; pT3, tumor with micro-metastasis of extra pelvic peritoneum confirmed by microscope. **(B)** Statistical comparison of p62 expression between normal and tumor tissues detected in **(A)**. **(C)** Statistical comparison of p62 expression between the pT1/2 group (tumors with or without pelvic spread) and the pT3 group (tumors with micro-metastasis of extra pelvic peritoneum confirmed by microscope). **(D)** Statistical analysis of p62 expression with tumor progression (clinical stage). **(E)** Representative PCBP1, p62 and caspase 8 expressions in a same ovary tumor sample with clinical stage IV. All images in are photographed with 200X amplification. Scale bars on equal to 100 μm. $^{*}P < 0.05$; $^{**}P < 0.01$; ns, no significance.

apoptosis in tumor samples, we use the same batch of ovary cancer samples to check these 3 protein expressions at same time (**Figure 7E** and **Supplementary Figure 6C**). IHC results also showed that PCBP1 expression was positively correlated with p62 expression (**Supplementary Figure 6D**, $R^2 = 0.506$) and Caspase-8 expression levels (**Supplementary Figure 6E**, $R^2 = 0.447$) in ovarian tumor samples, respectively. In line with the above observation, the correlation analysis indicated that p62

is positively correlated with Caspase-8 as well (**Supplementary Figure 6F**, R^2 = 0.598), which is upregulated in ovarian tumor adjacent tissues and is downregulated in ovarian tumor tissues. Overall, we found that PCBP1, p62, and Caspase-8 expression levels were correlated with each other in ovarian cancers (**Figure 7** and **Supplementary Figure 6**).

DISCUSSION

Accumulating evidence indicates that autophagy has complicated effects in tumorigenesis, which are dependent on the tumor developmental stage. For instance, it is documented that autophagy induction is well correlated to the worse prognosis in ovarian carcinoma (Zhao et al., 2014), but other study in turn indicates that autophagy accompanies higher overall survival of the ovary cancer patient, as high autophagic flux could sensitize the tumor cells to chemotherapy (Valente et al., 2014). Considering our previous result (Zhang et al., 2016), we assume that that once cells get transformed, the intrinsic autophagy would promote tumor cell survival in both nutrition-efficient and -deprived conditions, as long as the autophagy level (or intensity) is not too high. PCBP1 may naturally suppress or balance the intrinsic basal cell autophagy for cell homeostasis maintenance, as PCBP1 can affect various autophagy-related genes expression at the different autophagic stages, including the autophagy initiation-related genes ULK, ATG12 and ATG7, the autophagosome formation, LC3B, the autophagosome-lysome fusion, p62 in nutrient-efficient condition (**Figure 5**), which is usually characterized in the nutrient-deficient condition. As PCBP1 expression is shown to be broadly reduced in many types of tumor (Guo and Jia, 2018), our results here further demonstrate that PCBP1 represses intrinsic autophagic levels in both ovary tumors and colon tumors, indicating that p62 would be an general accompanying marker for other types of PCBP1-depleted tumors.

As an RNA-binding protein, PCBP1 can modulate multiple genes expressions on different expression levels, including RNA stability, RNA folding, splicing, translation, etc. Coordination regulation is also a conserved and efficient regulatory strategy to control expression of genes involved in sustaining organism homeostasis upon environmental changes. PCBP1, likely as a critical multifaceted gene may generally conduct such coordinated function via its RNA binding feature. Whether all these autophagy-related genes share a conserved RNA-binding sequence would be further investigated.

Currently, it is recognized that the exact role of autophagy in tumor progression is attributed to the component in tumor cells, tumor stages, physiological and tumor microenvironment (Levine, 2007; Mizushima, 2007; Galluzzi et al., 2015). At the early stages of tumorigenesis, when tumor vascular supply is limit, autophagy serves as a pro-survival mechanism for tumor cell by consuming its own unnecessary materials (Degenhardt et al., 2006; Levine, 2007; Huang et al., 2014). In line with this, our results thus have shown that for the tumor progression, in some circumstances, PCBP1 expression is inevitably repressed to induce relatively high autophagic level (**Figures 1B, 7B**).

Our results also demonstrate that there is no obvious PCBP1 expression difference among tumors from stage I + II groups to stage III + IV groups, but the clear deference between tumors and normal tissues (**Figure 1F**). Similarly, p62 expression is evidently reduced in tumors, compared with the normal tissues, but no clear alteration among tumors, regarding tumor stages and metastatic states (**Figures 7B–E**), indicating that PCBP1-p62 signaling axis may play repressive role only in the tumor initiation stage. Thus, PCBP1 would be an early tumor suppressor gene. However, the PCBP1 knock out animal model would be useful to validate whether a spontaneous tumorigenesis will occur along with aging. In addition, PCBP1 not only functions via autophagy, but other manners to interfere with tumor progression, for example, tumor cell cycle modulation (Shi et al., 2018), as tumor cell cycle can be modulated by particular autophagy during cancer development and by therapy (Zheng et al., 2019).

Interplay between autophagy and apoptosis is another complex phenomenon. Recent study disclosed that p62 plays a bridge role in autophagy and apoptosis through its adaptor protein feature (Moscat and Diaz-Meco, 2009). p62, as a substrate of autophagy, is usually used as a marker to examine autophagy completion status. Because p62 contains several interaction domains to many signaling molecules for their proteasomal degradation (Bardag-Gorce et al., 2005; Moscat and Diaz-Meco, 2009; Zheng et al., 2019), monitoring p62 degradation

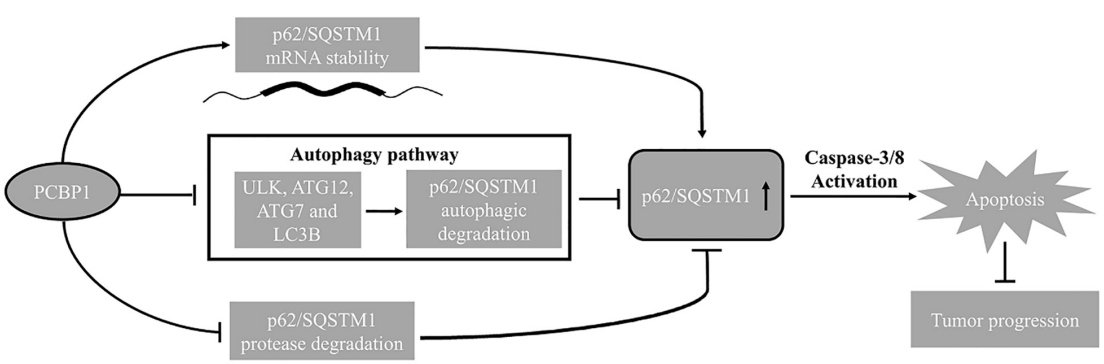


FIGURE 8 | Schematic model of PCBP1 in regulating p62/SQSTM1 to repress tumor progression.

thus cannot accurately evaluate the exact autophagic flux or autophagy outcome. In line with this notion, our results here show more evidence that p62 can be regulated by three different ways by PCBP1, including inhibitions of both autophagy- and proteasome-mediated p62 degradation, as well as increase p62 mRNA stability (**Figures 4**, **8**), which further indicates that p62 is not an ideal marker for detection of autophagy maturation. On the other hand, it is known that p62 can associate with caspase-8 and the subsequent apoptotic pathway (Jin et al., 2009; Yan et al., 2019). Likewise, our results here explain the reason why PCBP1 downregulation or depletion would be an important strategy to tumor cell survival by downregulations of p62 and caspase-8 (**Figure 7** and **Supplementary Figure 6**). Whether this hypothesis is valid, future work needs to be further studied in the appropriate tumor models.

Overall, our results provide a potential concept that autophagy inhibition could be an efficient strategy to induce malignant tumors to undergo apoptotic cell death, when PCBP1 depletion. In line with our results, it is reported that inhibition of autophagy with both 3-MA and CQ leads to apoptosis in both ovarian and colorectal cancer cells (Huang et al., 2014; Lu et al., 2014; Tang et al., 2015).

In summary, combining our current and previous results, we conclude that PCBP1 plays a coordinate role in inhibiting tumorigenesis via blocking autophagy, and in turn promoting tumor cell apoptosis (**Figure 8**). Therefore, inhibition of the autophagy in PCBP1-depleted tumors could be a promising therapeutic strategy.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Sun Yat-sen University Research Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HW, SY, and WZ conceived and designed the project. WZ, SZ, WG, and ZH conducted and collected the experiments and data interpretation. WZ, JK, and CH performed the statistical analysis. WZ, SZ, SY, and HW wrote and edited the manuscript. HW and SY supervised the project. All authors read and gave their approval for the final version of the manuscript.

FUNDING

This work was supported by the National Key R&D Program of China (No. 2018YFA0801005 to SY) and National Science Foundation of China (Nos. 31271481 and 81472730 to HW).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00930/full#supplementary-material

FIGURE S1 | Scores of PCBP1, p62, and caspase 8 expression based on IHC staining intensity.

FIGURE S2 | PCBP1 protein expression in the paired adjacent normal and tumor regions of human colorectal cancer. **(A)** Representative IHC staining of PCBP1 protein expression in the paired adjacent normal and tumor regions of colorectal. Scale bars are equal to 50 μm. **(B)** Statistical analysis of PCBP1 in 10 paired fresh normal and malignant tissues of colorectal tissues. The paired Student's *t*-test was carried out to check the group difference. *p*-values are shown.

FIGURE S3 | Endogenous PCBP1 knockdown increases autophagic flux. **(A)** Immunoblots of PCBP1 in the A2780, SKOV3, DLD-1, and HCT-116 cell lines. **(B)** Immunoblots of PCBP1 by the specific siRNA transfection. **(C)** Western blots of LC3 B in the indicated cells with GFP-PCBP1 (PCBP1) and the GFP-expressing control cells. GAPDH is used as a loading control. Protein bands' intensity ratio of LC3II to GAPDH were quantified and normalized, and shown under each lane, respectively.

FIGURE S4 | Genes precipitated in PCBP1 antibody-mediated RNA IP. The bound mRNAs are identified by RNA sequencing. Copy number of each gene is indicated in *Y*-axis, and the total number of PCBP1-bound mRNAs are shown in *X*-axis. –Autophagy-related ATG7 and ATG 12 are shown in the lower panel in red.

FIGURE S5 | (A,B) CCK8 cell proliferation analyses of A2780 **(A)** and DLD-1 **(B)** cells with exogenous PCBP1 in the absence or presence of autophagic inhibitor CQ at 50 μM. Data presented are mean ± SD. NS, No Significance; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$, $n = 3$.

FIGURE S6 | Relevance of PCBP1 to p62 and caspase 8 in ovary cancers. **(A)** Representative immunohistochemistry (IHC) analysis of p62 expression in the paired adjacent normal tissues and ovarian tumor samples. All images are photographed with 200X amplification. Scale bars in **Supplementary Figures 6A,E** equal to 100 μm. **(B)** Statistical comparison of p62 expression between primary and metastatic tumor tissues. **(C)** Representative immunohistochemistry (IHC) analysis of PCBP1, p62, and caspase 8 expressions in the same typical patient stages I–III. All Images are photographed with 200X amplification. Scale bars equal to 100 μm. **(D)** Correlationship of PCBP1 expression to p62. **(E)** Correlationship of PCBP1 expression to caspase-8. **(F)** Correlationship of p62 expression to caspase-8. **(G)** Correlationships among PCBP1, p62, and caspase 8 expressions in 90 ovary tumor samples.

TABLE S1 | Primers used for RT-PCR amplification in this study.

REFERENCES

Bardag-Gorce, F., Francis, T., Nan, L., Li, J., He, L. Y., French, B. A., et al. (2005). Modifications in P62 occur due to proteasome inhibition in alcoholic liver disease. *Life Sci.* 77, 2594–2602. doi: 10.1016/j.lfs.2005.04.020

Bialik, S., Dasari, S. K., and Kimchi, A. (2018). Autophagy-dependent cell death - where, how and why a cell eats itself to death. *J. Cell Sci.* 131:152. doi: 10.1242/jcs.215152

Bjorkoy, G., Lamark, T., Brech, A., Outzen, H., Perander, M., Overvatn, A., et al. (2005). p62/SQSTM1 forms protein aggregates degraded by autophagy and has

a protective effect on huntingtin-induced cell death. *J. Cell Biol.* 171, 603–614. doi: 10.1083/jcb.200507002

D'Arcy, M. S. (2019). Cell death: a review of the major forms of apoptosis, necrosis and autophagy. *Cell Biol. Int.* 43, 582–592. doi: 10.1002/cbin.11137

Degenhardt, K., Mathew, R., Beaudoin, B., Bray, K., Anderson, D., Chen, G., et al. (2006). Autophagy promotes tumor cell survival and restricts necrosis, inflammation, and tumorigenesis. *Cancer Cell* 10, 51–64. doi: 10.1016/j.ccr.2006.06.001

Fulda, S., and Debatin, K. M. (2006). Extrinsic versus intrinsic apoptosis pathways in anticancer chemotherapy. *Oncogene* 25, 4798–4811. doi: 10.1038/sj.onc.1209608

Galluzzi, L., Pietrocola, F., Bravo-San, P. J., Amaravadi, R. K., Baehrecke, E. H., Cecconi, F., et al. (2015). Autophagy in malignant transformation and cancer progression. *EMBO J.* 34, 856–880. doi: 10.15252/embj.201490784

Giles, K. M., Daly, J. M., Beveridge, D. J., Thomson, A. M., Voon, D. C., Furneaux, H. M., et al. (2003). The 3'-untranslated region of p21WAF1 mRNA is a composite cis-acting sequence bound by RNA-binding proteins from breast cancer cells, including HuR and poly(C)-binding protein. *J. Biol. Chem.* 278, 2937–2946. doi: 10.1074/jbc.M208439200

Guo, J., and Jia, R. (2018). Splicing factor poly(rC)-binding protein 1 is a novel and distinctive tumor suppressor. *J. Cell. Physiol.* 234, 33–41. doi: 10.1002/jcp.26873

Guo, J., Zhu, C., Yang, K., Li, J., Du, N., Zong, M., et al. (2017). Poly(C)-binding protein 1 mediates drug resistance in colorectal cancer. *Oncotarget* 8, 13312–13319. doi: 10.18632/oncotarget.14516

Huang, S., Okamoto, K., Yu, C., and Sinicrope, F. A. (2013). p62/sequestosome-1 up-regulation promotes ABT-263-induced caspase-8 aggregation/activation on the autophagosome. *J. Biol. Chem.* 288, 33654–33666. doi: 10.1074/jbc.M113.518134

Huang, Y. H., Al-Aidaroos, A. Q., Yuen, H. F., Zhang, S. D., Shen, H. M., Rozycka, E., et al. (2014). A role of autophagy in PTP4A3-driven cancer progression. *Autophagy* 10, 1787–1800. doi: 10.4161/auto.29989

Ishii, T., Hayakawa, H., Igawa, T., Sekiguchi, T., and Sekiguchi, M. (2018). Specific binding of PCBP1 to heavily oxidized RNA to induce cell death. *Proc. Natl. Acad. Sci. U.S.A.* 115, 6715–6720. doi: 10.1073/pnas.1806912115

Janku, F., McConkey, D. J., Hong, D. S., and Kurzrock, R. (2011). Autophagy as a target for anticancer therapy. *Nat. Rev. Clin. Oncol.* 8, 528–539. doi: 10.1038/nrclinonc.2011.71

Ji, F. J., Wu, Y. Y., An, Z., Liu, X. S., Jiang, J. N., Chen, F. F., et al. (2017). Expression of both poly r(C) binding protein 1 (PCBP1) and miRNA-3978 is suppressed in peritoneal gastric cancer metastasis. *Sci. Rep.* 7:15488. doi: 10.1038/s41598-017-15448-15449

Jin, Z., Li, Y., Pitti, R., Lawrence, D., Pham, V. C., Lill, J. R., et al. (2009). Cullin3-based polyubiquitination and p62-dependent aggregation of caspase-8 mediate extrinsic apoptosis signaling. *Cell* 137, 721–735. doi: 10.1016/j.cell.2009.03.015

Klionsky, D. J., Abdalla, F. C., Abeliovich, H., Abraham, R. T., Acevedo-Arozena, A., Adeli, K., et al. (2012). Guidelines for the use and interpretation of assays for monitoring autophagy. *Autophagy* 8, 445–544.

Kruidering, M., and Evan, G. I. (2000). Caspase-8 in apoptosis: the beginning of "the end"? *Iubmb. Life* 50, 85–90. doi: 10.1080/713803693

Levine, B. (2007). Cell biology: autophagy and cancer. *Nature* 446, 745–747. doi: 10.1038/446745a

Li, J., and Yuan, J. (2008). Caspases in apoptosis and beyond. *Oncogene* 27, 6194–6206. doi: 10.1038/onc.2008.297

Lu, Z., Baquero, M. T., Yang, H., Yang, M., Reger, A. S., Kim, C., et al. (2014). DIRAS3 regulates the autophagosome initiation complex in dormant ovarian cancer cells. *Autophagy* 10, 1071–1092. doi: 10.4161/auto.28577

Maiuri, M. C., Criollo, A., and Kroemer, G. (2010). Crosstalk between apoptosis and autophagy within the Beclin 1 interactome. *EMBO J.* 29, 515–516. doi: 10.1038/emboj.2009.377

Mizushima, N. (2007). Autophagy: process and function. *Genes Dev.* 21, 2861–2873. doi: 10.1101/gad.1599207

Moscat, J., and Diaz-Meco, M. T. (2009). p62 at the crossroads of autophagy, apoptosis, and cancer. *Cell* 137, 1001–1004. doi: 10.1016/j.cell.2009.05.023

Nishinakamura, H., Minoda, Y., Saeki, K., Koga, K., Takaesu, G., Onodera, M., et al. (2007). An RNA-binding protein alphaCP-1 is involved in the STAT3-mediated suppression of NF-kappaB transcriptional activity. *Int. Immunol.* 19, 609–619. doi: 10.1093/intimm/dxm026

Pankiv, S., Clausen, T. H., Lamark, T., Brech, A., Bruun, J. A., Outzen, H., et al. (2007). p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy. *J. Biol. Chem.* 282, 24131–24145. doi: 10.1074/jbc.M702824200

Ravikumar, B., Sarkar, S., Davies, J. E., Futter, M., Garcia-Arencibia, M., Green-Thompson, Z. W., et al. (2010). Regulation of mammalian autophagy in physiology and pathophysiology. *Physiol. Rev.* 90, 1383–1435. doi: 10.1152/physrev.00030.2009

Shi, H., Li, H., Yuan, R., Guan, W., Zhang, X., Zhang, S., et al. (2018). PCBP1 depletion promotes tumorigenesis through attenuation of p27(Kip1) mRNA stability and translation. *J. Exp. Clin. Cancer Res.* 37:187. doi: 10.1186/s13046-018-0840-841

Singh, S. S., Vats, S., Chia, A. Y., Tan, T. Z., Deng, S., Ong, M. S., et al. (2018). Dual role of autophagy in hallmarks of cancer. *Oncogene* 37, 1142–1158. doi: 10.1038/s41388-017-0046-46

Tang, Y., Li, M., Wang, Y. L., Threadgill, M. D., Xiao, M., Mou, C. F., et al. (2015). ART1 promotes starvation-induced autophagy: a possible protective role in the development of colon carcinoma. *Am. J. Cancer Res.* 5, 498–513.

Tao, M., Liu, T., You, Q., and Jiang, Z. (2020). p62 as a therapeutic target for tumor. *Eur. J. Med. Chem.* 193:112231. doi: 10.1016/j.ejmech.2020.112231

Tripathi, V., Sixt, K. M., Gao, S., Xu, X., Huang, J., Weigert, R., et al. (2016). Direct Regulation of Alternative Splicing by SMAD3 through PCBP1 Is Essential to the Tumor-Promoting Role of TGF-beta. *Mol. Cell.* 64, 549–564. doi: 10.1016/j.molcel.2016.09.013

Valente, G., Morani, F., Nicotra, G., Fusco, N., Peracchio, C., Titone, R., et al. (2014). Expression and clinical significance of the autophagy proteins BECLIN 1 and LC3 in ovarian cancer. *Biomed. Res. Int.* 2014:462658. doi: 10.1155/2014/462658

Villa, E., Proics, E., Rubio-Patino, C., Obba, S., Zunino, B., Bossowski, J. P., et al. (2017). Parkin-independent mitophagy controls chemotherapeutic response in cancer cells. *Cell Rep.* 20, 2846–2859. doi: 10.1016/j.celrep.2017.08.087

Wang, H., Vardy, L. A., Tan, C. P., Loo, J. M., Guo, K., Li, J., et al. (2010). PCBP1 suppresses the translation of metastasis-associated PRL-3 phosphatase. *Cancer Cell* 18, 52–62. doi: 10.1016/j.ccr.2010.04.028

Wang, X., Guo, J., Che, X., and Jia, R. (2019). PCBP1 inhibits the expression of oncogenic STAT3 isoform by targeting alternative splicing of STAT3 exon 23. *Int. J. Biol. Sci.* 15, 1177–1186. doi: 10.7150/ijbs.33103

Wang, Z., Yin, W., Zhu, L., Li, J., Yao, Y., Chen, F., et al. (2018). Iron drives T helper cell pathogenicity by promoting RNA-binding protein PCBP1-mediated proinflammatory cytokine production. *Immunity* 49, 80–92. doi: 10.1016/j.immuni.2018.05.008

Wu, H., Che, X., Zheng, Q., Wu, A., Pan, K., Shao, A., et al. (2014). Caspases: a molecular switch node in the crosstalk between autophagy and apoptosis. *Int. J. Biol. Sci.* 10, 1072–1083. doi: 10.7150/ijbs.9719

Yan, X. Y., Zhong, X. R., Yu, S. H., Zhang, L. C., Liu, Y. N., Zhang, Y., et al. (2019). p62 aggregates mediated Caspase 8 activation is responsible for progression of ovarian cancer. *J. Cell Mol. Med.* 23, 4030–4042. doi: 10.1111/jcmm.14288

Young, M. M., Takahashi, Y., Khan, O., Park, S., Hori, T., Yun, J., et al. (2012). Autophagosomal membrane serves as platform for intracellular death-inducing signaling complex (iDISC)-mediated caspase-8 activation and apoptosis. *J. Biol. Chem.* 287, 12455–12468. doi: 10.1074/jbc.M111.309104

Zhang, M. P., Zhang, W. S., Tan, J., Zhao, M. H., Lian, L. J., and Cai, J. (2017a). Poly r(C) binding protein (PCBP) 1 expression is regulated at the post-translation level in thyroid carcinoma. *Am. J. Transl. Res.* 9, 708–714.

Zhang, M. P., Zhang, W. S., Tan, J., Zhao, M. H., Lian, L. J., and Cai, J. (2017b). Poly r(C) binding protein (PCBP) 1 expression is regulated by the E3 ligase UBE4A in thyroid carcinoma. *Biosci. Rep.* 37:114. doi: 10.1042/BSR20170114

Zhang, T., Huang, X. H., Dong, L., Hu, D., Ge, C., Zhan, Y. Q., et al. (2010). PCBP-1 regulates alternative splicing of the CD44 gene and inhibits invasion in human hepatoma cell line HepG2 cells. *Mol. Cancer* 9:72. doi: 10.1186/1476-4598-9-72

Zhang, W., Shi, H., Zhang, M., Liu, B., Mao, S., Li, L., et al. (2016). Poly C binding protein 1 represses autophagy through downregulation of LC3B to promote tumor cell apoptosis in starvation. *Int. J. Biochem. Cell Biol.* 73, 127–136. doi: 10.1016/j.biocel.2016.02.009

Zhao, L., Wang, Z. G., Zhang, P., Yu, X. F., and Su, X. J. (2019). Poly r(C) binding protein 1 regulates posttranscriptional expression of the ubiquitin ligase TRIM56 in ovarian cancer. *IUBMB Life* 71, 177–182. doi: 10.1002/iub.1948

Zhao, Y., Chen, S., Gou, W. F., Xiao, L. J., Takano, Y., and Zheng, H. C. (2014). Aberrant Beclin 1 expression is closely linked to carcinogenesis, differentiation, progression, and prognosis of ovarian epithelial carcinoma. *Tumour Biol.* 35, 1955–1964. doi: 10.1007/s13277-013-1261-1266

Zheng, K., He, Z., Kitazato, K., and Wang, Y. (2019). Selective autophagy regulates cell cycle in cancer therapy. *Theranostics* 9, 104–125. doi: 10.7150/thno.30308

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 August 2020
doi: 10.3389/fgene.2020.00992

Identification of the Six-RNA-Binding Protein Signature for Prognosis Prediction in Bladder Cancer

Yucai Wu[1,2,3,4†], Yi Liu[5†], Anbang He[1,2,3,4], Bao Guan[1,2,3,4], Shiming He[1,2,3,4], Cuijian Zhang[1,2,3,4], Zhengjun Kang[5], Yanqing Gong[1,2,3,4], Xuesong Li[1,2,3,4*] and Liqun Zhou[1,2,3,4]*

[1] Department of Urology, Peking University First Hospital, Beijing, China, [2] Institute of Urology, Peking University, Beijing, China, [3] National Urological Cancer Center, Beijing, China, [4] Urogenital Diseases (Male) Molecular Diagnosis and Treatment Center, Peking University, Beijing, China, [5] Department of Urology, The Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou, China

RNA-binding proteins (RBPs) are a kind of gene regulatory factor that presents a significant biological effect in the initiation and development of various tumors, including bladder cancer (BLCA). However, the RBP-based prognosis signature for BLCA has not been investigated. In this study, we attempted to develop an RBP-based classifier to predict overall survival (OS) for BLCA based on transcriptome analysis. We extracted data of BLCA patients from The Cancer Genome Atlas database (TCGA) and UCSC Xena. Finally, a total of 398 cases without missing clinical data were enrolled and six RBPs (*FLNA*, *HSPG2*, *AHNAK*, *FASTKD3*, *POU5F1*, and *PCSK9*) associated with OS of BLCA were identified through univariate and multivariate Cox regression analysis. Online analyses and immunohistochemistry validated the prognostic value and expression of six RBPs. Risk scores were calculated to divide patients into high-risk and low-risk level, and patients in the high-risk group tended to have a poor prognosis. In addition, the receiver operating characteristic (ROC) curve analysis was performed to assess the prognostic value of RBPs, and the area under the curve (AUC) values were 0.711 and 0.706, respectively, in the training set and validating set. The findings were further validated in an external validation set. Subsequently, the 6-RBP-based signature and pathological stage were used to construct the nomogram to predict the 3- and 5-years OS of BLCA patients. Also, this 6-RBP-based signature was highly related to recurrence-free survival of BLCA. Weighted co-expression network analysis (WGCNA) combined with functional enrichment analysis contributed to study the potential pathways of six RBPs, including keratinocyte differentiation, RHO GTPases activate PNKs, epithelial tube morphogenesis, establishment or maintenance of cell polarity, and so on. In summary, the 6-RBP-based signature holds the potentiality to serve as a novel prognostic predictor of OS for BLCA.

Keywords: bladder cancer, RNA-binding proteins, overall survival, recurrence-free survival, prognosis

OPEN ACCESS

Edited by:
Yongsheng Kevin Li,
Harbin Medical University, China

Reviewed by:
Yunyan Gu,
Harbin Medical University, China
Zhaoying Yang,
Jilin University, China

***Correspondence:**
Yanqing Gong
yqgong@bjmu.edu.cn
Xuesong Li
pineneedle@sina.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *11 May 2020*
Accepted: *05 August 2020*
Published: *28 August 2020*

Citation:
Wu Y, Liu Y, He A, Guan B, He S, Zhang C, Kang Z, Gong Y, Li X and Zhou L (2020) Identification of the Six-RNA-Binding Protein Signature for Prognosis Prediction in Bladder Cancer. Front. Genet. 11:992. doi: 10.3389/fgene.2020.00992

Abbreviations: AUCs, Area under the curves; BLCA, Bladder cancer; circRNAs, Circular RNAs; DEGs, Differently expressed genes; DERBPs, DEGs coding for RNA-binding proteins; FD, Fold change; FDR, False discovery rate; GEO, Gene Expression Omnibus; HPA, Human Protein Atlas; HR, Hazard ratio; lncRNAs, Long non-coding RNAs; MIBC, Muscle-invasive bladder cancer; miRNAs, MicroRNAs; NMIBC, Non-muscle-invasive bladder cancer; OS, Overall survival; RBPs, RNA-binding proteins; RFS, Recurrence-free survival; ROC, Receiver operating characteristic; TCGA, The Cancer Genome Atlas; WGCNA, Weighted co-expression network analysis.

INTRODUCTION

Bladder cancer (BLCA) is the 10th most prevalent cancer and the most frequently diagnosed malignancy of the urinary system all over the world (Bray et al., 2018). It has been estimated that there will be 81,440 cases of newly diagnosed BLCA and 17,980 people will die for BLCA in 2020 in the United States (Siegel et al., 2020). Non-metastatic BLCA is separated into non-muscle-invasive bladder cancer (NMIBC) and muscle-invasive bladder cancer (MIBC) and approximately 70% of BLCA patients belong to NMIBC when initially diagnosed (Dobruch et al., 2016). MIBC patients have a more favorable prognosis than those with locally advanced and metastatic BLCA due to the limited effects of surgery on advanced BLCA. In addition, BLCA is the cancer with high recurrence and about half of patients after radical surgery relapse and present with metastases (Alfred Witjes et al., 2017). However, no specific symptoms appeared in the early stage of tumor, which makes it urgent to develop novel biomarkers to predict the survival of BLCA.

RNA-binding proteins (RBPs) are a kind of key factors regulating the process of tumorigenesis, and each step that led to the initiation of malignancy may involve one or more RBPs. Mechanisms of RBPs regulation have been identified in cancer cells, including alternative splicing, polyadenylation, stability, subcellular localization, and translation (Pereira et al., 2017). Post-transcriptional regulation is an essential way of promoting or suppressing oncogenesis. RBPs can interact with other proteins and coding or non-coding RNA to form the ribonucleoprotein complexes. For example, RBPs can interact with microRNAs (miRNAs) (Liang et al., 2020), long non-coding RNAs (lncRNAs) (Jiang et al., 2020), and circular RNAs (circRNAs) (Chen et al., 2019) to affect tumor progression.

Initial assessment of BLCA has been explored in recent times. In clinical practice, lncRNAs, miRNAs, and clinicopathological factors including TNM stage and lymph node status have been gradually used to assess BLCA prognosis. Recently increasing researches have revealed that RBPs were associated with the prognosis of patients (Busa et al., 2007; Vo et al., 2012). Therefore, we aimed to identify a number of RBPs as potential biomarkers based on transcriptome analysis to predict the outcome of BLCA. We constructed a 6-RBP-based classifier for OS by using the multivariable Cox regression, which could optimize the predictivity of the current TNM staging system. Patients with gene sequencing data from the GSE13507 database were adopted as the external validation. In addition, this 6-RBP-based classifier was also highly relevant to recurrence-free survival (RFS) in BLCA. Our results demonstrated that the 6-RBP-based classifier could be used as reliable prognostic predictors of BLCA survival and recurrence.

MATERIALS AND METHODS

Data Acquisition

The TCGA database was used to obtain transcriptome profiling data of tumor and normal tissues. Then, 19 normal samples and 411 BLCA samples were obtained. The matrix of mRNA expression was extracted separately by annotations using Gencode (GENCODE v 26) GTF file. Those genes whose expression was "zero" in 90% of BLCA patients were eliminated. Clinical data were downloaded from the UCSC Xena website[1]. To analyze the correlation of gene expression signatures with the OS of BLCA patients, we filtered out samples without clinical survival information; then, we selected a total of 398 patients and these patients were divided into training (n = 265) and validating set (n = 133) randomly at a 2:1 ratio for further analysis. Microarray study and its clinical information (GSE13507) in Gene Expression Omnibus (GEO) database[2] (n = 165) were extracted, profiled by the Illumina human-6 v2.0 expression BeadChip platform. A total of 1348 genes coding for RBPs including those with high confidence for RNA binding and those annotated as RNA binding in Ensembl were summarized from the published literature (Baltz et al., 2012; Castello et al., 2012; Kwon et al., 2013; Cunningham et al., 2015).

Analysis of Differentially Expressed Genes

We used the R package edgeR to obtain differentially expressed genes (DEGs), where $|\log2$ fold change (FC)$|$ >1, P <0.05, and false discovery rate (FDR) <0.05 were used as the cutoffs. Then, we filtered the DEGs coding for RNA-binding proteins (DERBPs). R package "heatmap" was performed to display the selected six DERBPs.

[1]https://xena.ucsc.edu/

[2]http://www.ncbi.nlm.nih.gov/geo/

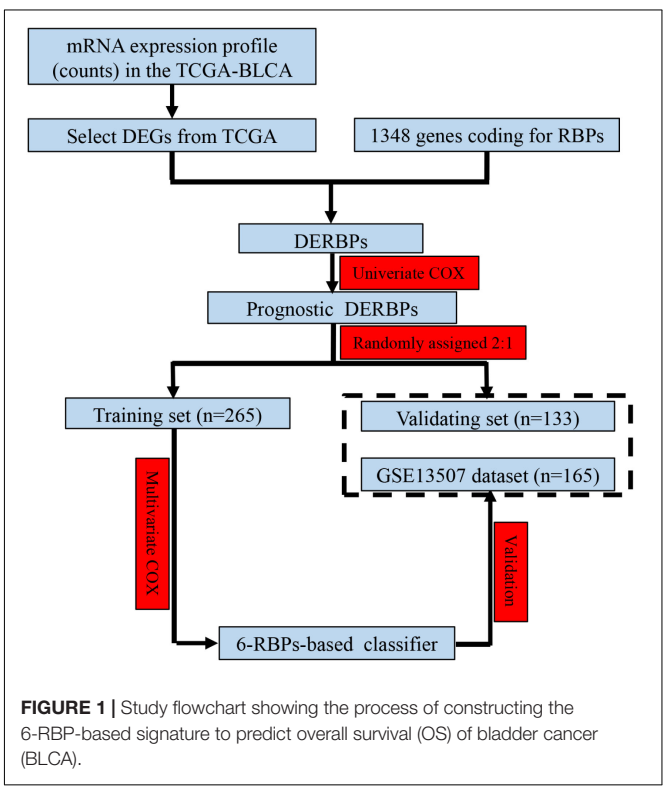


FIGURE 1 | Study flowchart showing the process of constructing the 6-RBP-based signature to predict overall survival (OS) of bladder cancer (BLCA).

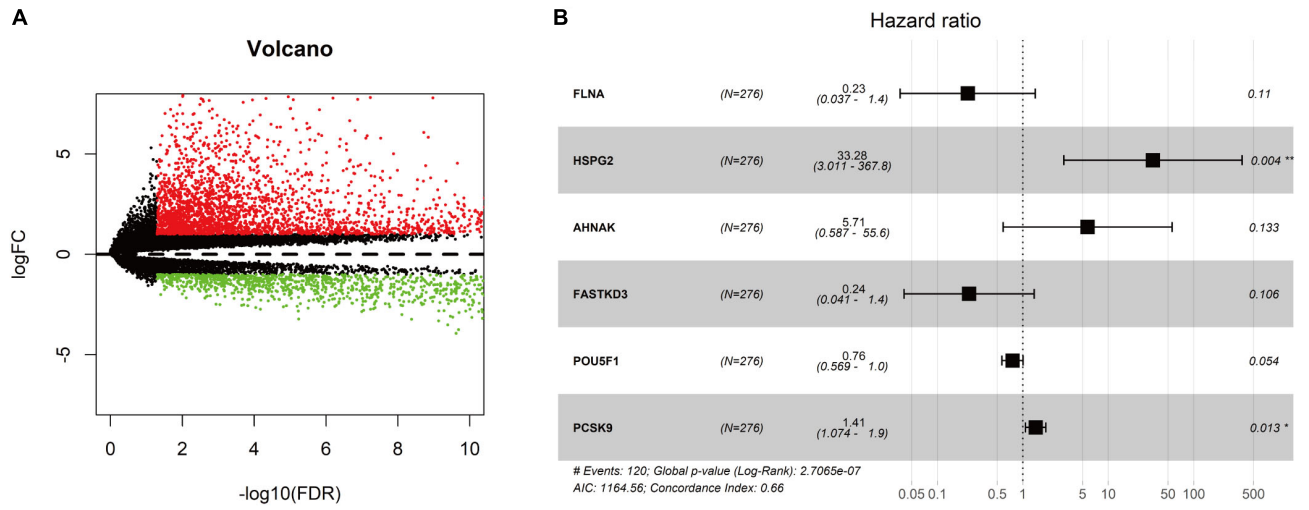


FIGURE 2 | Screening for differentially expressed genes in bladder cancer (BLCA). **(A)** Volcano plot of differentially expressed genes in the TCGA-BLCA cohort. The red dot represented the upregulated gene, while the green dot represented the downregulated gene. **(B)** Forest plot exhibited the hazard ratios (HRs) with 95% confidence interval (95% CI) of prognostic RBPs in BLCA on the basis of the multivariate Cox regression result.

Data Processing and Risk Score Calculation

The DERBPs were subjected to univariable Cox regression analysis to select DERBPs that were associated with OS of BLCA patients. We selected those DERBPs with P <0.001 into multivariable Cox regression to obtain the coefficients. Then, six DERBPs significantly correlated with OS were identified to build up the prediction model weighted by their coefficients. A risk score formula for OS was constructed, and each patient had been assigned a risk score according to this risk score formula that was a linear combination of the expression levels of the significant DERBPs weighted by their respective Cox regression coefficients. Then, we divided patients into low-risk group and high-risk group according to the median risk score.

Weighted Co-expression Network Analysis (WGCNA)

Considering our risk score model was built based on the expression of six RBPs, we constructed a weight co-expression network of risk score gene with DEGs in BLCA to explore the biological function by the R package "WGCNA" (Langfelder and Horvath, 2008). We selected 3 as the soft thresholding power to produce a scale-free network and the enrolled genes were hierarchically clustered into 16 modules, where the red module was found to be most relevant to risk score.

Pathway Enrichment Analysis

In order to explore the potential functions of the 6-RBP signature, genes in the red module were picked up for enrichment analysis. Pathway enrichment was conducted using an online web tool "Metascape[3]." The significance threshold of FDR for significantly enriched biological processes and pathways was set at 0.05.

Statistical Analysis

We use Chi-squared test or Fisher's exact test to measure the difference between training and validating sets and the relationship between clinical data and risk score. Both univariable and multivariable Cox regression analysis were performed using the R package "survival." The Kaplan–Meier survival curve was drawn to demonstrate the relationship between DERBPs and OS or RFS. The log-rank test was constructed to test the significance of the difference between the two groups. ROC analysis was performed to measure prognostic accuracy. All statistical tests were two-sided, and P <0.05 was considered statistically significant. All analyses were performed in SPSS version 25.0 (SPSS Inc., Chicago, IL, United States) or R version 3.5.2[4] with the following packages: "edgeR" (Robinson et al., 2010), "glmnet," "gplot," and "survivalROC."

RESULTS

Data Source and Processing

As shown in **Figure 1**, we obtained 19 normal samples and 411 BLCA samples from TCGA database, and 4456 DEGs with | Log2FC| >1 and FDR <0.05 were identified using edgeR (**Figure 2A**). A total of 1348 genes coding known or predicted RBPs were matched with the 4456 DEGs and then 109 RBPs remained. After that, univariate Cox regression was performed to choose factors to predict prognostic of patients and 12 DERBPs with P <0.001 were retained for further analysis. Clinical characters of BLCA patients were downloaded from the UCSC database, and these cases were randomly divided into training set (n = 265) and validating set (n = 133) at a 2:1 ratio. There were no significant differences in age, gender, pathological

[3]http://metascape.org/

[4]http://www.r-project.org/

TABLE 1 | Clinical features of BLCA patients in the training and validating sets.

Features	Training set (*n* = 265)	Validating set (*n* = 133)	Pearson χ^2	*P*
Age (years), *n* (%)				
≤70	138 (52.1)	72 (54.1)		
>70	127 (47.9)	61 (45.9)	0.151	0.698
Gender, *n* (%)				
Male	189 (71.3)	100 (75.2)		
Female	76 (28.7)	33 (24.8)	0.666	0.414
Pathological stage, *n* (%)				
I + II	85 (32.1)	41 (30.8)		
III + IV	180 (67.9)	92 (69.2)	0.064	0.801
Histologic grade, *n* (%)				
Low	13 (4.9)	7 (5.3)		
High	252 (95.1)	126 (94.7)	0.024	0.878
Diagnosis subtype, *n* (%)				
Non-papillary	175 (66.0)	96 (72.2)		
Papillary	90 (34.0)	37 (27.8)	1.538	0.215

BLCA, Bladder cancer. Bold values are significant to p < 0.05.

stage, histologic grade, and diagnosis subtype between two sets (**Table 1**). Then, we identified six DERBPs, which were strongly associated with OS of BLCA by multivariate Cox regression analysis in the training set (**Figure 2B**), and the detailed information of these RBPs including *FLNA (Filamin A)*, *HSPG2 (Heparan Sulfate Proteoglycan 2)*, *AHNAK (AHNAK Nucleoprotein)*, *FASTKD3 (FAST Kinase Domains 3)*, *POU5F1 (POU Class 5 Homeobox 1)*, and *PCSK9 (Proprotein Convertase Subtilisin/Kexin Type 9)* were listed in **Table 2**. Among these genes, higher expression of *HSPG2*, *AHNAK*, and *PCSK9* was associated with decreased survival. On the contrary, higher expression of *FLNA*, *FASTKD3*, and *POU5F1* was related to increased survival.

Validation the Prognostic Value and Expression of Six RBPs

To further explore the prognostic value of six RBPs in BLCA, the Kaplan–Meier plotter was used to determine the relationship between six RBPs and OS. Five of the six RBPs *(AHNAK*, *HSPG2*, *PCSK9*, *POU5F1*, and *FASTKD3*) were identified. Results of log-rank test demonstrated that the high expressions of *AHNAK*, *HSPG2*, and *PCSK9* were associated with the low OS, while the high expression of *POU5F1* and *FASTKD3* was associated with the high OS of BLCA patients (**Figure 3**). To further validate the expression of these RBPs in BLCA, we analyzed immunohistochemistry data from the Human Protein Atlas (HPA) database[5] to show that *FLNA*, *FASTKD3*, and *POU5F1* were significantly decreased in BLCA compared with normal urinary bladder tissue (**Figures 4A–C**). Besides, the staining level of *HSPG2* was increased in BLCA (**Figure 4E**). However, the staining level of *AHNAK* was relatively reduced in normal urinary bladder tissue and the result of *PCSK9* protein expression was not detectable (**Figure 4D**). These results showed that expression of each of the six RBPs was related to prognosis of BLCA.

Development and Validation a 6-RBP-Based Classifier to Predict OS of BLCA

To assess the ability of the 6-RBP-based model predicting survival of BLCA, we created a risk score according to the expression of six RBPs as follows: Risk score = (3.50 * expression value of *HSPG2*) + (1.74 * expression value of *AHNAK*) + (0.35 * expression value of *PCSK9*) - (1.49 * expression value of *FLNA*) - (1.45 * expression value of *FASTKD3*) - (0.28 * expression value of *POU5F1*). Then, we calculated risk score according to this formula and cases were divided into high-risk and low-risk group based on the cutoff of median risk score (**Figure 5A**). The mortality was higher in the high-risk group than that in the low-risk group [HR: 2.274 (95% CI: 1.562–3.312), $p < 0.001$]. Moreover, *HSPG2*, *AHNAK*, and *PCSK9* were highly expressed in the high-risk group, while *FLNA*, *FASTKD3*, and *POU5F1* were highly expressed in the low-risk group. Results in the validating set were consistent with findings described above (**Figure 5B**). Kaplan–Meier curves revealed that patients in the high-risk group had shorter OS than those in the low-risk group ($p < 0.001$) in the training set (**Figure 5C**), and this result was further confirmed in the validating set ($p < 0.001$) (**Figure 5D**).

We also extracted BLCA samples from the GSE31507 database ($n = 165$) to validate the ability of the 6-RBP-based classifier predicting OS of BLCA. Results were compatible with that in the training and validating set derived from TCGA database. Patients in the high-risk group had shorter OS than those in the low-risk group ($p = 0.014$) (**Figure 6A**). We also found that our 6-RBP-based classifier performed well in predicting 5-year OS of BLCA

[5]https://www.proteinatlas.org/

TABLE 2 | Six RBPs significantly associated with the OS of BLCA patients in the training set.

Gene symbol	ENSG ID	Coefficient	*P*	HR	HR (95% CI)
FLNA	ENSG00000196924	−1.48536	0.11	0.23	0.037–1.4
HSPG2	ENSG00000142798	3.504967	**0.004**	33.28	3.011–367.8
AHNAK	ENSG00000124942	1.743093	0.133	5.71	0.587–55.6
FASTKD3	ENSG00000124279	−1.44735	0.106	0.24	0.041–1.4
POU5F1	ENSG00000204531	−0.27992	**0.054**	0.76	0.569–1.0
PCSK9	ENSG00000169174	0.346457	**0.013**	1.41	1.074–1.9

BLCA, Bladder cancer; OS, Overall survival; HR, Hazard ratio; 95% CI, 95% Confidence interval. Bold values are significant to p < 0.05.

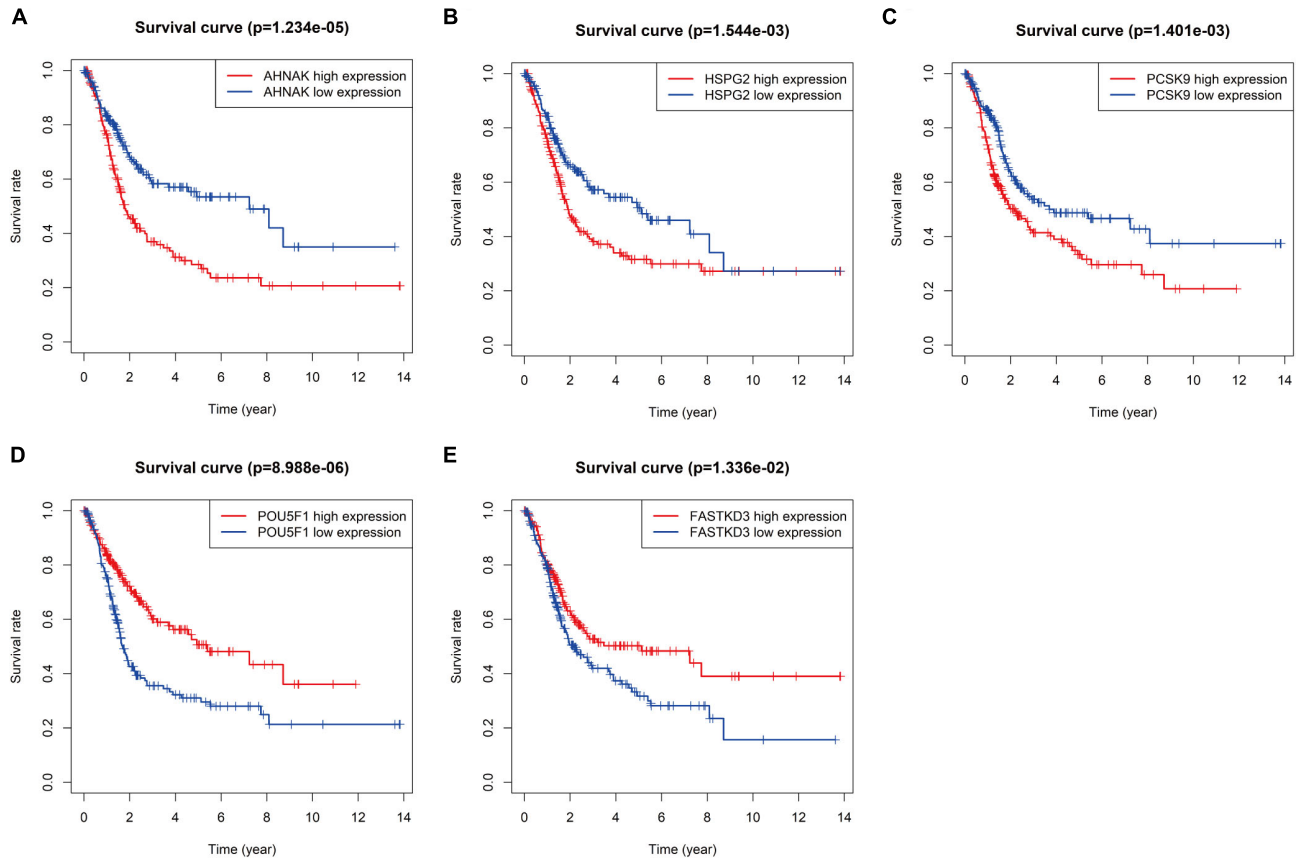


FIGURE 3 | Kaplan–Meier analysis for overall survival (OS) of bladder cancer (BLCA) patients based on the mRNA expression of *AHNAK* **(A)**, *HSPG2* **(B)**, *PCSK9* **(C)**, *POU5F1* **(D)**, and *FASTKD3* **(E)**.

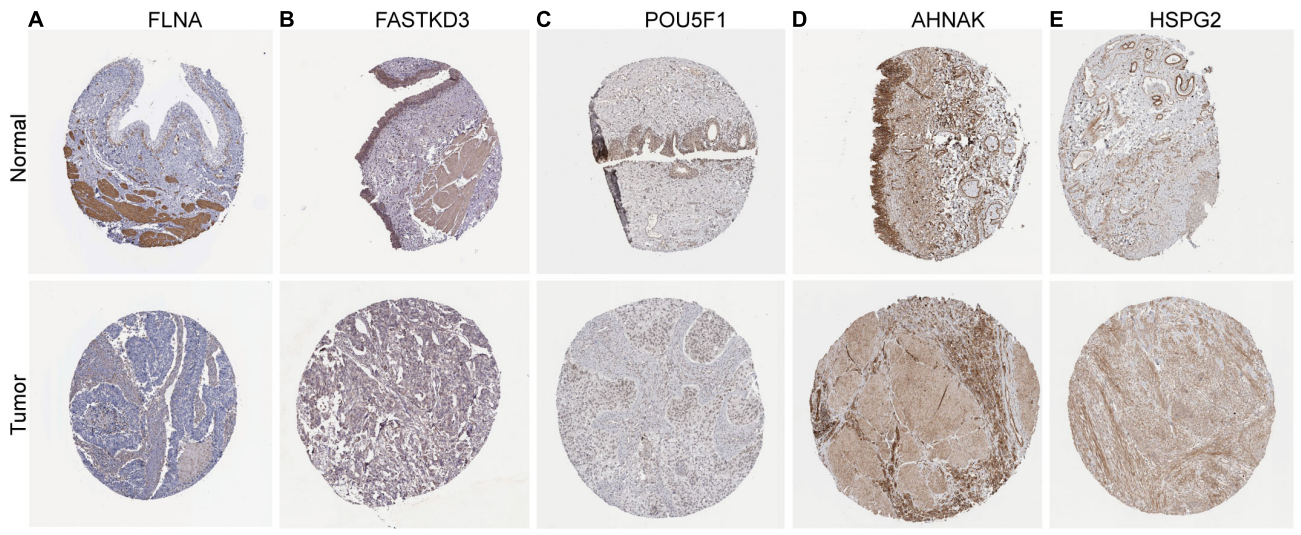


FIGURE 4 | Validation of the expression of RNA-binding proteins (RBPs) in bladder cancer (BLCA) and normal tissues in the Human Protein Atlas (HPA) database. **(A)** *FLNA*, **(B)** *FASTKD3*, **(C)** *POU5F1*, **(D)** *AHNAK*, and **(E)** *HSPG2*.

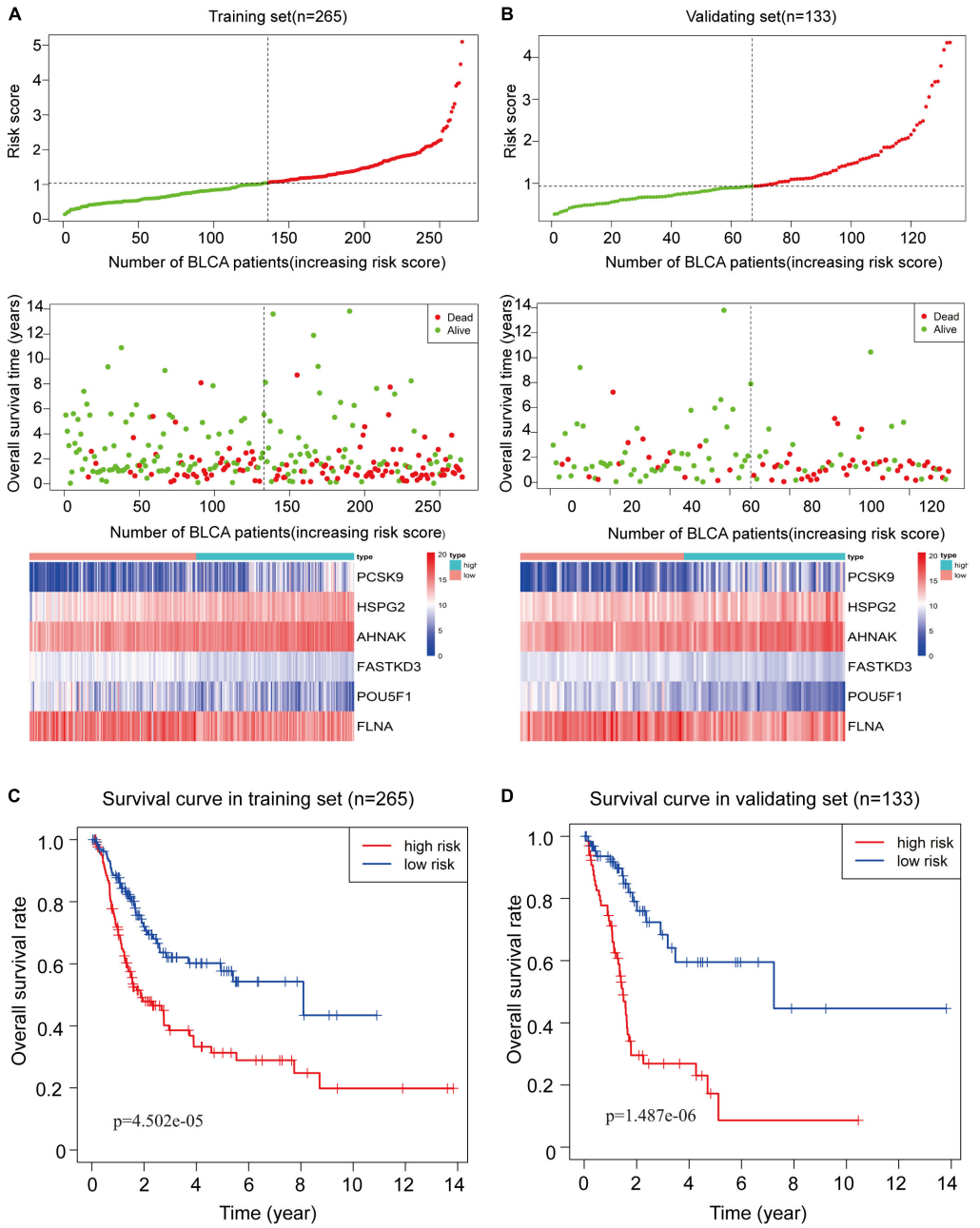


FIGURE 5 | Construction of the 6-RBP-based risk model of bladder cancer (BLCA). **(A)** The 6-RBP-based risk score distribution, living status of BLCA patients, and heatmap of six RBP expression profiles in the training set. **(B)** The 6-RBP-based risk score distribution, living status of BLCA patients, and heatmap of six RBP expression profiles in the validating set. **(C,D)** Kaplan–Meier analysis for overall survival (OS) of BLCA patients based on the risk stratification in the training set **(C)** and validating set **(D)**.

(AUCs = 0.721) (**Figure 6B**). The mortality was higher in the high-risk group than that in the low-risk group (**Figure 6C**).

Correlation Between RBPs Classifier and Clinicopathologic Characteristics

As shown in **Table 3**, clinicopathologic characteristics age (p = 0.032), pathological stage (p < 0.001), histologic grade (p = 0.011), and diagnosis subtype (p = 0.002) showed significant differences between the high-risk group and low-risk group in the training set. Only age (p = 0.046), pathological stage (p = 0.002), and histologic grade (p = 0.007) displayed distinct differences in the validating set. Patients with high pathological stage and histologic grade were prone to get a high-risk score in two sets (**Figure 7**).

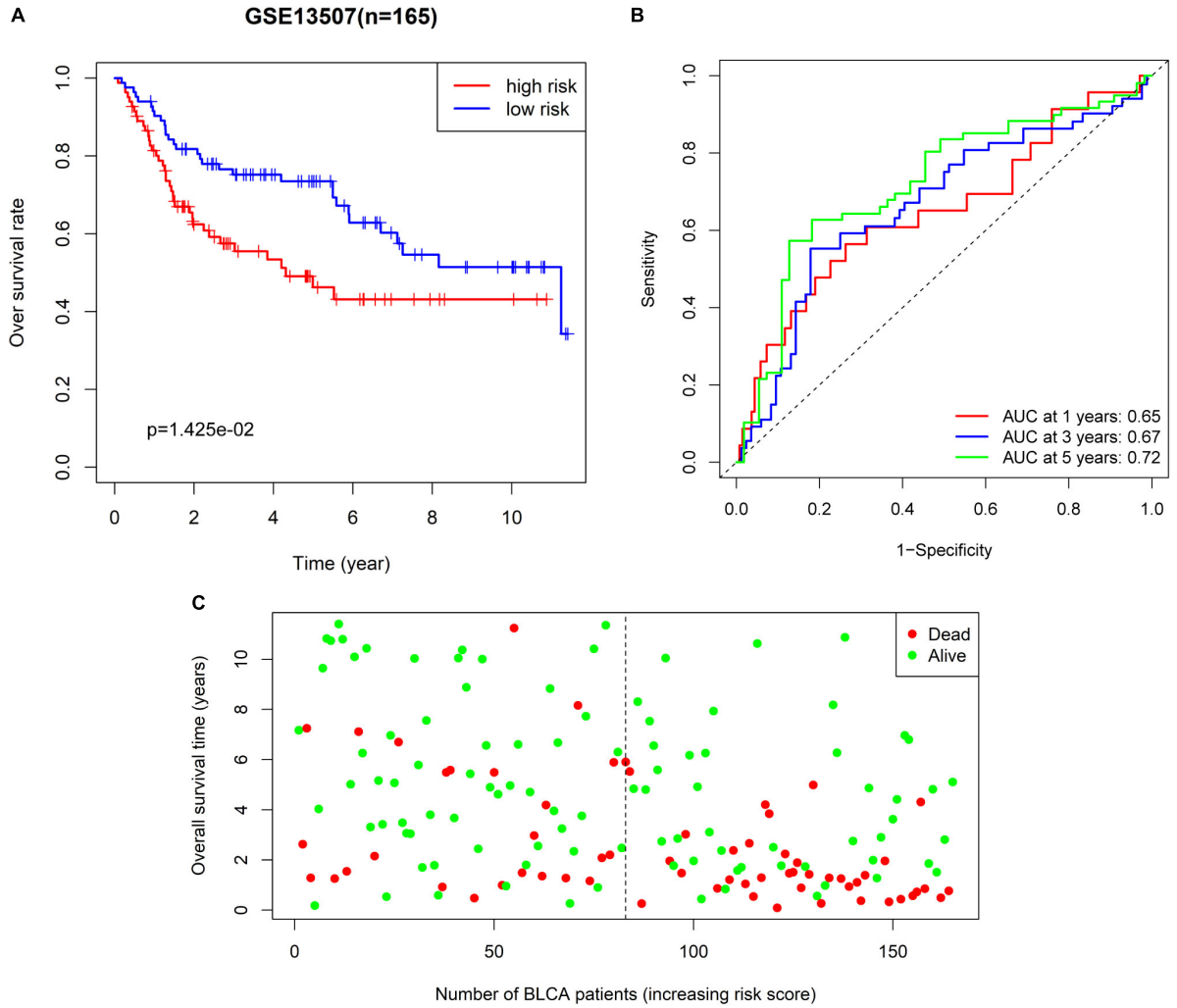


FIGURE 6 | Further validation of the 6-RBP-based risk model in the GSE13507 dataset. **(A)** Kaplan–Meier analysis for overall survival (OS) of bladder cancer (BLCA) patients based on the risk stratification. **(B)** Receiver operating characteristic (ROC) analysis for OS prediction including 1–, 3–, and 5–years of BLCA patients. **(C)** The living status of BLCA patients.

Prognostic Value of RBPs Classifier for Assessing Overall Survival

The 6-RBP-based classifier, age, and pathological stage were significantly related to the OS in the univariate Cox regression analysis. After the multivariate Cox regression analysis of the abovementioned factors, the 6-RBP-based classifier model and pathological stage were retained to be dependable factors for OS in the training set. Except for age, similar results were observed in the validating set (**Table 4**). Our result showed that the 6-RBP-based classifier was an independent prognostic factor for OS in BLCA in two sets.

In order to evaluate the capabilities of the 6-RBP-based signature to predict OS of BLCA, we plotted the ROC curves and AUC was also calculated in both cohorts. AUC values in the training set and validating set were 0.711 and 0.706, respectively, which showed that the RBP-based classifier model had an obvious better predictive accuracy compared with the TNM staging (0.670 and 0.674, respectively, in the training set and validating set) (**Figures 8A,B**). In consideration of the role of TNM staging in clinical practice, we combined the RBP-based model and TNM staging to predict OS of BLCA. AUC values of this joint prediction model in the training and validating set were 0.753 and 0.728, indicating that this model was more accurate than models enrolled in RBPs or TNM staging solely. Subsequently, the 6-RBP-based risk score and pathological stage were used to construct the nomogram to predict the 3- and 5-year OS of BLCA patients (**Figure 8C**).

Prognostic Value of the RBP-Based Classifier for Assessing RFS

We further explored whether this 6-RBP-based classifier was related to RFS of BLCA. As shown in **Table 5**, univariate and multivariate Cox regression analysis was conducted to identify prognostic factors for RFS. Finally, outcomes of univariate and

TABLE 3 | Clinicopathological characteristics of the 6-marker-based classifier with OS in the training set and validating set.

Features	Training set (*n* = 265)			Validating set (*n* = 133)		
	Low risk (*n* = 133)	High risk (*n* = 132)	*P*	Low risk (*n* = 67)	High risk (*n* = 66)	*P*
Age (years), *n* (%)						
≤70	78 (29.4)	60 (22.6)		42 (31.6)	30 (22.6)	
>70	55 (20.8)	72 (27.2)	**0.032**	25 (18.8)	36 (27.1)	**0.046**
Gender, *n* (%)						
Male	101 (38.1)	88 (33.2)		52 (39.1)	48 (36.1)	
Female	32 (12.1)	44 (16.6)	0.095	15 (11.3)	18 (13.5)	0.514
Pathological stage, *n* (%)						
I + II	58 (21.9)	27 (10.2)		29 (21.8)	12 (9.0)	
III + IV	75 (28.3)	105 (39.6)	**<0.001**	38 (28.6)	54 (40.6)	**0.002**
Histologic grade, *n* (%)						
Low	11 (4.2)	2 (0.8)		7 (5.3)	0 (0)	
High	122 (46.0)	130 (49.1)	**0.011**	60 (45.1)	66 (49.6)	**0.007**
Diagnosis subtype, *n* (%)						
Non-papillary	76 (28.7)	99 (37.4)		45 (33.8)	51 (38.3)	
Papillary	57 (21.5)	33 (12.5)	**0.002**	22 (16.5)	15 (11.3)	0.193

OS, Overall survival. Bold values are significant to p < 0.05.

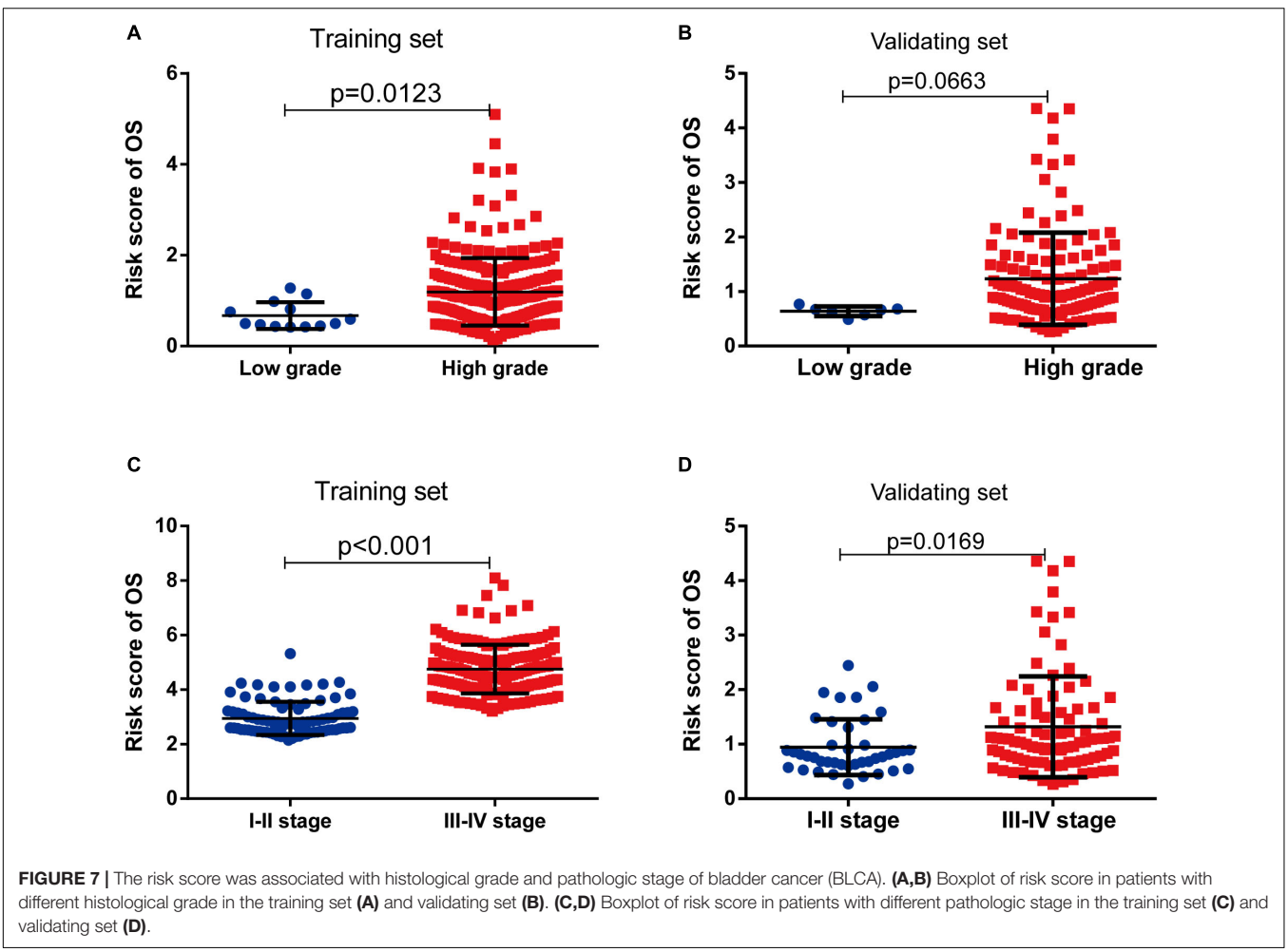


FIGURE 7 | The risk score was associated with histological grade and pathologic stage of bladder cancer (BLCA). **(A,B)** Boxplot of risk score in patients with different histological grade in the training set **(A)** and validating set **(B)**. **(C,D)** Boxplot of risk score in patients with different pathologic stage in the training set **(C)** and validating set **(D)**.

TABLE 4 | Univariate and multivariate Cox regression analysis of the 6-marker-based classifier with OS in the training set and validating set.

Features	Univariate COX		Multivariate COX	
	HR (95% CI)	*P*	HR (95% CI)	*P*
Training set				
Age (>70 vs. ≤70)	1.519 (1.049, 2.200)	**0.027**		
Gender (Male vs. Female)	0.831 (0.562, 1.231)	0.356		
Pathological stage (III + IV vs. I + II)	2.210 (1.387, 3.521)	**0.001**	1.826 (1.135, 2.937)	**0.013**
Histologic grade (High vs. Low)	1.942 (0.477, 7.899)	0.354		
Diagnosis subtype (Papillary vs. Non-papillary)	0.674 (0.441, 1.028)	0.067		
6-marker-based classifier (High risk vs. Low risk)	1.738 (1.435, 2.103)	**<0.001**	1.636 (1.342, 1.995)	**<0.001**
Validating set				
Age (>70 vs. ≤70)	1.853 (1.106, 3.105)	**0.019**	1.861 (1.084, 3.195)	**0.024**
Gender (Male vs. Female)	0.916 (0.508, 1.649)	0.769		
Pathological stage (III + IV vs. I + II)	2.573 (1.301, 5.089)	**0.007**	2.497 (1.234, 5.052)	**0.011**
Histologic grade (High vs. Low)	1.278 (0.019, 8.561)	0.394		
Diagnosis subtype (Papillary vs. Non-papillary)	0.571 (0.280, 1.163)	0.123		
6-marker-based classifier (High risk vs. Low risk)	3.781 (2.117, 6.754)	**<0.001**	3.166 (1.751, 5.727)	**<0.001**

OS, Overall survival; HR, Hazard ratio; 95% CI, 95% Confidence interval. Bold values are significant to $p < 0.05$.

multivariate analysis indicated that pathological stage and the 6-RBP-based classifier were independent risk factors of RFS in BLCA patients. Survival analysis revealed that the RFS of patients in the high-risk group were considerably shorter than that in the low-risk group (**Figure 8D**). Our results demonstrated that this 6-RBP-based classifier could also be used as a reliable prognostic predictor of BLCA recurrence.

Pathway Enrichment Analysis of DERBPs

To explore the biological function of the 6-RBP signature in BLCA, the WGCNA method was performed to choose genes associated with risk score. Sixteen modules were identified by hierarchical clustering analysis, and the red module was significantly associated risk score (**Figure 9A**). In addition, genes in the red module served to perform pathway enrichment analysis. Results of enrichment analysis showed that genes were mostly enriched in keratinocyte differentiation, and RHO GTPases activate PNKs, epithelial tube morphogenesis, establishment or maintenance of cell polarity, and so on, suggesting that these pathways were correlated with the disease progression of BLCA with high-risk score (**Figures 9B,C**).

DISCUSSION

RBPs were regarded acting as amplifiers of oncogenic driver mutations. Increasing literature demonstrated that RBPs were of vital importance in the initiation, development, and recurrence of multi-malignancies. For instance, *LIN28B* overexpression promoted tumorigenesis and metastasis of colon cancer via repressing the level of let-7 microRNAs (King et al., 2011). RBPs also played a critical role in the initiation of BLCA. It was reported that *HuR* was upregulated in the BLCA tissue compared with adjacent normal tissue and it could promote BLCA progression by competitively binding to the long non-coding HOX transcript antisense RNA with miR-1 (Yu et al., 2017). RBPs also exhibit potentiality as novel biomarkers. Pancreatic ductal adenocarcinoma patients with higher levels of *ESRP1* showed longer survival than those with low *ESRP1* expression (Ueda et al., 2014). In conclusion, RBPs can regulate the biology of cancer and hold potential as novel biomarkers.

Diverse models to predict the outcome of BLCA have been created, including miRNA-based signature (Hofbauer et al., 2018; Yin et al., 2019), clinical character-based nomogram (Zhang et al., 2019), and lncRNA-based model (Wang et al., 2020). Parts of them performed well in predicting the survival of BLCA. However, no RBP-based classifier for predicting survival in BLCA has been established. RBPs are a subset molecule of regulating progression and development of malignancies. Considering the limited capabilities of single RBP for prediction prognosis, we constructed a predicted model based on mRNA expression of six RBPs by univariate and multivariate Cox regression analysis. Patients were divided into two categories based on the median-risk score and high-risk patients have shorter survival than low-risk patients, suggesting that our model displayed powerful strength to forecast the OS for BLCA. Clinicopathologic characteristics including age, pathological stage, and histologic grade were associated with the expression of six RBPs and patients with high stage or grade tended to get high-risk score. Our 6-RBP-based classifier with AUCs being 0.711 and 0.706 in the training and validating sets presented a strong ability to predict OS of BLCA. In addition, TNM stage is the most commonly used index to assess the treatment way and outcome of BLCA. We also combined the 6-RBP-based signature together with TNM stage to assess the prognosis and the AUCs showed that this model was more accurate than the 6-RBP-based model. Furthermore, we explored the efficiency of the 6-RBP-based classifier in predicting RFS of BLCA. Results showed that pathological stage and the 6-RBP-based classifier were independent factors of RFS in BLCA patients and patients in the low-risk group showed a significantly longer RFS than those in the high-risk group.

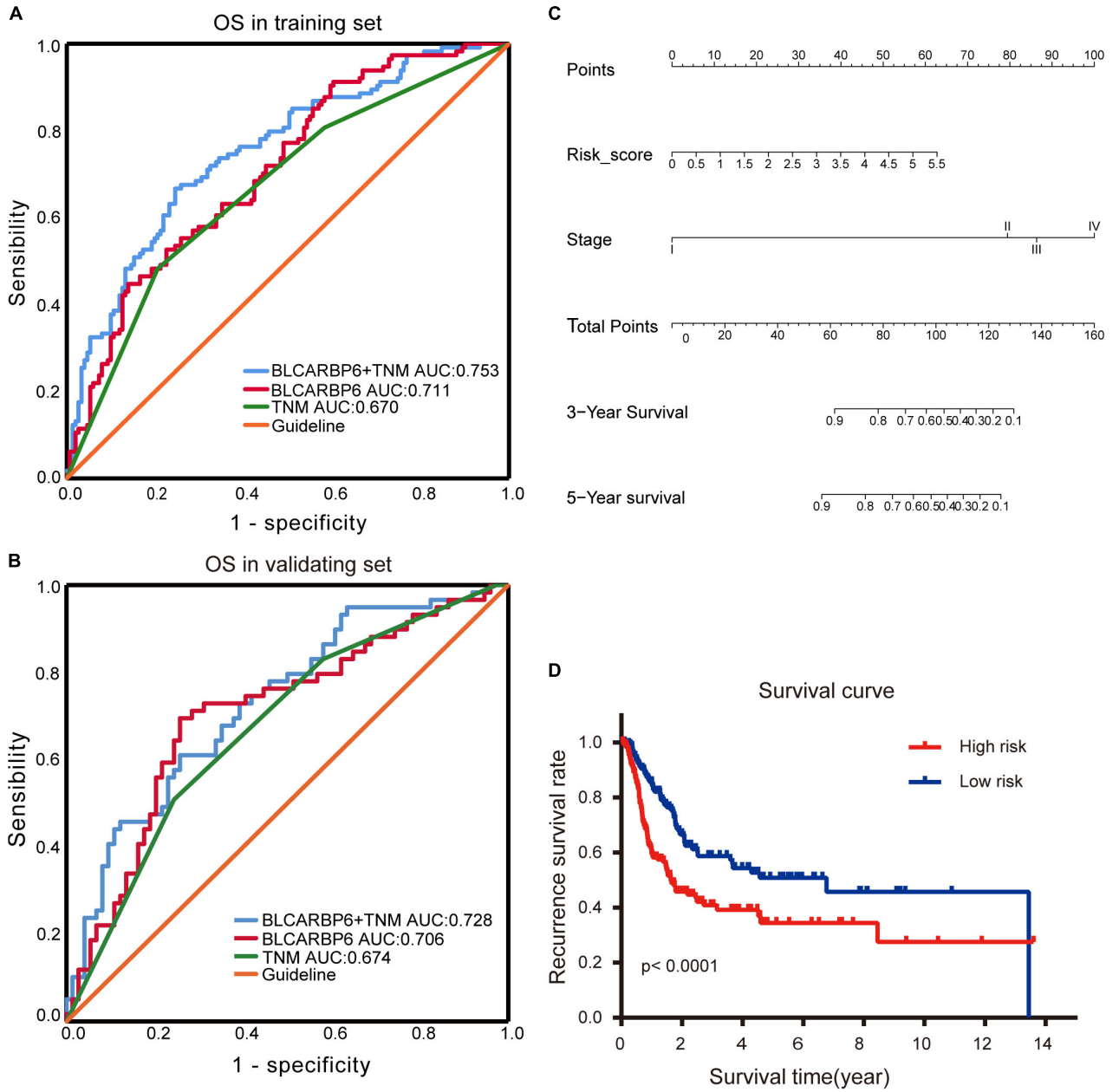


FIGURE 8 | Estimate the predictive ability of RBP-based classifier to predict survival in bladder cancer (BLCA) patients. **(A,B)** The receiver operating characteristic (ROC) for TNM stage only, the RBP-based risk score (BLCARBP6), and the RBP-based risk score combined with TNM stage for overall survival (OS) in the training set **(A)** and validating set **(B)**. **(C)** Nomogram to predict the 3- and 5-years OS. **(D)** Kaplan–Meier analysis for recurrence-free survival (RFS) of BLCA patients based on the risk stratification.

Six prognosis-related RBPs including *FLNA*, *HSPG2*, *AHNAK*, *FASTKD3*, *POU5F1*, and *PCSK9* were selected to build a classifier. It has been confirmed that *FLNA* was downregulated in BLCA tissue and overexpression of *FLNA* repressed migration, invasion, and migration of BLCA by regulating autophagy (Wang et al., 2018). *AHNAK* downregulated in BLCA performed accurately on discriminating between benign urothelial lesion and bladder urothelial carcinoma using voided-urine liquid-based cytology (Lee et al., 2018). In addition, low expression of *POU5F1* was associated with shorter cancer-related survival and might be a novel biomarker for BLCA (Chang et al., 2008). However, functions of *HSPG2*, *FASTKD3*, and *PCSK9* in BLCA have not been explored yet. These confirmed or predicted prognosis-related RBPs support evidence that our model is capable of assessing the outcome of BLCA.

In order to explore the biological function of the 6-RBP signature, we performed WGCNA and pathway enrichment analysis. Results showed that those genes relevant to risk

TABLE 5 | Univariate and multivariate Cox regression analysis of the 6-marker-based classifier with RFS in TCGA database.

Features	Univariate COX		Multivariate COX	
	HR (95% CI)	*P*	HR (95% CI)	*P*
Age (>70 vs. ≤70)	1.117 (0.802, 1.557)	0.513		
Gender (Male vs. Female)	0.857 (0.596, 1.232)	0.404		
Pathological stage (III + IV vs. I + II)	2.879 (1.810, 4.578)	**<0.001**	2.623 (1.643, 4.187)	**<0.001**
Histologic grade (High vs. Low)	3.345 (0.826, 13.544)	0.091		
Diagnosis subtype (Papillary vs. Non-papillary)	0.642 (0.431, 0.956)	**0.029**		
6-marker-based classifier (high risk vs. low risk)	1.911 (1.365, 2.677)	**<0.001**	1.696 (1.207, 2.382)	**0.002**

RFS, Recurrence-free survival; TCGA, The Cancer Genome Atlas; HR, Hazard ratio; 95% CI, 95% Confidence interval. Bold values are significant to p < 0.05.

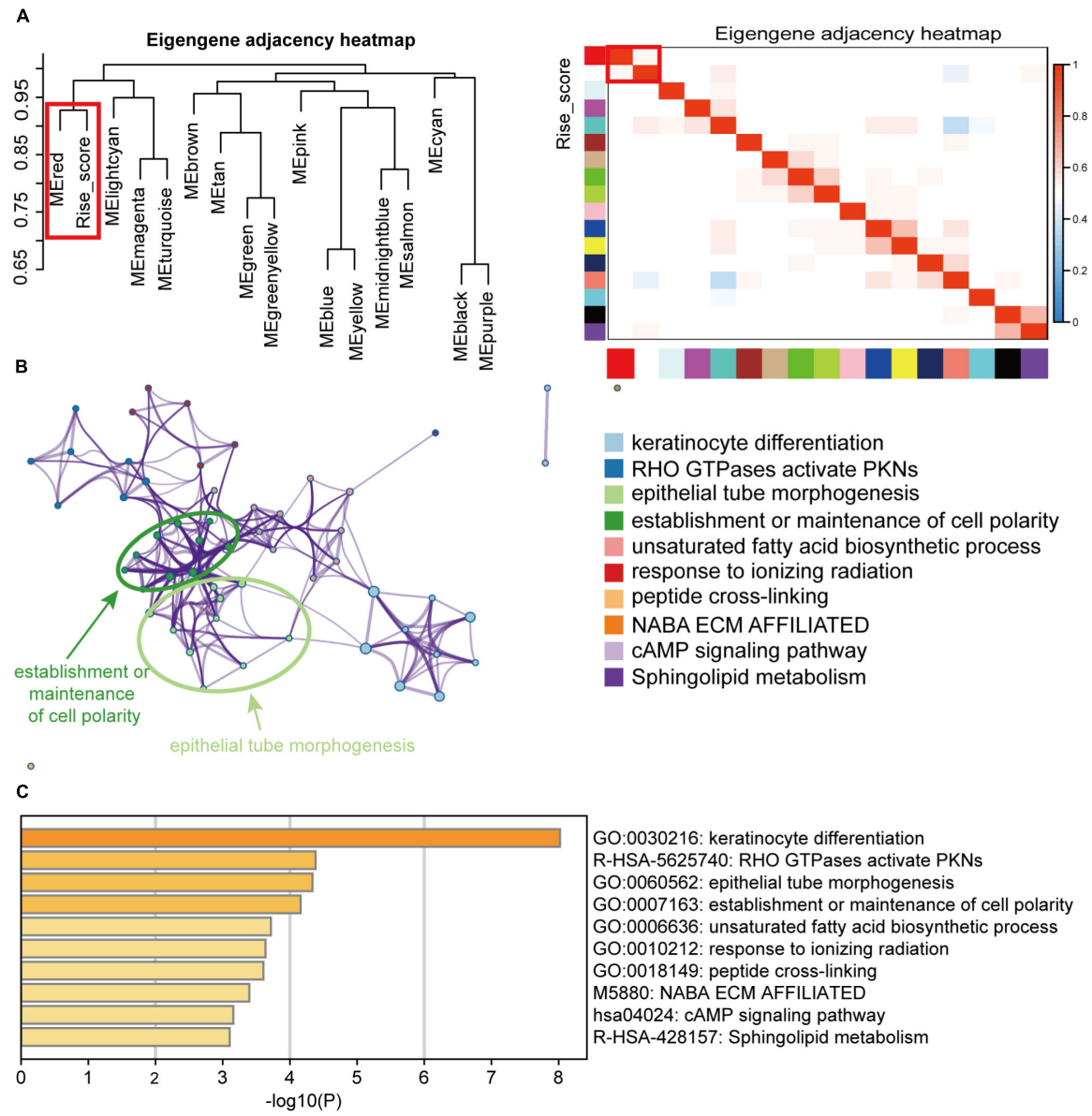


FIGURE 9 | Pathway enrichment analysis of the RBPs signature. **(A)** Weighted co-expression network analysis (WGCNA) plot was performed to cluster genes associated with the 6-RBP-based risk score, and the data of the risk score were added to construct the eigengene adjacency heatmap. **(B)** Pathways associated with the 6-RBP-based signature were enriched by Metascape. **(C)** The histogram of the top 10 enriched pathways associated with the 6-RBP-based risk score. The abscissa was the value of -Log10P, and the longitudinal axis denotes different enrichment pathways, sorted by the value of -Log10P.

score were mainly enriched in keratinocyte differentiation, and RHO GTPases activate PNKs, epithelial tube morphogenesis, establishment or maintenance of cell polarity, etc. Interestingly, we compared pathways predicted with annotations of these six RBPs in GeneCards[6] and found that *HSPG2* acted as an anti-angiogenic and anti-tumor peptide that inhibited endothelial cell migration and collagen-induced endothelial tube morphogenesis. Angiogenesis is thought to be a critical procedure of promoting BLCA progression and associated with poor survival (Bochner et al., 1995; Roudnicky et al., 2017). Among these prognostic RBPs, *HSPG2* demonstrated an anti-angiogenesis effect via binding to α2β1 integrin and interacting with *VEGFR2* at the surface of endothelial cells (Woodall et al., 2008; Goyal et al., 2012; Poluzzi et al., 2016). A previous study has demonstrated that the breakdown of cell polarity programs could promote the occurrence of aggressive, invasive tumors (VanderVorst et al., 2018). In BLCA, *BMP4* could induce monocyte/macrophage polarization toward M2 phenotype macrophages, which promoted the progression of BLCA (Martinez et al., 2017). *FLNA* also regulated cell polarity by interacting with FilGAP, a Rac-specific GTPase-activating protein (Nakamura et al., 2009), but this mechanism had not been elucidated in BLCA. Exact mechanisms of these RBPs remain largely unknown and more research is required to investigate their roles in BLCA.

CONCLUSION

In conclusion, we identified six RBPs associated with prognosis of BLCA and constructed the 6-RBP-based classifier to help clinical decision, while optimizing the predictive ability of the current TNM staging system. This study takes the initiative report that the RBP-based classifier could predict the prognosis in human BLCA. Nevertheless, large-scale, multi-center, and prospective studies are necessary to confirm our results before the 6-RBP-based signature can be applied in the clinic.

[6] https://www.genecards.org/

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

ETHICS STATEMENT

This study is a secondary data analysis, and the raw data come from public databases. The institutional and/or national research ethics committee has approved the data collection and management process.

AUTHOR CONTRIBUTIONS

YW and YL: design, analysis, and interpretation of data, drafting of the manuscript, and critical revision of the manuscript. AH and BG: acquisition of data and statistical analysis. SH, CZ, and ZK: research direction. YG, XL, and LZ: critical revision of the manuscript for important intellectual content, administrative support, obtaining funding, and supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the National Natural Science Foundation of China (Nos. 81772703 and 81972380) and the Natural Science Foundation of Beijing (7172219).

REFERENCES

Alfred Witjes, J., Lebret, T., Comperat, E. M., Cowan, N. C., De Santis, M., Bruins, H. M., et al. (2017). Updated 2016 EAU Guidelines on muscle-invasive and metastatic bladder Cancer. *Eur. Urol.* 71, 462–475. doi: 10.1016/j.eururo.2016.06.020

Baltz, A. G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* 46, 674–690. doi: 10.1016/j.molcel.2012.05.021

Bochner, B. H., Cote, R. J., Weidner, N., Groshen, S., Chen, S. C., Skinner, D. G., et al. (1995). Angiogenesis in bladder cancer: relationship between microvessel density and tumor prognosis. *J. Natl. Cancer Inst.* 87, 1603–1612. doi: 10.1093/jnci/87.21.1603

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Busa, R., Paronetto, M. P., Farini, D., Pierantozzi, E., Botti, F., Angelini, D. F., et al. (2007). The RNA-binding protein Sam68 contributes to proliferation and survival of human prostate cancer cells. *Oncogene* 26, 4372–4382. doi: 10.1038/sj.onc.1210224

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149, 1393–1406. doi: 10.1016/j.cell.2012.04.031

Chang, C. C., Shieh, G. S., Wu, P., Lin, C. C., Shiau, A. L., and Wu, C. L. (2008). Oct-3/4 expression reflects tumor progression and regulates motility of bladder cancer cells. *Cancer Res.* 68, 6281–6291. doi: 10.1158/0008-5472.CAN-08-0094

Chen, Y., Yang, F., Fang, E., Xiao, W., Mei, H., Li, H., et al. (2019). Circular RNA circAGO2 drives cancer progression through facilitating HuR-repressed functions of AGO2-miRNA complexes. *Cell Death. Differ.* 26, 1346–1364. doi: 10.1038/s41418-018-0220-6

Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., et al. (2015). Ensembl 2015. *Nucleic Acids Res.* 43, D662–D669. doi: 10.1093/nar/gku1010

Dobruch, J., Daneshmand, S., Fisch, M., Lotan, Y., Noon, A. P., Resnick, M. J., et al. (2016). Gender and bladder Cancer: a collaborative review of etiology. biology, and outcomes. *Eur. Urol.* 69, 300–310. doi: 10.1016/j.eururo.2015.08.037

Goyal, A., Poluzzi, C., Willis, C. D., Smythies, J., Shellard, A., Neill, T., et al. (2012). Endorepellin affects angiogenesis by antagonizing diverse vascular endothelial growth factor receptor 2 (VEGFR2)-evoked signaling pathways: transcriptional repression of hypoxia-inducible factor 1alpha and VEGFA and concurrent inhibition of nuclear factor of activated T cell 1 (NFAT1) activation. *J. Biol. Chem.* 287, 43543–43556. doi: 10.1074/jbc.M112.401786

Hofbauer, S. L., de Martino, M., Lucca, I., Haitel, A., Susani, M., Shariat, S. F., et al. (2018). A urinary microRNA (miR) signature for diagnosis of bladder cancer. *Urol Oncol.*36, 531.e1–531.e8. doi: 10.1016/j.urolonc.2018.09.006

Jiang, D., Zhang, Y., Yang, L., Lu, W., Mai, L., Guo, H., et al. (2020). Long noncoding RNA HCG22 suppresses proliferation and metastasis of bladder cancer cells by regulation of PTBP1. *J. Cell. Physiol.* 235, 1711–1722. doi: 10.1002/jcp.29090

King, C. E., Cuatrecasas, M., Castells, A., Sepulveda, A. R., Lee, J. S., and Rustgi, A. K. (2011). LIN28B promotes colon cancer progression and metastasis. *Cancer Res.* 71, 4260–4268. doi: 10.1158/0008-5472.CAN-10-4637

Kwon, S. C., Yi, H., Eichelbaum, K., Fohr, S., Fischer, B., You, K. T., et al. (2013). The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol. Biol.* 20, 1122–1130. doi: 10.1038/nsmb.2638

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559

Lee, H., Kim, K., Woo, J., Park, J., Kim, H., Lee, K. E., et al. (2018). Quantitative proteomic analysis identifies AHNAK (Neuroblast Differentiation-associated Protein AHNAK) as a novel candidate biomarker for bladder urothelial carcinoma diagnosis by liquid-based cytology. *Mol. Cell. Proteomics* 17, 1788–1802. doi: 10.1074/mcp.RA118.000562

Liang, G., Meng, W., Huang, X., Zhu, W., Yin, C., Wang, C., et al. (2020). 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.* 117, 4347–4357. doi: 10.1073/pnas.1917531117

Martinez, V. G., Rubio, C., Martinez-Fernandez, M., Segovia, C., Lopez-Calderon, F., Garin, M. I., et al. (2017). BMP4 Induces M2 macrophage polarization and favors tumor progression in Bladder Cancer. *Clin. Cancer Res* 23, 7388–7399. doi: 10.1158/1078-0432.CCR-17-1004

Nakamura, F., Heikkinen, O., Pentikainen, O. T., Osborn, T. M., Kasza, K. E., Weitz, D. A., et al. (2009). Molecular basis of filamin A-FilGAP interaction and its impairment in congenital disorders associated with filamin A mutations. *PLoS One* 4:e4928. doi: 10.1371/journal.pone.0004928

Pereira, B., Billaud, M., and Almeida, R. (2017). RNA-binding proteins in cancer: old players and new actors. *Trends Cancer* 3, 506–528. doi: 10.1016/j.trecan.2017.05.003

Poluzzi, C., Iozzo, R. V., and Schaefer, L. (2016). Endostatin and endorepellin: a common route of action for similar angiostatic cancer avengers. *Adv. Drug Deliv. Rev.* 97, 156–173. doi: 10.1016/j.addr.2015.10.012

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Roudnicky, F., Dieterich, L. C., Poyet, C., Buser, L., Wild, P., Tang, D., et al. (2017). High expression of insulin receptor on tumour-associated blood vessels in invasive bladder cancer predicts poor overall and progression-free survival. *J. Pathol.* 242, 193–205. doi: 10.1002/path.4892

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. *CA J. Cancer Clin.* 70, 7–30. doi: 10.3322/caac.21590

Ueda, J., Matsuda, Y., Yamahatsu, K., Uchida, E., Naito, Z., Korc, M., et al. (2014). Epithelial splicing regulatory protein 1 is a favorable prognostic factor in pancreatic cancer that attenuates pancreatic metastases. *Oncogene* 33, 4485–4495. doi: 10.1038/onc.2013.392

VanderVorst, K., Hatakeyama, J., Berg, A., Lee, H., and Carraway, K. L. III (2018). Cellular and molecular mechanisms underlying planar cell polarity pathway contributions to cancer malignancy. *Semin. Cell Dev. Biol.* 81, 78–87. doi: 10.1016/j.semcdb.2017.09.026

Vo, D. T., Subramaniam, D., Remke, M., Burton, T. L., Uren, P. J., Gelfond, J. A., et al. (2012). 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.* 181, 1762–1772. doi: 10.1016/j.ajpath.2012.07.031

Wang, Y., Du, L., Yang, X., Li, J., Li, P., Zhao, Y., et al. (2020). A nomogram combining long non-coding RNA expression profiles and clinical factors predicts survival in patients with bladder cancer. *Aging* 12, 2857–2879. doi: 10.18632/aging.102782

Wang, Z., Li, C., Jiang, M., Chen, J., Yang, M., and Pu, J. (2018). Filamin A (FLNA) regulates autophagy of bladder carcinoma cell and affects its proliferation, invasion and metastasis. *Int. Urol. Nephrol.* 50, 263–273. doi: 10.1007/s11255-017-1772-y

Woodall, B. P., Nystrom, A., Iozzo, R. A., Eble, J. A., Niland, S., Krieg, T., et al. (2008). Integrin alpha2beta1 is the required receptor for endorepellin angiostatic activity. *J. Biol. Chem.* 283, 2335–2343. doi: 10.1074/jbc.M708364200

Yin, X. H., Jin, Y. H., Cao, Y., Wong, Y., Weng, H., Sun, C., et al. (2019). Development of a 21-miRNA signature associated with the prognosis of patients with bladder Cancer. *Front. Oncol.* 9:729. doi: 10.3389/fonc.2019.00729

Yu, D., Zhang, C., and Gui, J. (2017). RNA-binding protein HuR promotes bladder cancer progression by competitively binding to the long noncoding HOTAIR with miR-1. *Onco Targets Ther.* 10, 2609–2619. doi: 10.2147/OTT.S132728

Zhang, Y., Hong, Y. K., Zhuang, D. W., He, X. J., and Lin, M. E. (2019). Bladder cancer survival nomogram: development and validation of a prediction tool, using the SEER and TCGA databases. *Medicine* 98:e17725. doi: 10.1097/MD.0000000000017725

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 02 October 2020
doi: 10.3389/fgene.2020.568192

Development and Validation of Nine-RNA Binding Protein Signature Predicting Overall Survival for Kidney Renal Clear Cell Carcinoma

Weimin Zhong[1], Chaoqun Huang[1], Jianqiong Lin[1], Maoshu Zhu[1], Hongbin Zhong[1], Ming-Hsien Chiang[2], Huei-Shien Chiang[2], Mei-Sau Hui[3], Yao Lin[4] and Jiyi Huang[1,5]**

[1] The Fifth Hospital of Xiamen, Xiamen, China, [2] Taiwan LinkMed Asia Public Health & Healthcare Management Research Association, Taipei, Taiwan, [3] Far Eastern Polyclinic, Zhongzheng, Taiwan, [4] Key Laboratory of Optoelectronic Science and Technology for Medicine of Ministry of Education, College of Life Sciences, Fujian Normal University, Fuzhou, China, [5] Xiang'an Branch, The First Affiliated Hospital of Xiamen University, Xiamen University, Xiamen, China

OPEN ACCESS

Edited by:
Juan Xu,
Harbin Medical University, China

Reviewed by:
Zhaoying Yang,
Jilin University, China
Dongjun Lee,
Pusan National University,
South Korea

***Correspondence:**
Jiyi Huang
hjy0602@163.com
Yao Lin
yaolin@fjnu.edu.cn;
yaolinfjfz@gmail.com

Specialty section:
This article was submitted to RNA,
a section of the journal Frontiers in Genetics

Received: 03 June 2020
Accepted: 31 August 2020
Published: 02 October 2020

Citation:
Zhong W, Huang C, Lin J, Zhu M, Zhong H, Chiang M-H, Chiang H-S, Hui M-S, Lin Y and Huang J (2020) Development and Validation of Nine-RNA Binding Protein Signature Predicting Overall Survival for Kidney Renal Clear Cell Carcinoma. Front. Genet. 11:568192. doi: 10.3389/fgene.2020.568192

Cumulative studies have shown that RNA binding proteins (RBPs) play an important role in numerous malignant tumors and are related to the occurrence and progression of tumors. However, the role of RBPs in kidney renal clear cell carcinoma (KIRC) is not fully understood. In this study, we first downloaded gene expression data and corresponding clinical information of KIRC from the Cancer Genome Atlas (TCGA) database, International Cancer Genome Consortium (ICGC), and Gene Expression Omnibus (GEO) database, respectively. A total of 137 differentially expressed RBPs (DERBPs) were then identified between normal and tumor tissue, including 38 downregulated and 99 upregulated RBPs. Nine RBPs (EIF4A1, RPL36A, EXOSC5, RPL28, RPL13, RPS19, RPS2, EEF1A2, and OASL) were served as prognostic genes and exploited to construct a prognostic model through survival analysis. Kaplan-Meier curves analysis showed that the low-risk group had a better survival outcome when compared with the high-risk group. The area under the curve (AUC) value of the prognostic model was 0.713 in the TCGA data set (training data set), 0.706 in the ICGC data set, and 0.687 in the GSE29609 data set, respectively, confirming a good prognostic model. The prognostic model was also identified as an independent prognostic factor for KIRC survival by performing cox regression analysis. In addition, we also built a nomogram relying on age and the prognostic model and internal validation in the TCGA data set. The clinical benefit of the prognostic model was revealed by decision curve analysis (DCA). Gene set enrichment analysis revealed several crucial pathways (ERBB signaling pathway, pathways in cancer, MTOR signaling pathway, WNT signaling pathway, and TGF BETA signaling pathway) that may explain the underlying mechanisms of KIRC. Furthermore, potential drugs for KIRC treatment were predicted by the Connectivity Map (Cmap) database based on DERBPs, including

several important drugs, such as depudecin and vorinostat, that could reverse KIRC gene expression, which may provide reference for the treatment of KIRC. In summary, we developed and validated a robust nine-RBP signature for KIRC prognosis prediction. A nomogram with risk score and age can be applied to promote the individualized prediction of overall survival in patients with KIRC. Moreover, the two drugs depudecin and vorinostat may contribute to KIRC treatment.

Keywords: kidney renal clear cell carcinoma, differentially expressed RBP, protein-protein interaction network, survival analysis, nomogram, drugs

INTRODUCTION

Renal cell carcinoma (RCC) is one of the most common cancers in people and mainly classified as three types: kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), and malignancies of the chromophobe. It has been reported that about 14,240 people died and 62,700 newly validated patients with kidney cancer were discovered in the United States in 2016 (Siegel et al., 2015). According to morphology, RCC can be mainly divided into three subtypes: KIRC, KIRP, and malignancies of the chromophobe (Fernandez-Pello et al., 2017; Foshat and Eyzaguirre, 2017). Among them, KIRC accounts for about 70%–80% kidney carcinoma. Moreover, KIRC patients have no obvious symptoms in the early stage, and about 30% of KIRC cases show metastasis when it is detected because of the sophisticated KIRC tumorigenesis in advanced stages (Ezz El Din, 2016; Zhao et al., 2016). Although some well-known biomarkers of KIRC, such as VHL/HIF, PI3K/Akt/mTOR, and Ras/Raf/MEK/ERK, have been identified, the underlying molecular mechanism of KIRC still remains uncertain (Elfiky et al., 2011; Colbert et al., 2015). Regarding the KIRC treatment, PD-1/PD-L1 blocking agents have been approved in the treatment of KIRC and in inhibiting the immune checkpoint have achieved some progress (Hahn et al., 2020). However, some patients still respond poorly, showing resistance to progress (Stein et al., 2020). Thus, it is necessary to reveal the underlying mechanism of KIRC to develop effective drugs or methods for its diagnosis and treatment.

RNA binding proteins (RBPs) are a class of proteins that interact with multiple types of RNAs. At present, it is reported that nearly 1500 RBPs were identified in the human genome (Gerstberger et al., 2014). The RBPs play a crucial role in preserving the physiological balance of cells, especially in the process of cell development and stress response (Masuda and Kuwano, 2019). Given the importance of post-transcriptional regulation, abnormal RBPs could lead to numerous human diseases. A previous study reveals that aberrant RBPs are associated with the occurrence and development of disease or cancers. For example, SRF1 and HuR can mediate post-transcriptional events to control the occurrence and progression of cardiovascular diseases (de Bruin et al., 2017). HuR can control mRNA stability to boost proliferation and metastasis of gastric cancer (Xie et al., 2019).

Currently, the potential role of RBP in KIRC is not fully understood, and a comprehensive functional study of RBP will help us fully understand its importance in the occurrence and development of KIRC. Thus, we firstly downloaded RNA sequencing data and the corresponding clinical information of KIRC from the TCGA, GEO, and ICGA databases. We then identified disregulated RBPs between normal and tumor tissue and systematically explored their prognostic values and molecular mechanisms in KIRC. Our study validated several prognostic RBPs that elevate our knowledge of the molecular mechanisms underlying KIRC.

MATERIALS AND METHODS

Data Processing

We downloaded the read count of KIRC, including 72 normal and 539 tumor tissues with its corresponding clinical information from TCGA[1] (**Table 1**). In order to identify DERBPs, we employed the edgeR R package to perform analyses (Robinson et al., 2010). The DERBPs were screened with the cutoff: $|\log \text{fold change (FC)}| \geq 1$ and false discovery rate (FDR) < 0.05. Moreover, we also downloaded 91 KIRC samples as a validation data set from the ICGC[2].

KEGG Pathway, GO Enrichment Analysis, GSEA Enrichment, and PPI Network Construction

The potential function of the DERBPs was further applied to GO enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis using clusterProfiler R package (Yu et al., 2012). Both p and FDR values less than 0.05 were statistically significant. To further screen the key module for RBPs, the DERBPs were uploaded to the STRING database[3] first (Szklarczyk et al., 2019). The Cytoscape software was further employed to build a ppi network (Smoot et al., 2011). The crucial modules were screened by using the Molecular Complex Detection (MCODE) module with the criterion score ≥ 2. GSEA enrichment analysis was performed among different risk groups, and a significant pathway was selected with the NOM-p value < 0.05 and FDR < 0.05.

[1]https://portal.gdc.cancer.gov/

[2]https://icgc.org/

[3]http://www.string-db.org/

TABLE 1 | Statistics of clinical information in high risk group and low risk group.

Characteristic		High risk (*N* = 256)	Low risk (*N* = 256)	Total (*N* = 512)	*P* value
Age	<65	153	173	326	0.08082557
	> = 65	103	83	186	
Stage	Stage I	96	160	256	3.232756e-09
	Stage II	26	30	56	
	Stage III	73	45	118	
	Stage IV	61	21	82	
T	T1	99	161	260	4.090107e-08
	T2	35	33	68	
	T3	112	61	173	
	T4	10	1	11	
M	M0	184	222	406	4.780686e-06
	M1	58	20	78	
	MX	14	14	28	
N	N0	119	109	228	0.3058459
	N1	11	5	16	
	NX	126	142	268	
Gender	Female	83	93	176	0.4023477
	Male	173	163	336	
Grade	G1	2	9	11	2.814139e-06
	G2	91	128	219	
	G3	109	94	203	
	G4	54	19	73	
	GX	0	6	6	
Smoking	1-year	135	130	265	0.8603639
	2-year	11	14	25	
	3-year	91	92	183	
	4-year	14	12	26	
	5-year	5	8	13	

Survival Analysis

By integrating clinical information and RBP expression profiles, we first performed univariate cox regression analysis using the survival R package and selected those significant RBPs with its p value smaller than or equal to 0.05. Then, in order to increase the feasibility and reliability of the clinical prognosis based on RBPs, we conducted a robust likelihood-based survival analysis to further selected target RBPs by using the Rbsurv R package (Renaud et al., 2015). The procedure was as follows:

1. All the samples were randomly categorized into the training set with N*(1-p) samples and a testing data set with N*p samples. We fitted a gene into the training data set and obtained its parameter estimation. Subsequently, we estimated the log likelihood with the parameter estimate and the validation set of samples. This evaluation was repeated 10 times for each gene, and we obtained 10 log likelihoods for each gene.
2. The best gene, g (Siegel et al., 2015), with the corresponding largest mean log likelihood was selected. We then searched the next best gene by evaluating every two-gene model and selected an optimal one with the largest mean log likelihood. A series of predictive models was built based on the above procedure, and the Akaike information criterion (AIC) value for each gene was calculated. The optimal model was screened with the lowest AIC value. Using this model, the prognostic RBPs were strictly selected.

After selecting most predictive genes, Multivariate cox regression analysis was conducted on these RBPs to calculate the corresponding coefficient. According to the coefficient, we constructed the risk score system and the formula as follows: Risk score = ΣCoef $_{RBPs}$ x Exp $_{RBPs.}$ In the risk score formula, the Coef $_{RBPs}$ represent the regression coefficients of each RBP, and Exp $_{RBPs}$ is the expression level of each prognostic RBP. Subsequently, we calculated the risk score for each patient and further categorized the patient into a high- or low-risk score group based on the median score. In addition, we performed an ROC curve analysis by using the survivalROC R package to estimate the sensitivity and specificity of the prognostic RBP risk model[4]. Log-rank $p < 0.05$ was considered significant for survival analysis.

Independence of the Risk Model of Other Clinical Parameters in TCGA

In order to evaluate the independence of the risk model, we compared clinical features, such as age, gender, grade, and stage with the risk model using the univariate and multivariate cox regression analyses, and $p < 0.05$ were considered statistically significant.

Building and Validating a Predictive Nomogram

A nomogram was built by including all significantly independent prognostic factors (Iasonos et al., 2008). The calibration plot was applied to explore the calibration and the discrimination of the nomogram. The age, prognostic, and combined models (age and prognostic model) were compared with ROC and decision curve analyses (DCA) (Vickers and Elkin, 2006).

External Validation of the Prognostic Gene Signature

The validation data sets were downloaded from ICGC with 91 samples and GSE29609 with 39 patient samples. We then calculated the risk score for each patient based on the prognostic model. Then the ROC and Kaplan-Meier analyses, respectively, were performed in the ICGA data set. In addition, the protein expression of the prognostic genes in the risk model was further validated in the Human Protein Atlas (HPA, https://www.proteinatlas.org/) (Nwosu et al., 2017). The online tool cBioportal was used to explore the genetic alterations of the prognostic genes[5].

[4]https://cran.rproject.org/web/packages/survivalROC/index.html
[5]https://www.cbioportal.org/

Identification of Candidate Small Molecules

The CMap database[6] was applied to predict a potential drug that may reverse or induce the biological states of KIRC based on the gene expression (Lamb et al., 2006). We first uploaded the DERBPs to the CMAP in the "query" module and then searched for small molecular drugs that may treat KIRC. The enrich scores ranging from -1 to 1 represent the correlation level between drugs and DERBPs. Drugs with a greater negative correlation value are more beneficial for the treatment of KIRC. Therefore, drugs with a score of $\leq$-0.75 were considered as candidate drugs for KIRC treatment. In addition, we also performed mode-of-action (MoA) analysis for the drugs to search for the potential mechanism.

RESULTS

Identification of DERBPs in KIRC

In the study, we collected 72 normal tissues and 539 tumors of KIRC from TCGA database. To explore the DERBPs, we compared the RBP gene expression between normal and tumor tissue using the edegR R package. As a result, a total of 137 DERBPs were obtained with the cutoff: $|\log FC| > 1$ and FDR < 0.05, of which 38 RBPs were downregulated and 99 were upregulated. The expression distribution of these differently expressed RBPs is shown in **Figures 1A,B**.

GO and KEGG Enrichment for DERBPs

In order to explore the potential function of the DERBPs, we use the clusterProfiler R package to perform functional enrichment analysis. As a result, these RBPs were mainly enriched in translational initiation, mRNA catabolic processes, RNA catabolic processes, nuclear-transcribed mRNA catabolic processes, SRP-dependent co-translational protein targeting to membrane, and co-translational protein targeting to membrane (**Supplementary Figure 1A**). Moreover, we also discovered that these RBPs were involved in ribosome and legionellosis pathways in the KEGG result, which is consistent with the previous study (**Supplementary Figure 1B**).

[6]http://www.broadinstitute.org

Construction Protein–Protein Interaction (PPI) Network and Crucial Modules Screening

To explore the role of DERBPs, we uploaded the RBPs to the String database and identified a PPI network. We further used the Cytoscape software to visualize it (**Figure 2A**). For the purpose of searching the key modules from the PPI network, we used the MCODE module to identify the important modules. As a result, the top two important modules were acquired, which consist of 26 potential key RBPs (**Figure 2B**).

Identification and Selection of Prognostic Related RBPs

In order to obtain a reliable survival result for KIRC, we first excluded samples with a survival time less than 30 days. As a result, a set of 26 RBPs with 512 patients were exploited into univariate cox regression analysis, and a total of 10 significant RBPs were identified ($p < 0.05$) (**Supplementary Table 1**). To ensure the stability and feasibility of clinical prognosis based on 10 RBPs, we further made a selection on the 10 RBPs using the robust likelihood-based survival analysis. As shown in **Table 2**, nine genes, including EIF4A1, RPL36A, EXOSC5, RPL28, RPL13, RPS19, RPS2, EEF1A2,

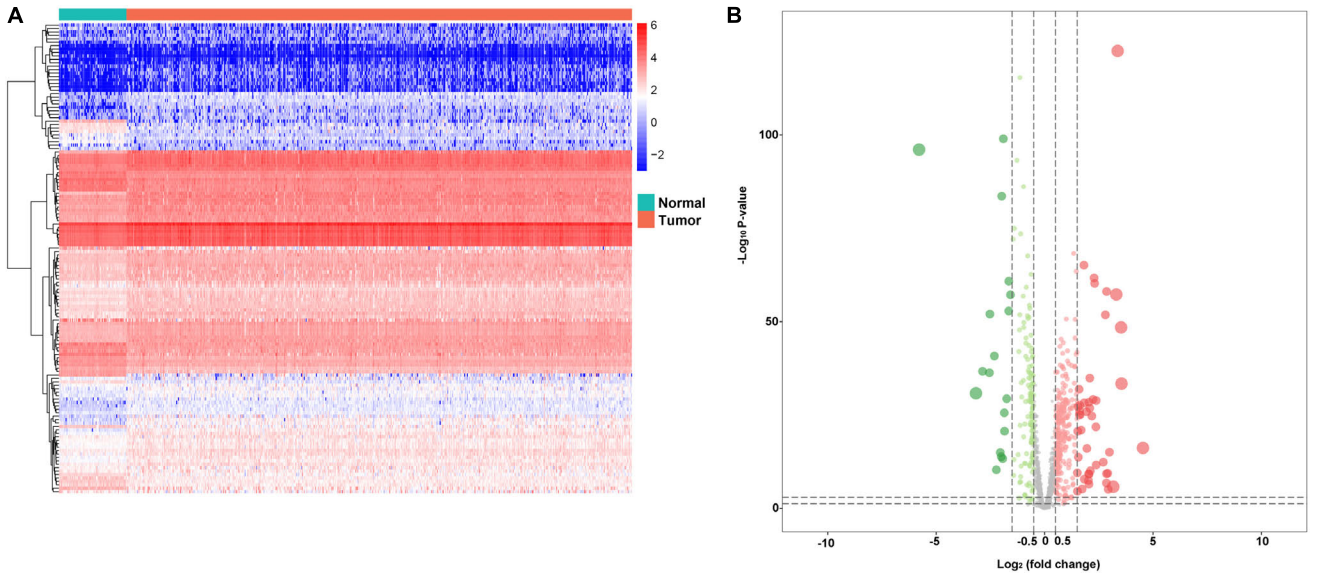


FIGURE 1 | Differentially expressed RBPs. **A** Heat maps of differentially expressed RBPs between tumor and normal tissues in the TCGA data set. **B** Volcano plot of differentially expressed RBPs; red dots represent upregulated RBPs, and green dots represent downregulated RBPs.

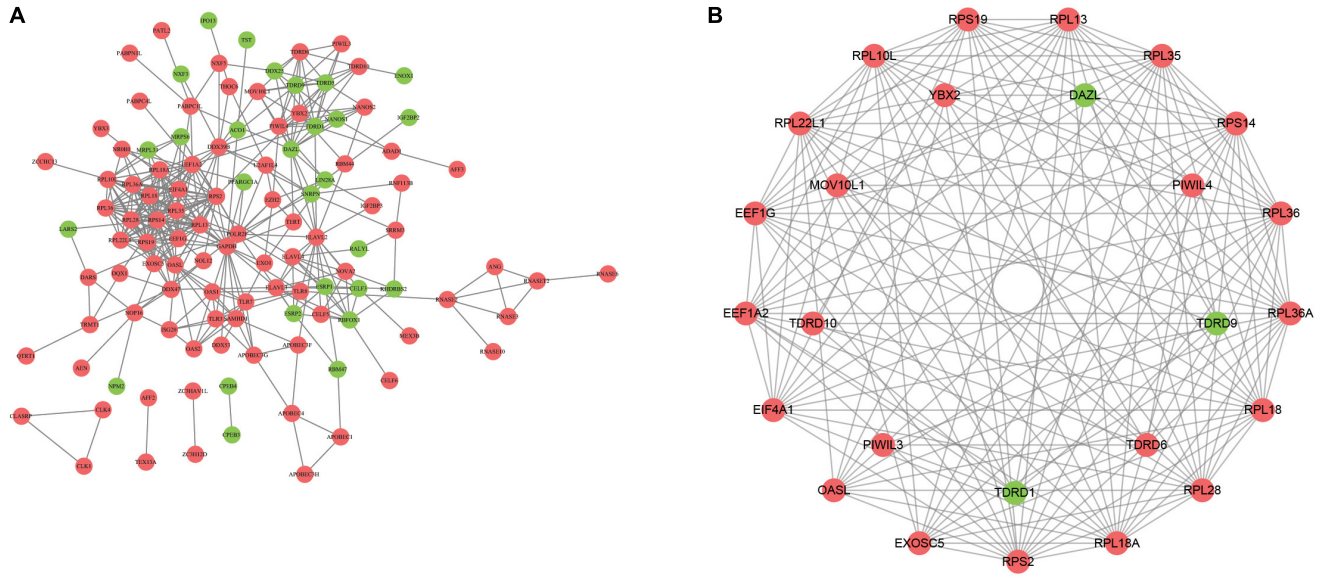


FIGURE 2 | Construction of protein–protein interaction network. **A** The network visualization using cytoscape for all differentially expressed RBPs. **B** The network of two key modules visualized by cytoscape.

and OASL, were picked out. To systemically investigate the relationship between these nine RBPs and prognosis of KIRC, we developed a nine-RBP signature-based risk score based on their cox coefficient:

Risk score = (0.005121079 * EIF4A1)
+ (−0.065342266 * RPL36A)
+ (−0.074842527 * EXOSC5)
+ (−0.007007688 * RPL28)
+ (0.003365894 * RPL13)
+ (−0.000184204 * RPS19)
+ (0.000510318 * RPS2)
+ (0.017008893 *EEF1A2)
+ (0.118627759 * OASL)

We then calculated the risk score for each patient based on the risk formula and ranked them according to the risk score. **Figure 3A** shows that survival time of patients with KIRC was affected adversely with their risk score. Numerous cases of death were related to a high-risk score, and patients with a low-risk score have prolonged survival time. The Kaplan-Meier curve and log-rank test indicated that patients in the low-risk group have a better survival time than in the high-risk group ($p < 0.01$) (**Figure 3B**). To compare the sensitivity and specificity of survival prediction, ROC analysis was performed for the nine-RBP signature. As shown in **Figure 3C**, the area under the curve (AUC) values reached 0.713, exhibiting a good accuracy. In addition, to further explore the function between the high- and low-risk group, we performed GSEA enrichment and found several important pathways, including the insulin signaling pathway, ERBB signaling pathway, renal cell carcinoma, pathways in cancer, MTOR signaling pathway, WNT signaling pathway, TGF BETA signaling pathway, and UBIQUITIN mediated proteolysis that were enriched in the low-risk group (**Figure 4**). We then assessed the alterations in nine genes by using the cBioPortal database as shown in **Figure 5**, and the RPL36A gene included six amplification samples; RPL28 and EEF1A2 were altered in 0.6% of cases, and EXOSC5, RPS19, and RPS2 were altered in 0.3% cases while EIF4A1 and OASL have no mutation cases.

TABLE 2 | Prognostic RBPs signature screened by performing forward selection analysis in the TCGA dataset.

Gene ID	nloglik	AIC
EIF4A1	943.86	1889.73*
RPL36A	935.21	1874.43*
EXOSC5	934.78	1875.56*
RPL28	933.65	1875.3*
RPL13	933.65	1877.3*
RPS19	933.58	1879.15*
RPS2	932.74	1879.49*
EEF1A2	929.78	1875.55*
OASL	926.6	1871.2*

Independent Prognostic Role of the Prognostic RBP Signature

To explore the independence of the signature, we compared the clinical features including gender, age, smoking, grade, stage, T, N, M, and RBP signature by performing univariate and multivariate cox regression analysis. As shown in **Figures 6A,B**, the age and RBP signature risk score were considered as the independent prognostic factor ($p < 0.05$). Then patients were stratified according to age (<65 and ≥65). Patients in the high-risk group shown significantly poorer OS than those in the low-risk group both in <65 and in ≥65 (**Figures 6C,D**).

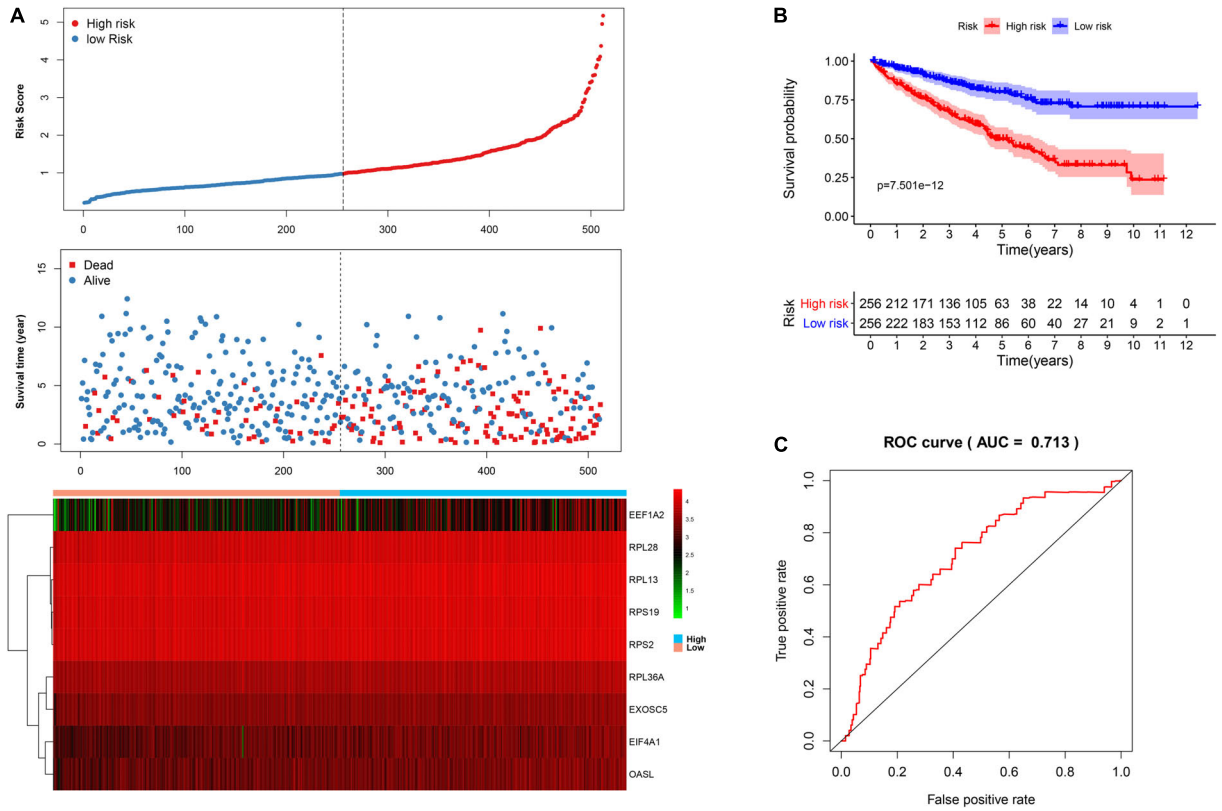


FIGURE 3 | The nine-RBP signature associated with overall survival of KIRC in the TCGA data set. **A** The upper panel represents the risk score distribution for each patient, the middle panel shows the patient distribution with increasing risk value, and the lower panel represents the expression of nine prognostic RBPs. **B** Kaplan-Meier curve analysis for the patients in KIRC between the high- and low-risk groups. **C** ROC curve analysis for the prognostic model.

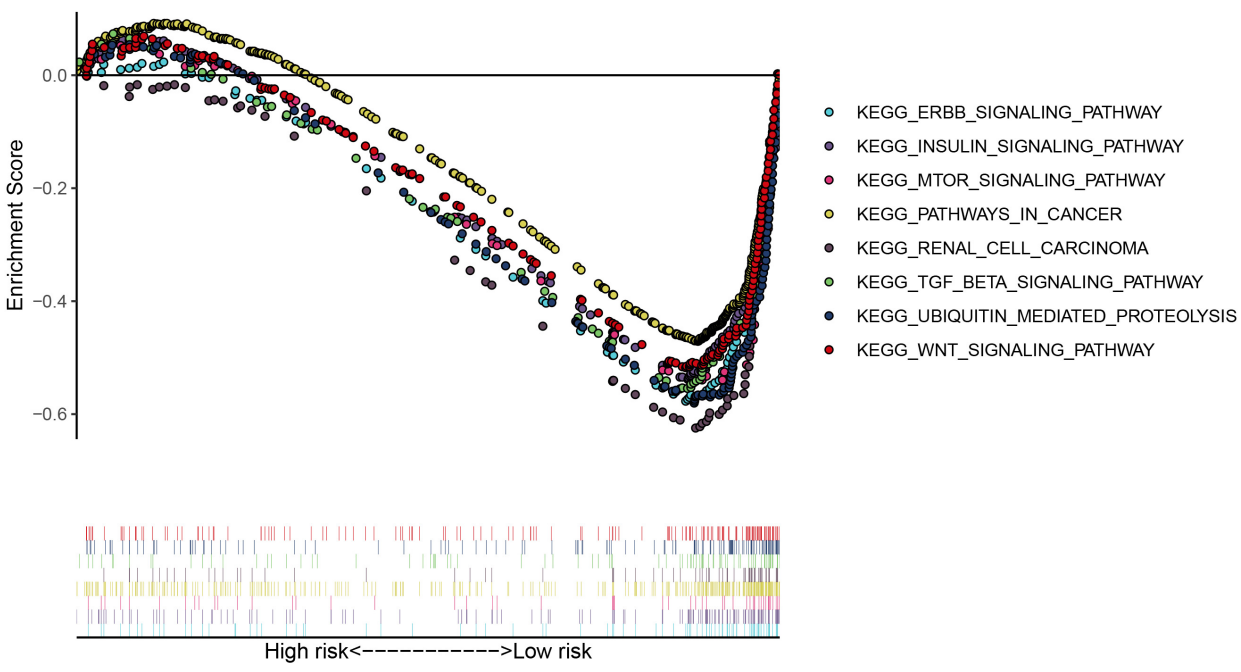


FIGURE 4 | The significant pathways were enriched in the low-risk group by performing the GSEA analysis based on the gene expression.

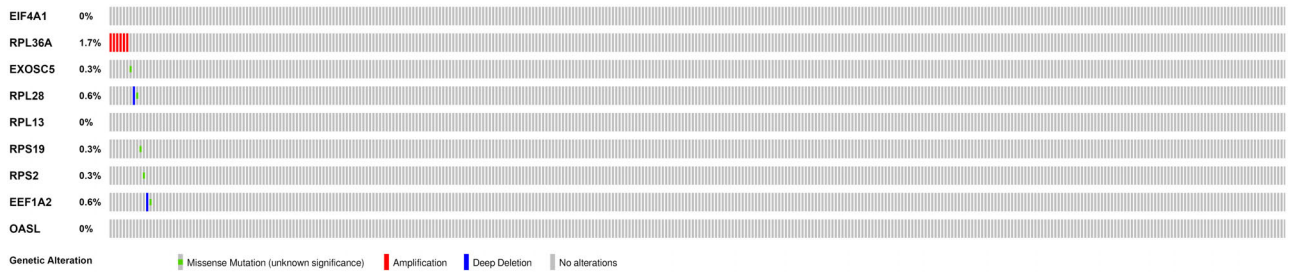


FIGURE 5 | Genetic alterations of the nine RBPs in KIRC patients; the data were retrieved from the cBioportal database.

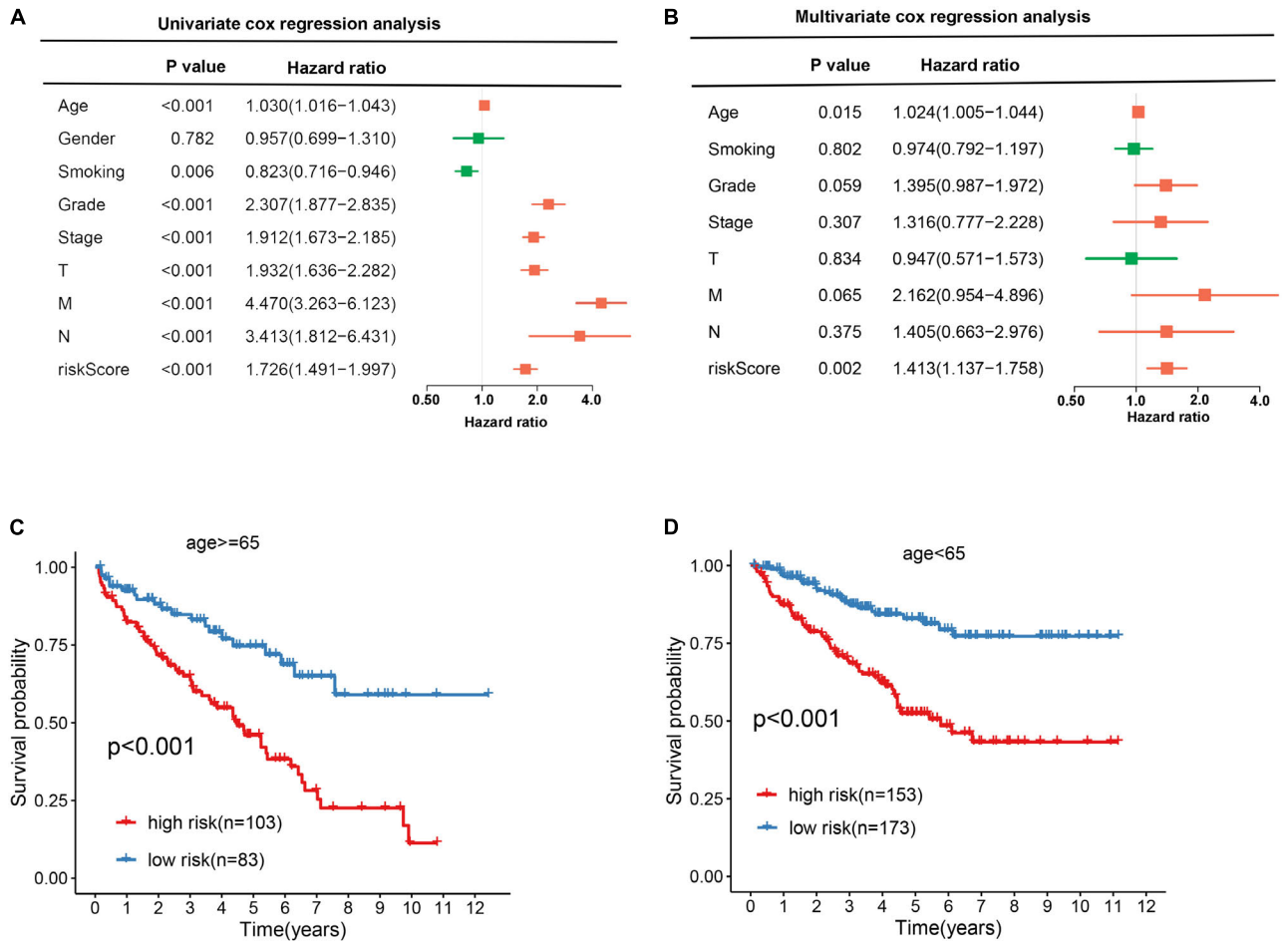


FIGURE 6 | Exploring the association between clinical traits and the prognostic model. **A** Univariate cox regression analysis between clinical traits and prognostic risk score. **B** Multivariate cox regression analysis between clinical traits and prognostic risk score. **C, D** The Kaplan–Meier curve shows the prognostic value of the nine-RBP signature for KIRC patients categorized by age.

Construction of a Nomogram Based on Prognostic Model and Clinical Features

In order to evaluate the clinical trait and prognostic model for KIRC prognosis, we integrated the prognostic model and age to build a nomogram (**Figure 7**). In addition, the corresponding calibration plots in 1, 3, and 5 year were also drawn, and it was discovered that the performance of the nomogram was best in predicting 1 year OS (**Figures 8A–C**). We further estimate the AUC value for the age and prognostic model, respectively. As shown in **Figures 9A–C**, the AUC values for 1-, 3-, and 5-year OS were 0.64, 0.57, and 0.59 in age, and in the prognostic model, the AUC value for 1-, 3-, and 5-year OS were 0.71, 0.66, and 0.69, respectively. Interestingly, when we incorporated the age and

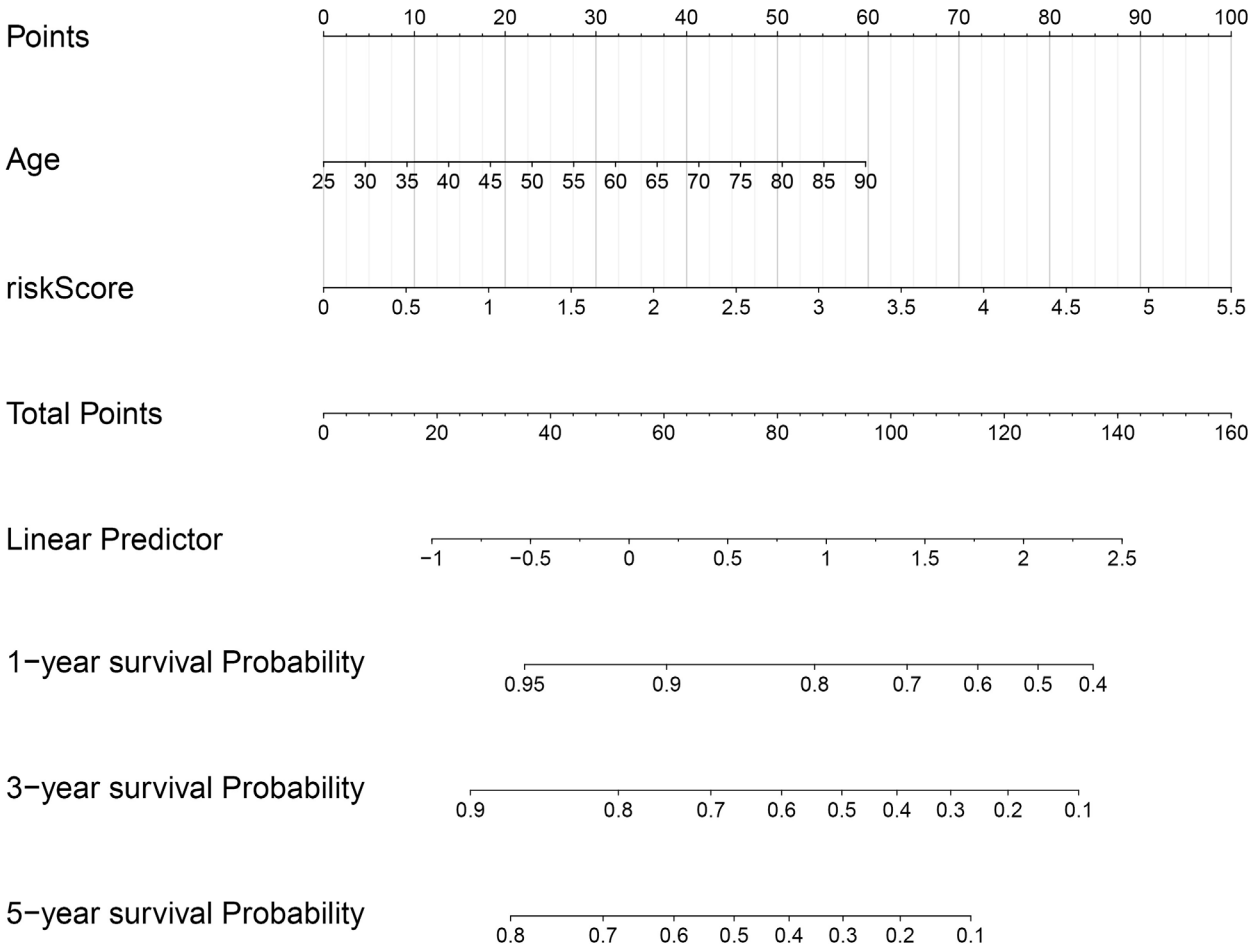


FIGURE 7 | A nomogram plot was constructed on the basis of two independent prognostic factors (age and prognostic risk score) in KIRC.

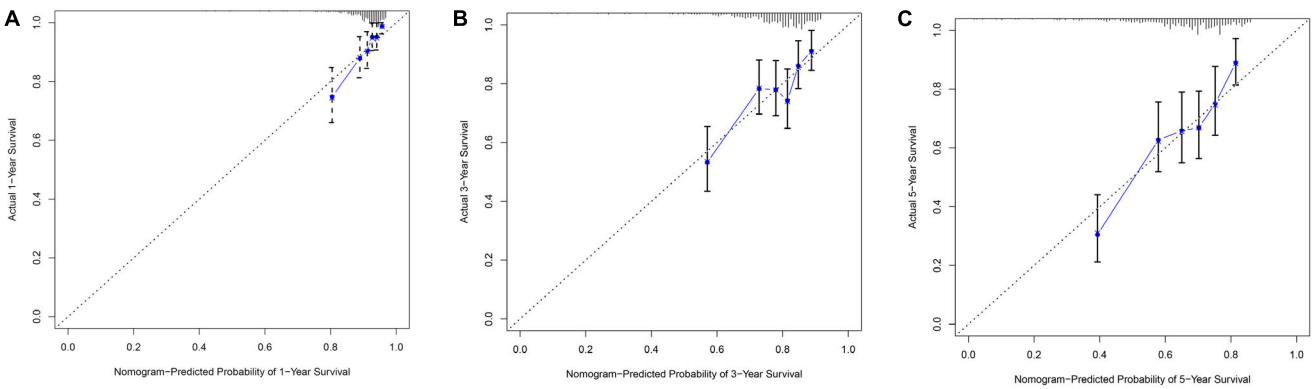


FIGURE 8 | The calibration plot for internal validation of the nomogram within 1-year **(A)**, 3-year **(B)** and 5-year **(C)**, respectively.

prognostic model into a combined model, the AUC value in 1, 3, and 5 years was increased, especially in 1-year OS (**Figures 9A–C**). Moreover, we also discovered that combining our prognostic model with age showed some net benefit for predicting OS (**Figures 9D–F**).

Validation of the Prognostic Model and Hub RBPs

In order to validate the stability and reliability of the prognostic model, we first downloaded 91 samples with complete clinical information as the validation data set from the ICGC database.

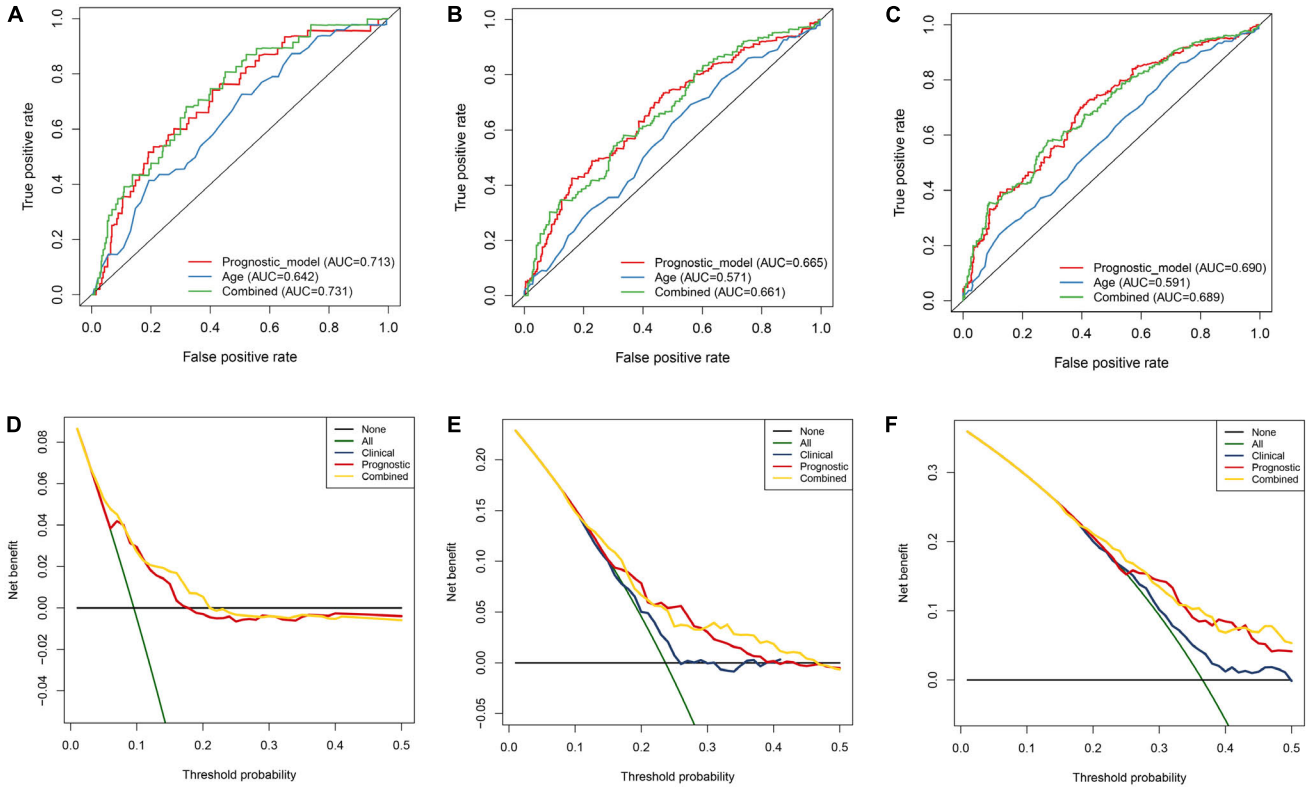


FIGURE 9 | Estimation of nomogram by performing ROC curve and DCA curve analyses within 1, 3, and 5 years, respectively. **A–C** ROC curves analysis of the nomogram compared for 1, 3, and 5 years. **D–F** DCA curve analysis of the nomogram compared for 1, 3, and 5 years.

Using the prognostic model, we calculated the risk score for each patient and divided patients into high- and low-risk group, respectively. We found that patients in the high-risk group corresponded to higher death rates (**Figure 10A**). The Kaplan-Meier curve and log-rank test suggest that patients in the high-risk group have a worse survival rate compared to the low-risk group ($p < 0.05$) (**Figure 10B**). Moreover, the AUC for overall survival was reached in 0.706, indicating good accuracy (**Figure 10C**). Similarly, we also downloaded a GSE29609 data set from the GEO database that included 39 samples. According to the risk model, we also calculated the risk score for each patient and then classified into them high- and low-risk group, respectively. The Kaplan-Meier curve and log-rank test suggest that patients have a significant difference between risk groups ($P < 0.05$) (**Supplementary Figure 2A**). The ROC analysis results indicate good accuracy of the risk model for the prognosis of KIRC (**Supplementary Figure 2B**).

To further explore the prognostic value of nine hub RBPs in KIRC, we used the Kaplan-Meier curve and log-rank test analyses to determine the association between hub RBPs and disease-free survival (DFS). As shown in **Figure 11**, the nine hub RBPs were significantly associated with the DFS in KIRC patients, and high expression of them corresponded to a lower survival probability ($p < 0.05$). We also evaluated the expression level of the nine hub RBPs between tumor and normal tissue. As shown in **Supplementary Figure 3**, most of the hub RBPs presented significant divergence between normal and tumor tissue except for EEF1A2. Interestingly, these RBPs show a high expression level in tumor tissue when compared to normal tissue.

In addition, we further explore the protein expression of nine hub RBPs. We employed immunohistochemistry results from the HPA database to discover that EIF4A1 was significantly increased in kidney tumor tissue compared with normal tissue (**Supplementary Figure 4**). However, the antibody staining level of EEF1A2 and RPL36A were decreased from normal tissue to kidney tumor tissue (**Supplementary Figure 4**). Moreover, the protein expression level of EXOSC5, RPL13, RPL28, and RPS2 were not significant between normal and tumor tissue, and EXOSC5 and RPS19 were not detected in the HPA database.

Related Drugs Screening for KIRC Treatment

To identify the potential drugs for KIRC, we uploaded the upregulated and downregulated RBPs to the CMAP database. As a result, 27 significant candidate drugs that score $\leq$ -0.50 and p value < 0.05 were considered as potential drugs for KIRC treatment (**Supplementary Table 2**). The mechanism of action for these drugs were further analyzed and are shown in **Figure 12**. We can discover that these drugs were enriched in the HDAC inhibitor, protein synthesis inhibitor,

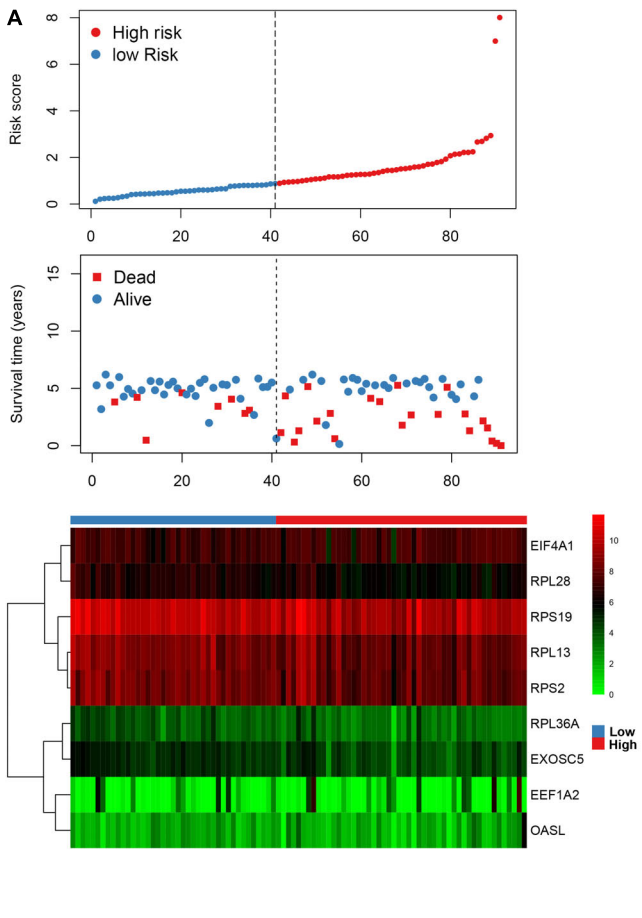


FIGURE 10 | The nine-RBP signature associated with overall survival of KIRC in the ICGC data set. **A** The upper panel represents the risk score distribution for each patient, the middle panel shows the patients' distribution with increasing risk value, and the lower panel represents the expression of nine prognostic RBPs. **B** Kaplan-Meier curves analysis for the patients in KIRC between the high- and low-risk groups. **C** ROC curve analysis for the prognostic model.

adrenergic receptor antagonist, cytokine production inhibitor, glucocorticoid receptor agonist, histamine receptor agonist, histamine receptor antagonist, lipoprotein lipase activator, local anesthetic, MAP kinase inhibitor, and Tricyclic antidepressant (**Figure 12**). These mechanisms of action and potential small molecule drugs might provide guidance for developing targeted drugs for KIRC.

DISCUSSION

Disorders of RBPs have been reported in numerous malignant tumors (Gerstberger et al., 2014; Masuda and Kuwano, 2019). However, fewer studies have comprehensively investigated the function and prognosis of RBPs. In the present study, we systemically explored the prognosis and function of hub RBPs in KIRC. A total of 137 DERBPs were identified between tumor and normal tissue of KIRC based on the TCGA RNAseq data. We comprehensively investigate the potential function and pathway and construct a PPI network for these RBPs. Furthermore, we constructed and validated a nine-RBP signature to predict KIRC prognosis based on the cox regression coefficient using the univariate cox regression analysis, robust likelihood-based survival analysis, multivariate cox regression analysis, and ROC analysis. We also identified some potential drugs that may contribute to treatment of KIRC. These findings might provide new insight into the pathogenesis of KIRC and potential therapeutic targets for KIRC.

Functional enrichment analysis results reveal that the DERBPs were mainly enriched in translation initiation, mRNA catabolic processes, RNA catabolic processes, nuclear-transcribed mRNA catabolic processes, SRP-dependent co-translational protein targeting to membrane, and cotranslational protein targeting to membrane, etc. Previous studies have demonstrated that regulation of translation, RNA processing, and the RNA metabolism process were the causes of the occurrence and development of the human disease (Jain et al., 2019; Siang et al., 2020). The KEGG analysis results indicate that the dysregulated RBPs were enriched in Ribosome and Legionellosis, which is consistent with previous studies (Li et al., 2020).

In addition, we constructed a PPI network for these DERBPs and identified two key modules with 26 hub RBPs. We further

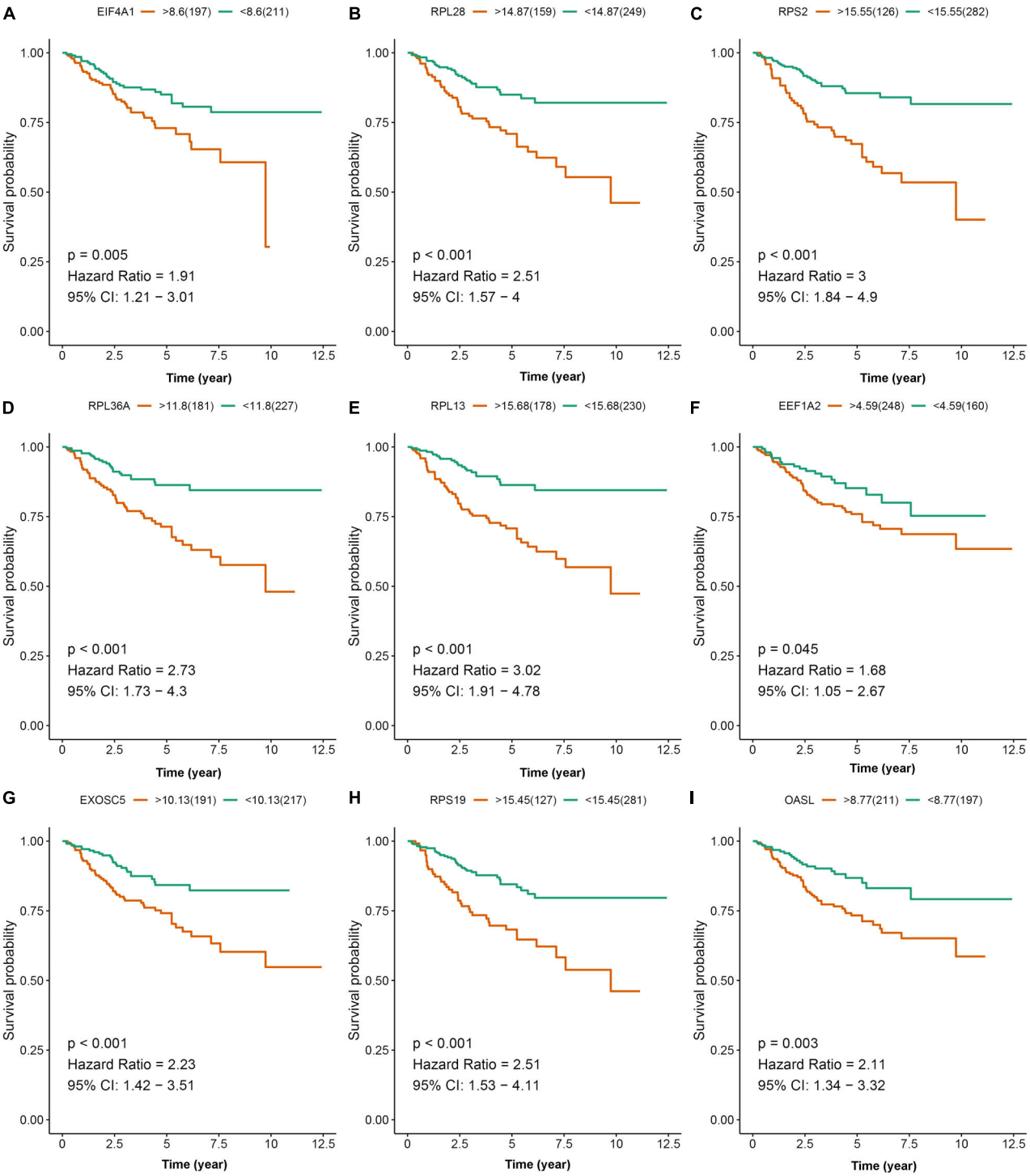


FIGURE 11 | Disease-free survival analysis of the nine prognostic RBPs including EIF4A1 **(A)**, RPL28 **(B)**, RPS2 **(C)**, RPL36A **(D)**, RPL13 **(E)**, EEF1A2 **(F)**, EXOSC5 **(G)**, RPS19 **(H)**, and OASL **(I)** in the TCGA data set.

explored the association between 26 RBPs and overall survival of KIRC by performing univariate Cox regression analysis, robust likelihood-based survival analysis, and multivariate Cox regression analysis. A total of nine RBPs, including EIF4A1, RPL36A, EXOSC5, RPL28, RPL13, RPS19, RPS2, EEF1A2, and OASL, were identified as prognostic RBPs. Among the nine RBPs, EEF1A2, and RPL13 have been reported to be associated with tumorigenesis and progression of kidney cancer patients

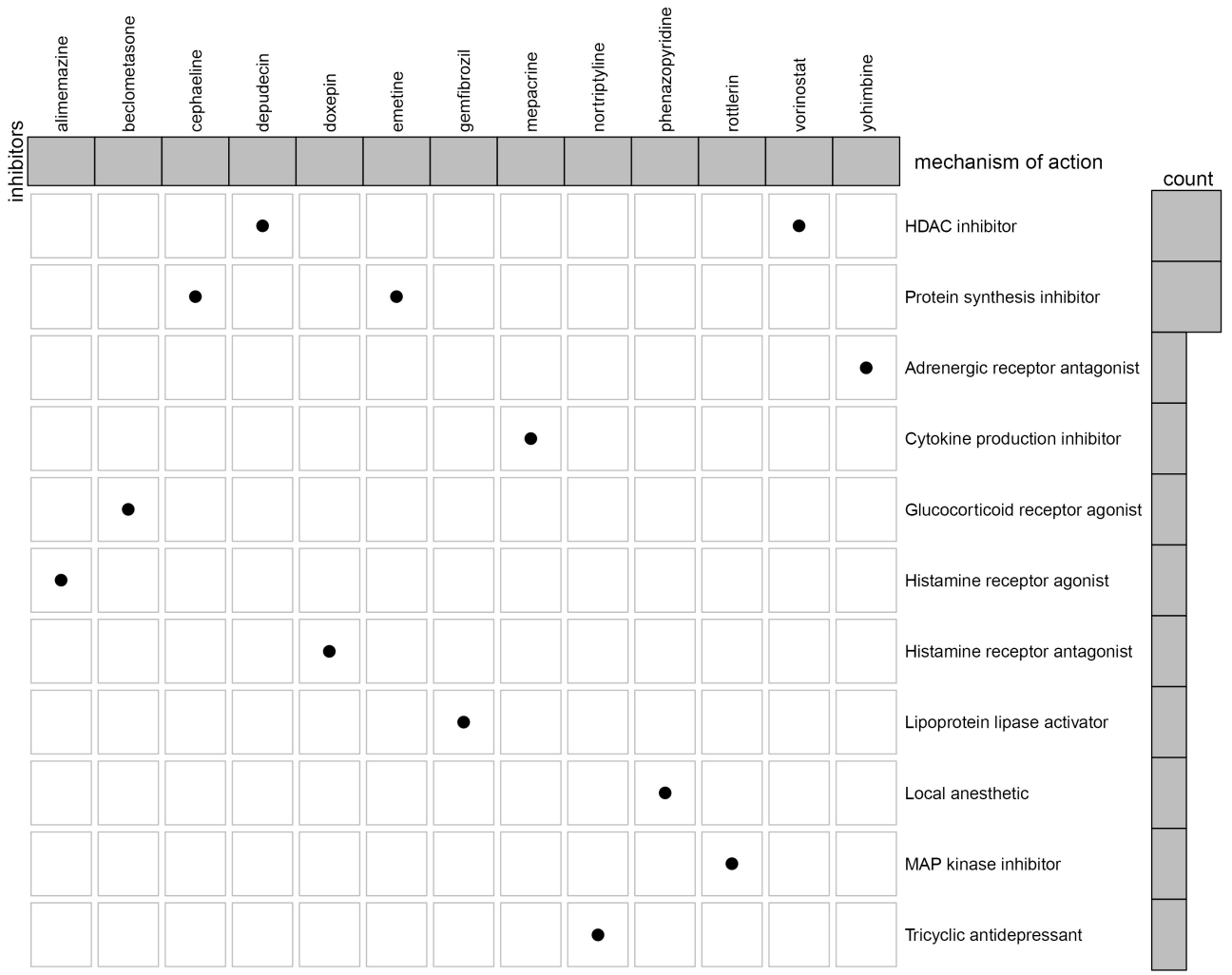


FIGURE 12 | Oncoplot of the mechanism of action for the potential drugs.

(Pflueger et al., 2013; Wierzbicki et al., 2014). Eukaryotic translation initiation factor 4A1 (EIF4A1) is a component of the translation initiation complex, and a high expression level of EIF4A1 is positively associated with poor tumor differentiation, late T stage, lymph node metastasis, advanced TNM stage, and poor prognosis in patients with gastric cancer (Gao et al., 2020). Overexpression of ribosomal protein L36a (RPL36A) has been reported to closely relate to cellular proliferation in hepatocellular carcinoma (Kim et al., 2004). The EXOSC5 was identified as a novel prognostic marker that can promote proliferation of colorectal cancer through activating the ERK and AKT pathways (Pan et al., 2019). The mutation of RPL28 was associated with shorter progression-free survival and overall survival in metastatic colorectal cancer (Labriet et al., 2019). Markiewski et al. found that the ribosomal protein S19 (RPS19) can contribute to generate regulatory T cells while reducing infiltration of CD8 + T cells into tumors. When the expression level of RPS19 is decreased, the tumor growth is impaired, and the development of tumors is also delayed in a transgenic model of breast cancer (Markiewski et al., 2017). The RPS2 and OASL were considered to be a potential therapeutic target in prostate cancer and lung cancer (Lv et al., 2018). According to the nine genes, we built a risk model with their coefficient. The ROC analysis results in the TCGA data set and ICGC data set revealed that our risk model has a good performance to predict survival of KIRC.

The GSEA result revealed many significant cancer-related pathways for the RBP signature, of which the insulin signaling pathway, ERBB signaling pathway, renal cell carcinoma, pathways in cancer, MTOR signaling pathway, WNT signaling pathway, TGF BETA signaling pathway, and UBIQUITIN mediated proteolysis were enriched in the low-risk group, and no significant pathway was enriched in the high-risk group. On one hand, these results demonstrate the robust connection of the RBP signature with tumorgenesis and progression of KIRC. On the other hand, the results might provide promising directions to elaborate the underlying molecular mechanisms of the signature.

To identify potential drugs for KIRC treatment, we obtained 27 compounds from the prediction of the CMAP database based on the DERBPs. Among these compounds, vorinostat, a histone deacetylase (HDAC) suppressor, has been reported to be a promising drug in the treatment of KIRP (Pang et al., 2019). The HDACs are a class of enzymes in the nucleus of eukaryotic organisms that promote histone deacetylation and, accordingly, allow histones to assemble and convert DNA into biologically active units (Valenzuela-Fernandez et al., 2008). According to the report, HDACs (HDAC1 and HDAC2) are necessary for the growth and survival of RCC tumor cells, and inhibition of HDACs might improve the response of oncologic chemotherapy for RCC (Aggarwal et al., 2017; Kiweler et al., 2018). Interestingly, depudecin is also an HDAC suppressor, which contributes to inducing morphological reversion of transformed fibroblasts and has been used to treat neuroendocrine tumor (Kwon et al., 1998; Kunnimalaiyaan and Chen, 2007). A recent study reports that depudecin can serve as a candidate drug for the treatment of pituitary adenomas (Zhou et al., 2016). The present study indicates a close reverse mechanistic association of depudecin and vorinostat with KIRC, suggesting that the two drugs may serve as suitable drugs for KIRC treatment. However, the mechanism and efficacy of the two drugs for treatment of KIRC remain to be elucidated in future studies.

Overall, we constructed an RBP prognostic model based on bioinformatics analysis, which have potentially substantial clinical significance. However, several limitations need to be pointed out. First, all the results were based on analysis, and further experimental verification is required. Second, the data sets did not provide complete clinical information, especially in the validation data set, which may reduce the statistical reliability and validity of the result.

CONCLUSION

In conclusion, our study presents the expression, function, and prognostic potential of RBPs in KIRC. We identified a novel nine-RBP signature for KIRC and proved that the prognostic model can serve as an independent predictor for KIRC. To our knowledge, this is the first attempt to develope an RBP prognostic model in KIRC. In addition, we also identified two prospective drugs for the treatment of KIRC.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov, https://www.ncbi.nlm.nih.gov/geo/, and https://dcc.icgc.org/.

AUTHOR CONTRIBUTIONS

YL and JH designed the study. CH, JL, MZ, and HZ collected the clinical information and expression data. WZ analysis data and wrote the manuscript. M-HC, H-SC, and M-SH revised and offered advice about the manuscripts. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Health Science Research Personnel Training Program of Fujian Province (2017-CXB-22), Provincial Natural Science Foundation of Fujian (2018D0022), and Xiamen Medical Advantage Subspecialty Vascular Access Construction Fund ([2018] 296).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.568192/full#supplementary-material

Supplementary Figure 1 | GO **(A)** and KEGG pathway **(B)** enrichment analysis for the differentially expressed RBPs.

Supplementary Figure 2 | The nine RBP signature associated with overall survival of KIRC in the GSE29609 data set; **(A)** Kaplan-Meier curve analysis for the patients in KIRC between the high- and low-risk groups. **B** ROC curve analysis for the prognostic model.

Supplementary Figure 3 | Exploration of the expression of the nine hub RBPs in normal and tumor tissue.

Supplementary Figure 4 | Validation of the seven RBPs on the protein expression level based on Immunohistochemical results in KIRC; the data was retrieved from HPA database.

Supplementary Table 1 | Univariate cox regression analysis for the hub RBPs.

Supplementary Table 2 | Enrichment analysis for the 27 potential drugs in KIRC using the CMAP database.

REFERENCES

Aggarwal, R., Thomas, S., Pawlowska, N., Bartelink, I., Grabowsky, J., Jahan, T., et al. (2017). Inhibiting histone deacetylase as a means to reverse resistance to angiogenesis inhibitors: phase I study of abexinostat plus pazopanib in advanced solid tumor malignancies. *J. Clin. Oncol.* 35, 1231–1239. doi: 10.1200/jco.2016.70.5350

Colbert, L. E., Fisher, S. B., Balci, S., Saka, B., Chen, Z., Kim, S., et al. (2015). High nuclear hypoxia-inducible factor 1 alpha expression is a predictor of distant recurrence in patients with resected pancreatic adenocarcinoma. *Int. J. Radiat. Oncol. Biol. Phys.* 91, 631–639. doi: 10.1016/j.ijrobp.2014.11.004

de Bruin, R. G., Rabelink, T. J., van Zonneveld, A. J., and van der Veer, E. P. (2017). Emerging roles for RNA-binding proteins as effectors and regulators of cardiovascular disease. *Eur. Heart J.* 38, 1380–1388.

Elfiky, A. A., Aziz, S. A., Conrad, P. J., Siddiqui, S., Hackl, W., Maira, M., et al. (2011). Characterization and targeting of phosphatidylinositol-3 kinase (PI3K) and mammalian target of rapamycin (mTOR) in renal cell cancer. *J. Transl. Med.* 9:133. doi: 10.1186/1479-5876-9-133

Ezz El Din, M. (2016). Utilization of sunitinib for renal cell cancer: an Egyptian university hospital experience. *Asian Pac. J. Cancer Prev.* 17, 3161–3166.

Fernandez-Pello, S., Hofmann, F., Tahbaz, R., Marconi, L., Lam, T. B., Albiges, L., et al. (2017). A systematic review and meta-analysis comparing the effectiveness

and adverse effects of different systemic treatments for non-clear cell renal cell carcinoma. *Eur. Urol.* 71, 426–436. doi: 10.1016/j.eururo.2016.11.020

Foshat, M., and Eyzaguirre, E. (2017). Acquired cystic disease-associated renal cell carcinoma: review of pathogenesis, morphology, ancillary tests, and clinical features. *Arch. Pathol. Lab. Med.* 141, 600–606. doi: 10.5858/arpa.2016-0123-rs

Gao, C., Guo, X., Xue, A., Ruan, Y., Wang, H., and Gao, X. (2020). High intratumoral expression of eIF4A1 promotes epithelial-to-mesenchymal transition and predicts unfavorable prognosis in gastric cancer. *Acta Biochim. Biophys. Sin.* 52, 310–319. doi: 10.1093/abbs/gmz168

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. *Nat. Rev. Genet.* 15, 829–845. doi: 10.1038/nrg3813

Hahn, A. W., Drake, C., Denmeade, S. R., Zakharia, Y., Maughan, B. L., Kennedy, E., et al. (2020). A Phase I study of Alpha-1,3-galactosyltransferase-expressing allogeneic renal cell carcinoma immunotherapy in patients with refractory metastatic renal cell carcinoma. *Oncologist* 25, 121–213. doi: 10.1634/theoncologist.2019-0599

Iasonos, A., Schrag, D., Raj, G. V., and Panageas, K. S. (2008). How to build and interpret a nomogram for cancer prognosis. *J. Clin. Oncol.* 26, 1364–1370. doi: 10.1200/jco.2007.12.9791

Jain, A., Brown, S. Z., Thomsett, H. L., Londin, E., and Brody, J. R. (2019). Evaluation of post-transcriptional gene regulation in pancreatic cancer cells: studying RNA binding proteins and their mRNA targets. *Methods Mol. Biol.* 1882, 239–252. doi: 10.1007/978-1-4939-8879-2_22

Kim, J. H., You, K. R., Kim, I. H., Cho, B. H., Kim, C. Y., and Kim, D. G. (2004). Over-expression of the ribosomal protein L36a gene is associated with cellular proliferation in hepatocellular carcinoma. *Hepatology* 39, 129–138. doi: 10.1002/hep.20017

Kiweler, N., Brill, B., Wirth, M., Breuksch, I., Laguna, T., Dietrich, C., et al. (2018). The histone deacetylases HDAC1 and HDAC2 are required for the growth and survival of renal carcinoma cells. *Arch. Toxicol.* 92, 2227–2243. doi: 10.1007/s00204-018-2229-5

Kunnimalaiyaan, M., and Chen, H. (2007). Tumor suppressor role of Notch-1 signaling in neuroendocrine tumors. *Oncologist* 12, 535–542. doi: 10.1634/theoncologist.12-5-535

Kwon, H. J., Owa, T., Hassig, C. A., Shimada, J., and Schreiber, S. L. (1998). Depudecin induces morphological reversion of transformed fibroblasts via the inhibition of histone deacetylase. *Proc. Natl. Acad. Sci. U.S.A.* 95, 3356–3361. doi: 10.1073/pnas.95.7.3356

Labriet, A., Levesque, E., Cecchin, E., De Mattia, E., Villeneuve, L., Rouleau, M., et al. (2019). Germline variability and tumor expression level of ribosomal protein gene RPL28 are associated with survival of metastatic colorectal cancer patients. *Sci. Rep.* 9:13008.

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., et al. (2006). The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929–1935. doi: 10.1126/science.1132939

Li, W., Gao, L. N., Song, P. P., and You, C. G. (2020). Development and validation of a RNA binding protein-associated prognostic model for lung adenocarcinoma. *Aging* 12, 3558–3573. doi: 10.18632/aging.102828

Lv, J., Wang, L., Shen, H., and Wang, X. (2018). Regulatory roles of OASL in lung cancer cell sensitivity to *Actinidia chinensis* planch root extract (acRoots). *Cell Biol. Toxicol.* 34, 207–218. doi: 10.1007/s10565-018-9422-4

Markiewski, M. M., Vadrevu, S. K., Sharma, S. K., Chintala, N. K., Ghouse, S., Cho, J. H., et al. (2017). The ribosomal protein S19 suppresses antitumor immune responses via the complement C5a receptor 1. *J. Immunol.* 198, 2989–2999. doi: 10.4049/jimmunol.1602057

Masuda, K., and Kuwano, Y. (2019). Diverse roles of RNA-binding proteins in cancer traits and their implications in gastrointestinal cancers. *Wiley Interdiscip. Rev. RNA* 10:e1520. doi: 10.1002/wrna.1520

Nwosu, Z. C., Megger, D. A., Hammad, S., Sitek, B., Roessler, S., Ebert, M. P., et al. (2017). Identification of the consistently altered metabolic targets in human hepatocellular carcinoma. *Cell Mol. Gastroenterol. Hepatol.* 4, 303–323. doi: 10.1016/j.jcmgh.2017.05.004

Pan, H., Pan, J., Song, S., Ji, L., Lv, H., and Yang, Z. (2019). EXOSC5 as a novel prognostic marker promotes proliferation of colorectal cancer via activating the ERK and AKT pathways. *Front. Oncol.* 9:643. doi: 10.3389/fonc.2019.00643

Pang, J. S., Li, Z. K., Lin, P., Wang, X. D., Chen, G., Yan, H. B., et al. (2019). The underlying molecular mechanism and potential drugs for treatment in papillary renal cell carcinoma: a study based on TCGA and Cmap datasets. *Oncol. Rep.* 41, 2089–2102.

Pflueger, D., Sboner, A., Storz, M., Roth, J., Comperat, E., Bruder, E., et al. (2013). Identification of molecular tumor markers in renal cell carcinomas with TFE3 protein expression by RNA sequencing. *Neoplasia* 15, 1231–1240. doi: 10.1593/neo.131544

Renaud, G., Stenzel, U., Maricic, T., Wiebe, V., and Kelso, J. (2015). deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics* 31, 770–772. doi: 10.1093/bioinformatics/btu719

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Siang, D. T. C., Lim, Y. C., Kyaw, A. M. M., Win, K. N., Chia, S. Y., Degirmenci, U., et al. (2020). The RNA-binding protein HuR is a negative regulator in adipogenesis. *Nat. Commun.* 11:213.

Siegel, R. L., Miller, K. D., and Jemal, A. (2015). Cancer statistics, 2015. *CA Cancer J. Clin.* 65, 5–29. doi: 10.3322/caac.21254

Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27, 431–432. doi: 10.1093/bioinformatics/btq675

Stein, J. E., Lipson, E. J., Cottrell, T. R., Forde, P. M., Anders, R. A., Cimino-Mathews, A., et al. (2020). Pan-tumor pathologic scoring of response to PD-(L)1 blockade. *Clin. Cancer Res.* 26, 545–551. doi: 10.1158/1078-0432.ccr-19-2379

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613.

Valenzuela-Fernandez, A., Cabrero, J. R., Serrador, J. M., and Sanchez-Madrid, F. (2008). HDAC6: a key regulator of cytoskeleton, cell migration and cell-cell interactions. *Trends Cell Biol.* 18, 291–297. doi: 10.1016/j.tcb.2008.04.003

Vickers, A. J., and Elkin, E. B. (2006). Decision curve analysis: a novel method for evaluating prediction models. *Med. Decis. Making* 26, 565–574. doi: 10.1177/0272989x06295361

Wierzbicki, P. M., Klacz, J., Rybarczyk, A., Slebioda, T., Stanislawowski, M., Wronska, A., et al. (2014). Identification of a suitable qPCR reference gene in metastatic clear cell renal cell carcinoma. *Tumour. Biol.* 35, 12473–12487. doi: 10.1007/s13277-014-2566-9

Xie, M., Ma, T., Xue, J., Ma, H., Sun, M., Zhang, Z., et al. (2019). The long intergenic non-protein coding RNA 707 promotes proliferation and metastasis of gastric cancer by interacting with mRNA stabilizing protein HuR. *Cancer Lett.* 443, 67–79. doi: 10.1016/j.canlet.2018.11.032

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118

Zhao, H., Leppert, J. T., and Peehl, D. M. (2016). A protective role for androgen receptor in clear cell renal cell carcinoma based on mining TCGA data. *PLoS One* 11:e0146505. doi: 10.1371/journal.pone.0146505

Zhou, W., Ma, C. X., Xing, Y. Z., and Yan, Z. Y. (2016). Identification of candidate target genes of pituitary adenomas based on the DNA microarray. *Mol. Med. Rep.* 13, 2182–2186. doi: 10.3892/mmr.2016.4785

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 15 October 2020
doi: 10.3389/fgene.2020.584624

Development and Validation of an RNA-Binding Protein-Based Prognostic Model for Ovarian Serous Cystadenocarcinoma

Yunan He[1†], Sen Zeng[2†], Shunjie Hu[3], Fengqian Zhang[1] and Nianchun Shan[4]*

[1] Department of Gynecology and Obstetrics, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China, [2] Department of Neurology, The Third Xiangya Hospital of Central South University, Changsha, China, [3] Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China, [4] Department of Gynecology and Obstetrics, Xiangya Hospital, Central South University, Changsha, China

Ribonucleic acid-binding proteins (RBPs) are reportedly involved in tumor progression and recurrence; however, the functions and mechanisms of action of RBPs in ovarian serous cystadenocarcinoma (OSC) are not known. To address these issues, gene expression profiles of OSC tissues from The Cancer Genome Atlas (TCGA) and normal tissues from the Genotype-Tissue Expression database were compared in order to identify RBPs that are differentially expressed in OSC. We also analyzed the biological functions of these RBPs and their relationship to clinical outcome. There were 190 RBPs that were differentially expressed between OSC and normal tissues, including 93 that were upregulated and 97 that were downregulated. Five of the RBPs were used to construct a prediction model that was evaluated by univariate and multivariate Cox regression analyses. TCGA data were randomly divided into training and test cohorts, and further categorized into high- and low-risk groups according to risk score in the model. The overall survival (OS) of the high-risk group was shorter than that of the low-risk group (training cohort $P = 0.0007596$; test cohort $P = 0.002219$). The area under the receiver operating characteristic curve of the training and test cohorts was 0.701 and 0.638, respectively, demonstrating that the model had good predictive power. A nomogram was established to quantitatively describe the relationship between the five prognostic RBPs and OS in OSC, which can be useful for developing individualized management strategies for patients.

Keywords: ovarian serous cystadenocarcinoma, RNA-binding proteins, prognostic model, overall survival, bioinformatics

OPEN ACCESS

Edited by:
Pradeep Chaluvally Raghavan, Medical College of Wisconsin, United States

Reviewed by:
Ji Heon Noh, Chungnam National University, South Korea
Jasmine George, Medical College of Wisconsin, United States

****Correspondence:***
Nianchun Shan
shannc@csu.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *20 July 2020*
Accepted: *18 September 2020*
Published: *15 October 2020*

Citation:
He Y, Zeng S, Hu S, Zhang F and Shan N (2020) Development and Validation of an RNA-Binding Protein-Based Prognostic Model for Ovarian Serous Cystadenocarcinoma. Front. Genet. 11:584624. doi: 10.3389/fgene.2020.584624

INTRODUCTION

Ovarian cancer, a common gynecologic cancer, accounts for just 3% of newly diagnosed tumors but is the fifth leading cause of cancer-related deaths in women; this is partly attributable to the difficulty of early diagnosis and high rates of metastasis and recurrence (Li et al., 2012; Xiong et al., 2018). Ovarian serous cystadenocarcinoma (OSC) is the most common subtype of ovarian cancer (60%–80% of ovarian epithelial tumors) (Li et al., 2012; Kaldawy et al., 2016). In most cases, OSC is detected at an advanced stage and recurrence after treatment is common

(Torre et al., 2017). There is therefore a need to clarify the molecular mechanisms underlying OSC pathogenesis and progression so that more effective therapeutic strategies can be developed.

Ribonucleic acid (RNA)-binding proteins (RBPs) participate in the formation of the ribonucleoprotein (RNP) complex for protein synthesis (Dreyfuss et al., 2002). Over 1500 RBPs have been identified to date (Gerstberger et al., 2014) and play a critical role in RNA processing by regulating mRNA stability, localization, alternative splicing, polyadenylation, and translation efficiency (Brinegar and Cooper, 2016; Protter and Parker, 2016; Masuda and Kuwano, 2019). Dysregulation of RBP expression has been implicated in numerous human diseases (Brinegar and Cooper, 2016; Newman et al., 2016). For example, mutations in the genes encoding the RBPs Fused in sarcoma (FUS) and TAR DNA-binding protein 43 (TDP-43) have been linked to the pathogenesis of amyotrophic lateral sclerosis, and the proteins were depleted from the nucleus and aggregated in the cytoplasm in affected neurons (Brinegar and Cooper, 2016). The RBPs Elav-like family (CELF) and Muscleblind-like (MBNL) contribute to the pathogenesis of myotonic dystrophy by reverting to fetal expression patterns and promoting fetal mRNA processing in adult tissues (Brinegar and Cooper, 2016).

RBPs are also associated with cancer development, as dysregulation of RBP expression alters the expression of oncogenes and tumor suppression genes (Pereira et al., 2017). Musashi 1 (MSI1) and MSI2 have been shown to increase the levels of Myc and estrogen receptor α1 (ESR1) oncogenes and reduce that of phosphatase and tensin homlog (PTEN) by modulating mRNA stability and protein translation, leading to various types of cancer (Kudinov et al., 2017). LIN-28 homolog B (LIN28B) promotes pluripotency and plays a critical role in colorectal carcinogenesis by interacting with microRNAs of the let-7 family (King et al., 2011; Balzeau et al., 2017). Quaking (QKI), a splicing factor that regulates cell proliferation, is downregulated in lung cancer, which is associated with poor survival (Zong et al., 2014). RNA-binding motif protein 10 (RBM10) is a regulator of alternative splicing in lung adenocarcinoma (Hernandez et al., 2016); and human antigen R (HuR) promotes cell dedifferentiation and proliferation by regulating the stability of target mRNAs in hepatocellular carcinoma (Fernandez-Ramos and Martinez-Chantar, 2015). However, the mechanisms by which most RBPs contribute to carcinogenesis remain unknown.

The aim of the present study was to clarify the role of RBPs in the pathogenesis of OSC. We retrieved RNA sequencing and clinicopathologic data for OSC from The Cancer Genome Atlas (TCGA) database and screened for differentially expressed RBPs. A functional analysis was also carried out in order to identify key RBPs in OSC that can potentially serve as prognostic biomarkers.

MATERIALS AND METHODS

Data Processing

Ribonucleic acid profiles of tumor tissue from OSC patients and normal tissues were obtained from TCGA database. For comparison, we obtained RNA sequences of normal ovarian tissue from the Genotype-Tissue Expression (GTEx) database. RBPs that were differentially expressed between tumor and normal tissues were screened using R v4.0.2 software (The R Project, Vienna, Austria).

Kyoto Encyclopedia of Genes and Genomes Pathway and Gene Ontology Analyses

To determine the biological function of differentially expressed RBPs, we used the R software packages clusterProfiler, org.Hs.eg.db, enrichplot, and ggplot2 to carry out KEGG and GO analyses, which included cellular component (CC), molecular function (MF), and biological process (BP) as functional domains. A q value or false discovery rate < 0.05 was taken as statistically significant.

Protein–Protein Interaction Network Construction

Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) was used to investigate the interactions of RBPs. A PPI network and visual subnetwork were constructed using Cytoscape v3.8.0 software (https://cytoscape.org/index.html). Functionally significant RBPs were identified using the Molecular Complex Detection (MCODE) algorithm. RBPs with MCODE score and node counts > 3 were deemed significant, and P-values < 0.05 were considered statistically significant.

Prognostic Model Construction and Validation

The survival package of R software was used for univariate Cox regression analysis of key RBPs; candidate RBPs were selected with the log-rank test and incorporated into a multivariate Cox regression model. The risk score was calculated according to the following formula: risk score = $\beta 1 \times \mathrm{Exp}_1 + \beta 2 \times \mathrm{Exp}_2 + \ldots + \beta \mathrm{i} \times \mathrm{Exp}_i$. We used R software to construct a nomogram to predict overall survival (OS) of OSC patients. The model was validated using data from TCGA database, which were randomly divided into training and test cohorts. With the median risk score as the cutoff, OSC patients were divided into high- and low-risk groups, and the log-rank test was used to compare differences in OS between them. $P < 0.05$ was considered statistically significant. Receiver operating characteristic (ROC) curve analysis was also performed to evaluate the predictive value of the model, which was validated using data from the Human Protein Atlas (HPA) database.

RESULTS

Identification of Differentially Expressed RBPs in OSC Patients

We investigated the functions and prognostic value of RBPs in OSC patients; the flow diagram of the study is shown in **Figure 1**. We downloaded RNA sequences of 379 OSC patients from TCGA

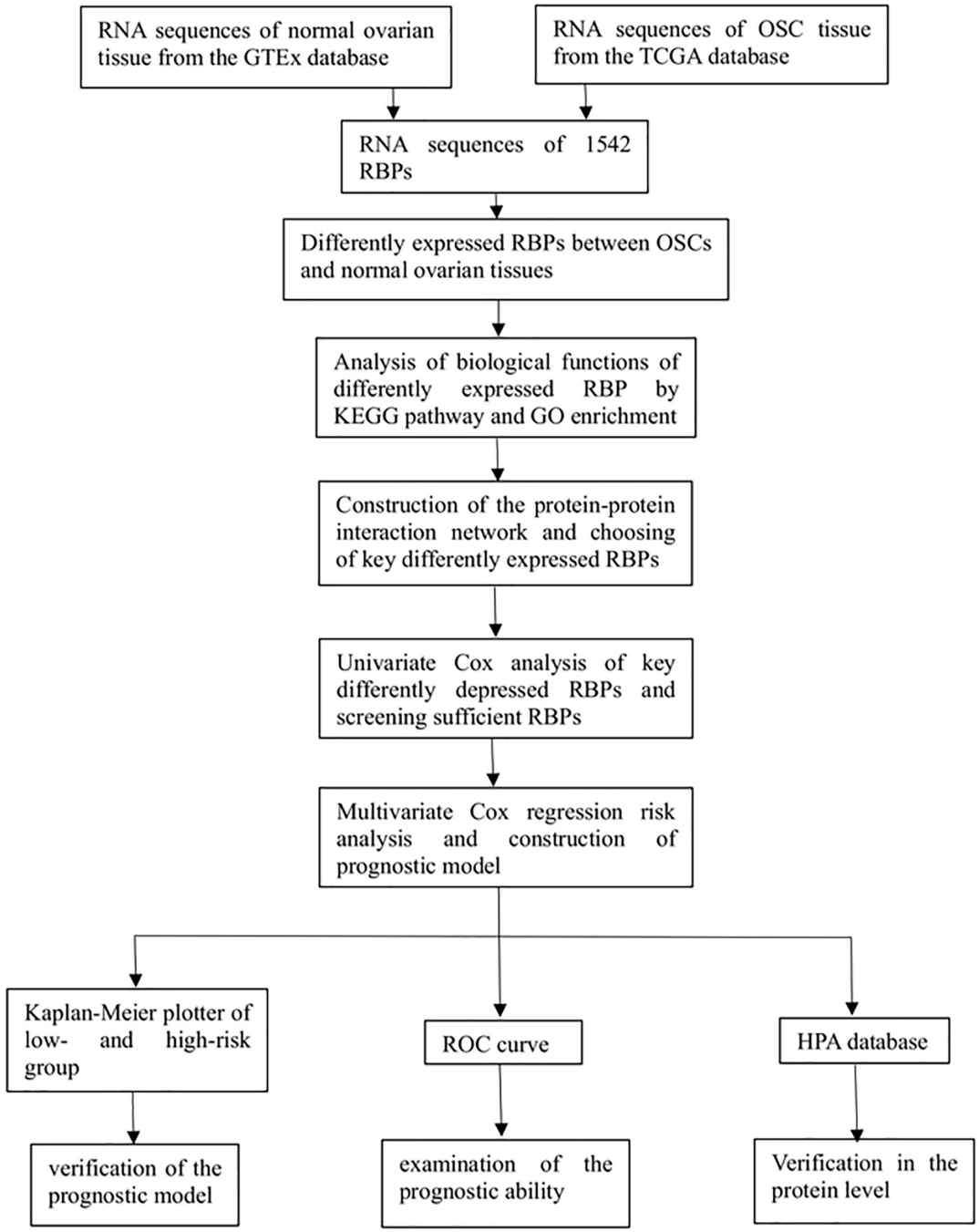


FIGURE 1 | Flow diagram of this study.

database; 88 normal ovarian tissue samples obtained from the GTEx database were used as a control. The RNA sequences of 1542 RBPs (Gerstberger et al., 2014) were ultimately included in the analysis; 190 sequences encoded RBPs that were differentially expressed between normal and tumor tissues ($P < 0.05$, $|\log2 \text{ fold change}| > 1.0$), including 93 upregulated and 97 downregulated RBPs (**Figure 2**). All up- or down-regulated RBP genes in OSC has been listed in the **supplementary files**.

KEGG Pathway Enrichment and GO Analyses of Differentially Expressed RBPs

We used R software to evaluate the enrichment of the identified RBP-encoding genes under biological processes, metabolic mechanisms, and molecular functions. The results of the KEGG analysis showed that the upregulated RBPs were significantly enriched in pathways related to RNA

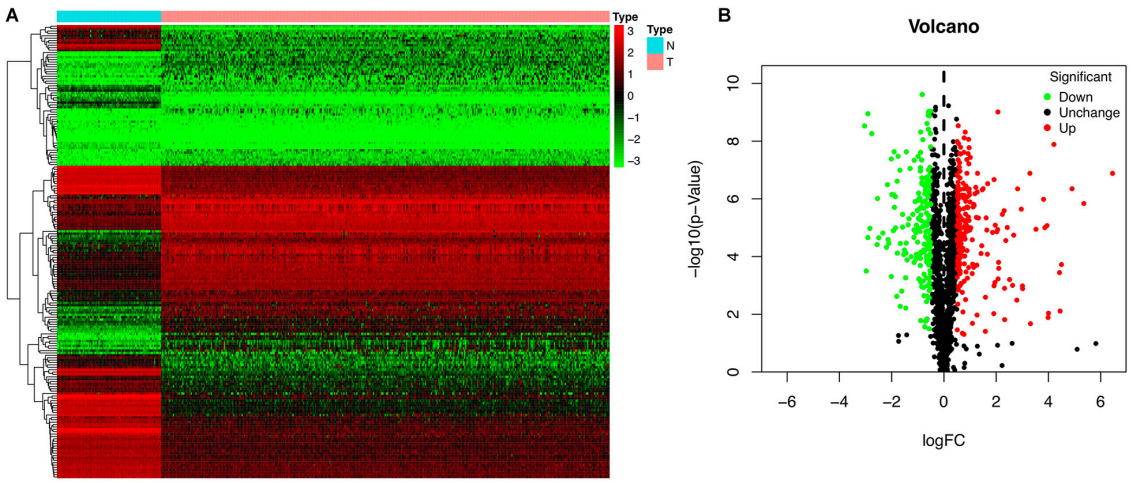


FIGURE 2 | Differentially expressed RBPs in OSC. **(A)** Heat map. **(B)** Volcano plot.

transport, ribosome biogenesis in eukaryotes, and ribosome (**Figure 3A**), whereas downregulated RBPs were enriched in RNA transport, spliceosome, and ribosome (**Figure 3B**). GO analysis revealed that under BP, upregulated RBPs were mainly involved in defense response to virus, RNA catabolic process, and non-coding RNA metabolic process. Meanwhile, downregulated RBPs were involved in RNA splicing; RNA splicing, via transesterification reactions with bulged adenosine as nucleophile; mRNA splicing; and mRNA splicing via spliceosome. Under CC, both upregulated and downregulated RBPs were enriched in RNP granule, cytoplasmic RNP granule, and P-body. Under MF, both upregulated and downregulated RBPs were enriched in catalytic activity, acting on RNA, and mRNA 3'-untranslated region (UTR) binding; upregulated RBPs were also enriched in double-stranded RNA binding (**Figure 3C**), and downregulated RBPs were enriched in translation regulator activity and nucleic acid binding (**Figure 3D**).

PPI Network Construction and Key Module Selection

To investigate the interactions of differentially expressed RBPs and identify key RBPs related to OSC, we constructed a PPI network using data from the STRING database and Cytoscape software. The PPI network included 190 nodes and 493 edges. A coexpression network was constructed using the MCODE tool and the top 3 modules and genes were selected and visualized according to their risk scores (**Figure 4**). The RBPs in the key modules were associated with the defense response to virus, translation, and RNA binding.

Construction of a RBP-Based Prediction Model for OSC

We analyzed the RNA sequences of 9 RBPs selected from the PPI network and evaluated their clinical and prognostic significance. The results of the univariate Cox regression analysis revealed that five of the RBPs (mitochondrial ribosomal protein L14 [MRPL14], zinc finger protein 239 [ZNF239], proteasome 20S subunit α6 [PSMA6], poly[RC]-binding protein 3 [PCBP3], and ribosomal protein S4 Y-linked 1 [RPS4Y1]) were related to prognosis in OSC. To further assess their influence on OS, we performed a multivariate Cox regression analysis and found that the five RBPs were independent predictors of OS in OSC patients (**Figure 5**). We constructed a prediction model by calculating the risk score for each patient using the following formula: risk score = (−0.34749 × Exp[MRPL14]) + (−0.17478 × Exp[ZNF239]) + (−0.47382 × Exp[PSMA6]) + (0.41487 × Exp[PCBP3]) + (3.46278 × Exp[RPS4Y1]). A total of 379 OSC patients in TCGA were randomly divided into training and test cohorts and further classified into low- and high-risk subgroups according to median risk score. To evaluate the predictive value of our model, we performed a survival analysis of the cases. In both the training and test cohorts, the high-risk group had shorter OS than the low-risk group (training cohort $P = 0.0007596$, test cohort $P = 0.002219$) (**Figures 6A**, **7A**). The heatmap of RBP expression, survival status, and risk scores of the low- and high-risk subgroups of the training and test cohorts are shown in **Figures 6C–E**, **7C–E**. The time-dependent ROC curve analysis showed that the area under the ROC curve of the RBP-based risk score model was 0.701 and 0.638 for the training and test cohorts, respectively (**Figures 6B**, **7B**), indicating a moderate predictive power.

Construction of a Nomogram Based on RBPs

A nomogram was constructed to quantitatively assess the role of the five RBPs in the prediction model for OSC patient survival (**Figure 8**). Based on the multivariate Cox analysis, we assigned scores of each variable to the scale of the nomogram, determined the score of each variable, and calculated the total scores of the five RBPs for each patient. The total score was normalized to a distribution ranging from 0 to 100 and used to calculate the 1-year, 3-year, and 5-year estimated OS rates of OSC patients.

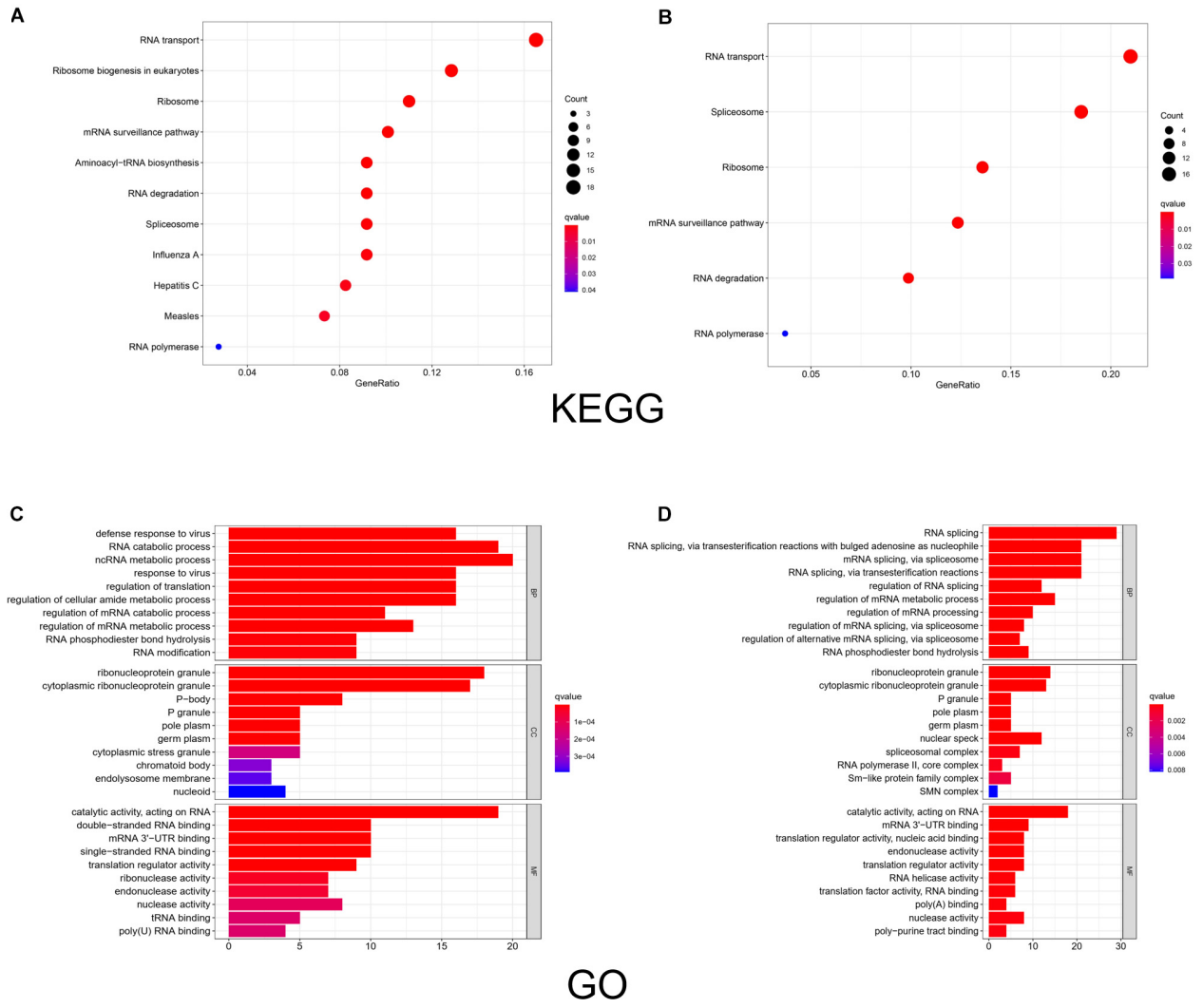


FIGURE 3 | KEGG pathway analysis and GO enrichment analysis of differentially expressed RBPs. The size of a dot in the KEGG analysis represents the relative abundance of the corresponding RBP. The color represents the *q* value, with a darker red color indicating a smaller *q* value. And the color bar in the legend of GO analysis represents the q value of the corresponding item. BP, biological process; CC, cellular component; MF, molecular function. **(A)** KEGG analysis of upregulated RBPs. **(B)** KEGG analysis of downregulated RBPs. **(C)** GO enrichment analysis of upregulated RBPs. **(D)** GO enrichment analysis of downregulated RBPs.

We also evaluated the prognostic significance of various clinical characteristics of OSC patients in TCGA by Cox regression analysis. The univariate analysis showed that risk scores were independent risk factors for OS (training cohort $P < 0.001$, test cohort $P = 0.010$), while age and tumor grade were unrelated to OS (**Figure 9**). The multivariate regression analysis showed that risk scores were independent prognostic factors for OS in OSC patients (training cohort $P < 0.001$, test cohort $P = 0.007$) (**Figure 10**).

Expression of Prognostic RBPs

In order to clarify the expression of the five prognostic RBPs in OSC patients, we examined immunohistochemistry data from the HPA database. MRPL14 was highly expressed in OSC tissue compared to normal tissue. In contrast, the immunoreactivity of PSMA6, PCBP3, and RPS4Y1 in OSC tissue was relatively low (**Figure 11**). ZNF239 protein expression data were not available in the HPA.

DISCUSSION

Only a small fraction of RBPs have been identified as being related to tumor recurrence and progression, and in most cases the mechanism of action has not been reported. Bioinformatics approaches allow investigation of the diagnostic or prognostic significance of changes in RBP expression. Our study identified 190 RBPs that were differentially expressed between OSC and normal tissues. Five of the RBPs were used to construct a risk prediction model,

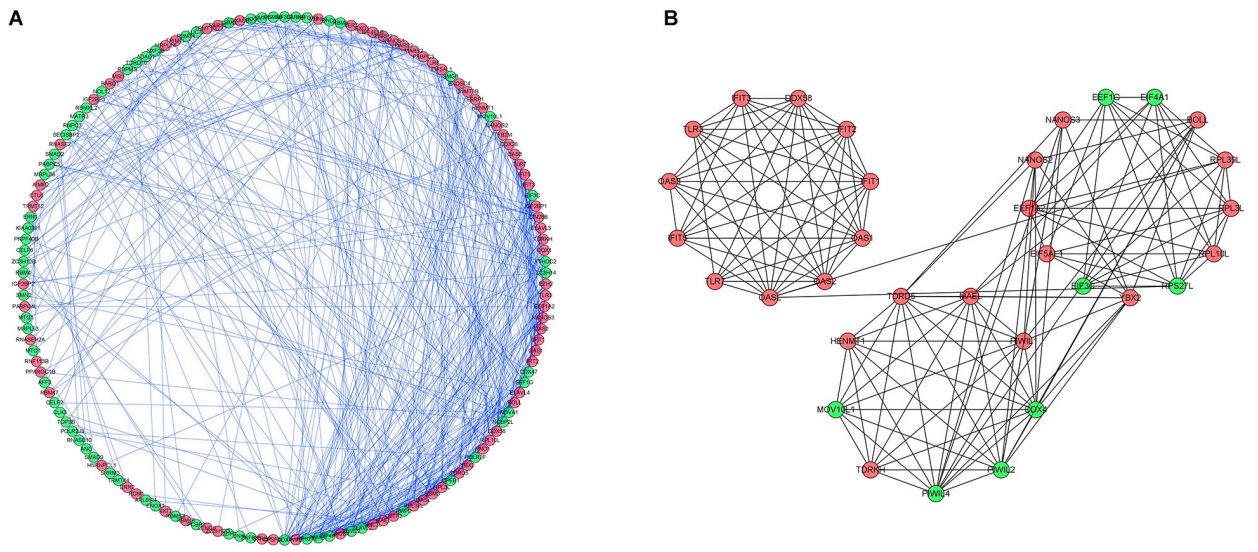


FIGURE 4 | Protein-protein interaction network and modules analysis. **(A)** Protein–protein interaction network of differentially expressed RBPs. **(B)** Three critical modules from PPI network. Green circles: downregulation with a fold change of more than 2. Red circles: upregulation with fold change of more than 2.

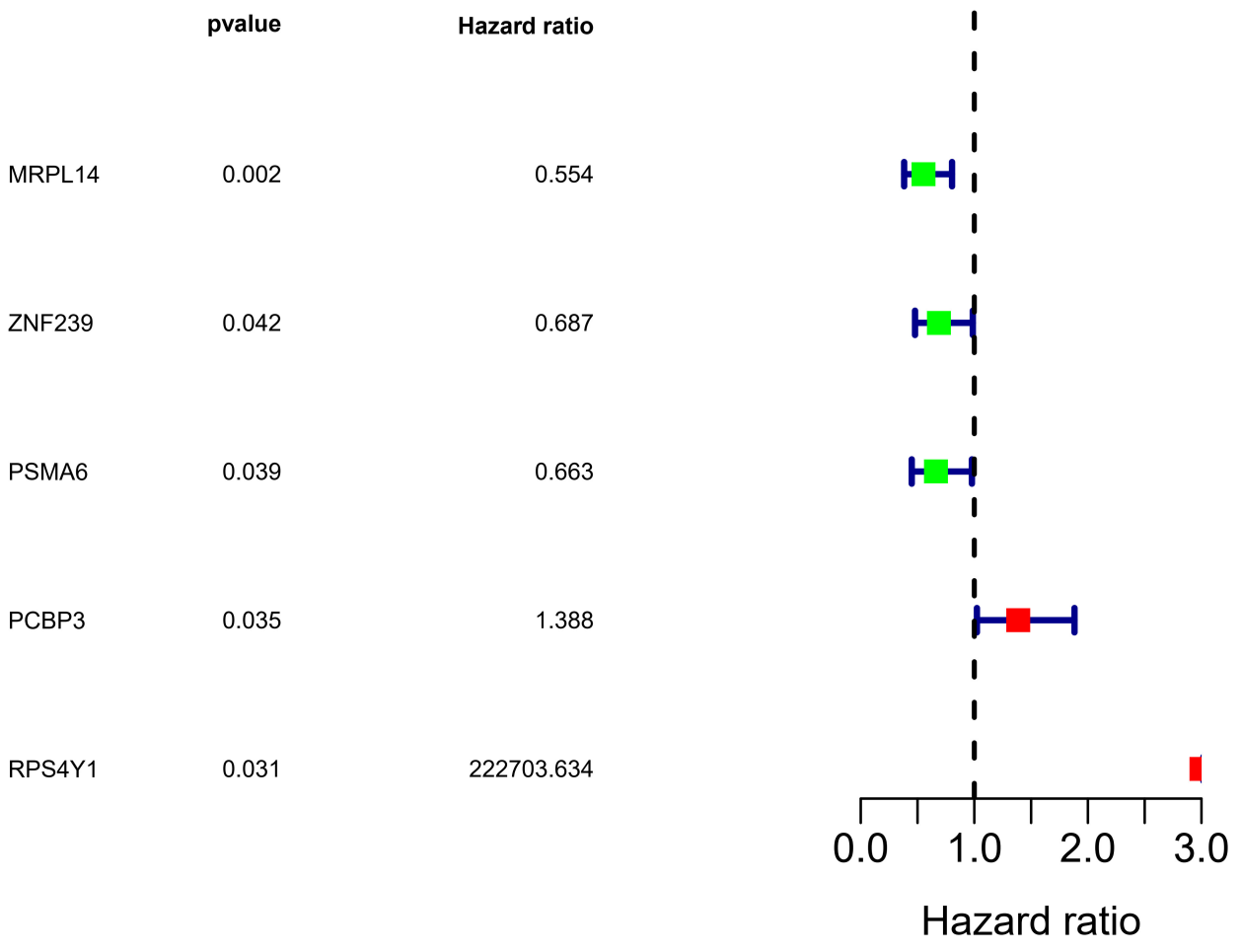


FIGURE 5 | Identification of prognosis-related RBPs by multivariate Cox regression analysis.

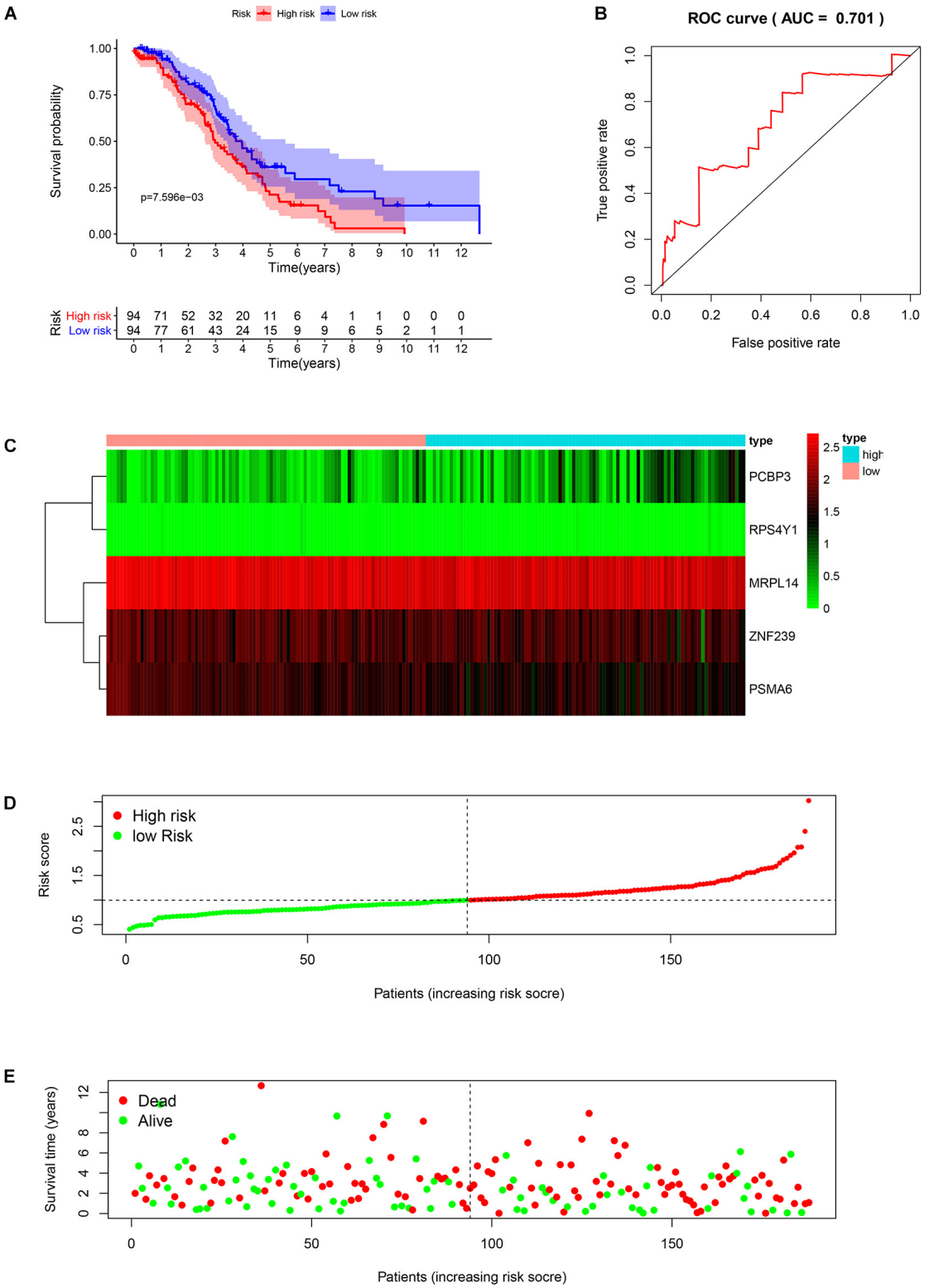


FIGURE 6 | Risk score analysis of the training cohort in TCGA using the 5-gene prognostic model. **(A)** Survival curve for low- and high-risk subgroups. **(B)** ROC curve for predicting OS based on risk score. **(C)** Expression heat map. **(D)** Risk score distribution. Patients were assigned to the training group based on risk score for determination of median risk score. **(E)** Survival status. The dashed line represents the median risk score; most patients on the right side had died, revealing a trend of greater risk of death with increasing risk score.

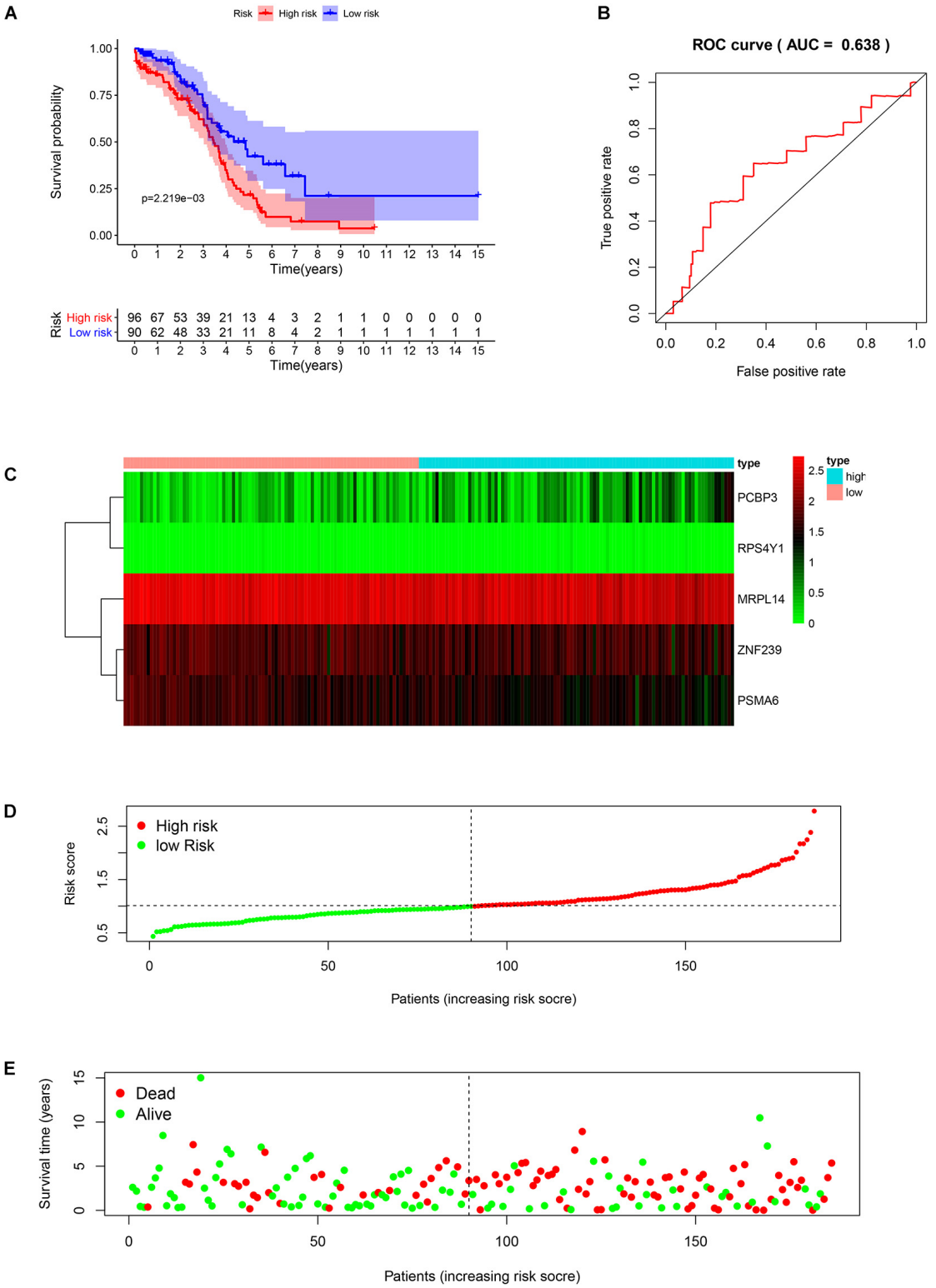


FIGURE 7 | Risk score analysis of the test cohort in TCGA using the 5-gene prognostic model. **(A)** Survival curve for low- and high-risk subgroups. **(B)** ROC curve for predicting OS based on risk score. **(C)** Expression heat map. **(D)** Risk score distribution. Patients were assigned to the training group based on risk score for determination of median risk score. **(E)** Survival status. The dashed line represents the median risk score; most patients on the right side had died, revealing a trend of greater risk of death with increasing risk score.

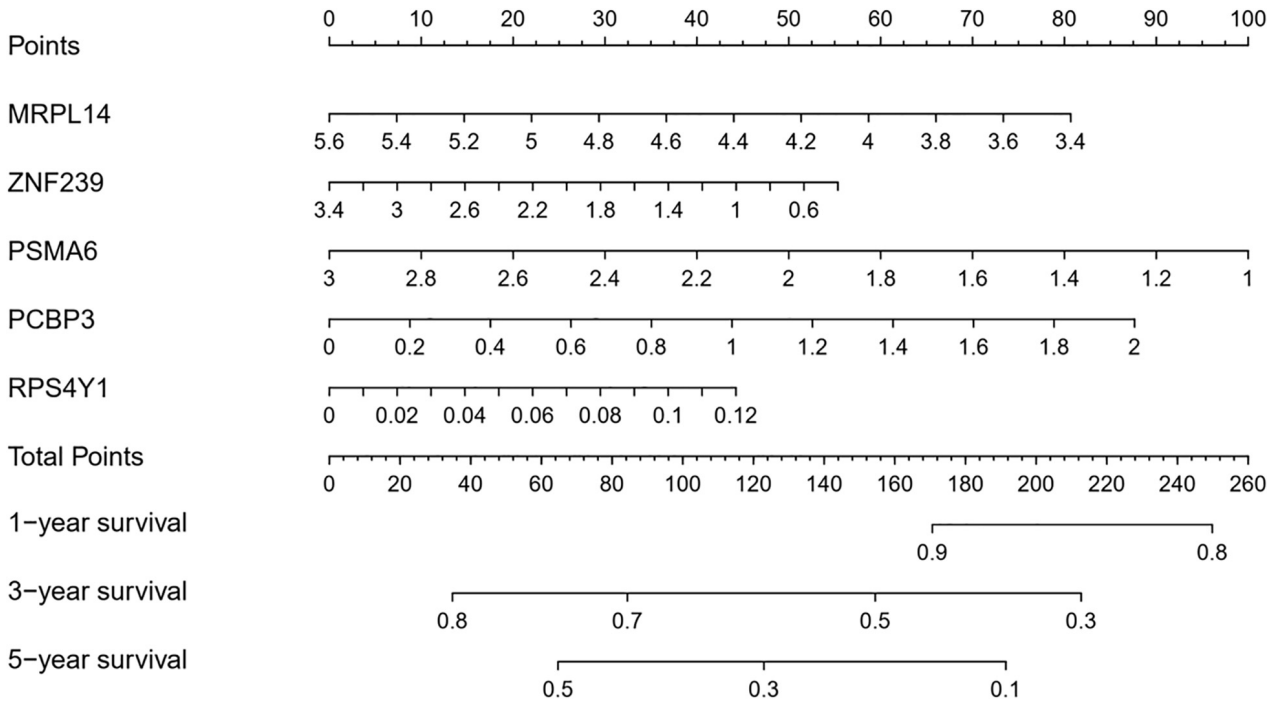


FIGURE 8 | Nomogram for predicting 1-year, 3-year, and 5-year OS of OSC patients in TCGA database.

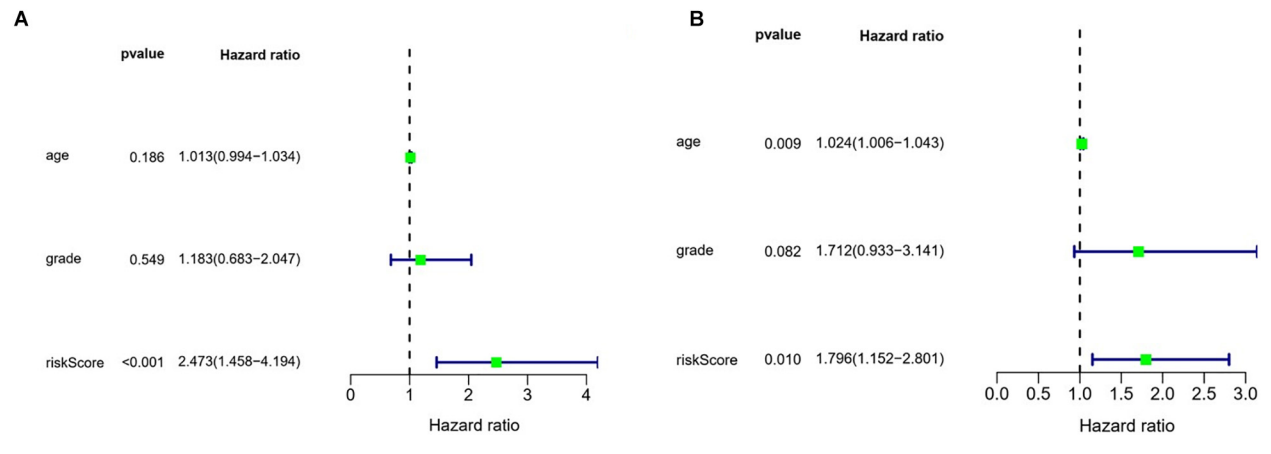


FIGURE 9 | Evaluation of the prognostic value of different clinical parameters by univariate analysis. **(A)** Training cohort. **(B)** Test cohort.

which showed moderate performance in predicting OSC patient survival.

The results of the GO and KEGG pathway analyses revealed that the differentially expressed RBPs were significantly enriched in the defense response to virus as well as RNA- and protein translation-related processes that have been linked to the pathogenesis of various human diseases (Scotti and Swanson, 2016; Anastasiadou et al., 2018; Grafanaki et al., 2019). RBPs form the RNP complex that regulates RNA stability and hence, gene expression; dysfunction of the RNP complex can lead to cancer development and progression (Carotenuto et al., 2019). The RBP ribonucleoprotein 1, translational regulator 1 (LARP1) promotes ovarian cancer progression and by altering the stability of its target mRNAs B cell lymphoma 2 (BCL2) and BCL-2–interacting killer (BIK) (Hopkins et al., 2016). LIN28B inhibits the apoptosis of ovarian cancer cells and promotes cancer progression by binding to AKT2 mRNA and increasing the expression of the protein (Lin et al., 2018).

The PPI of the differentially expressed RBPs identified in this study reveals an important role for these proteins in tumorigenesis and cancer progression. Eukaryotic translation elongation factor 1 α2 (EEF1A2) is an oncogene that promotes ovarian carcinogenesis and

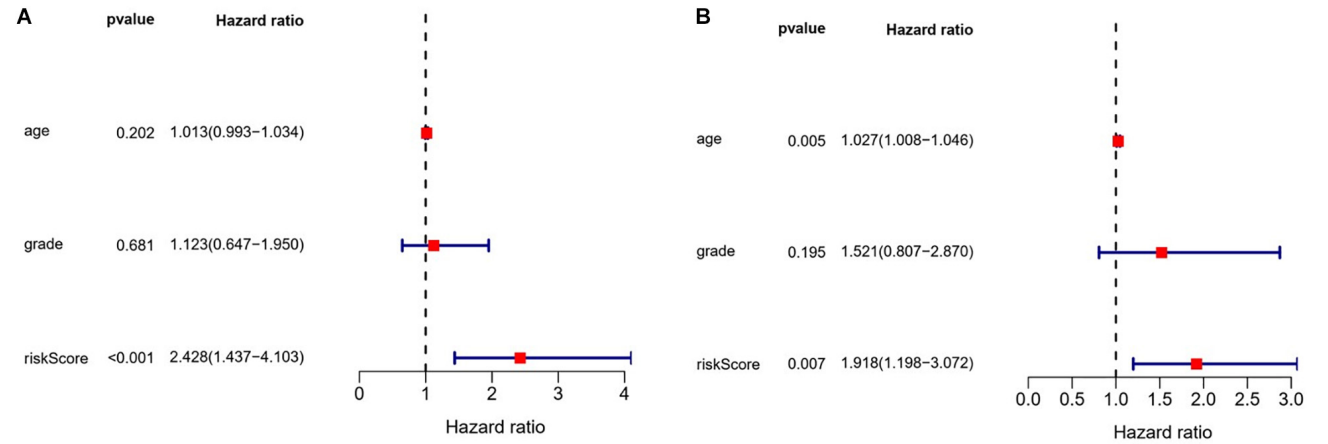


FIGURE 10 | Evaluation of the prognostic value of different clinical parameters by multivariate analysis. **(A)** Training cohort. **(B)** Test cohort.

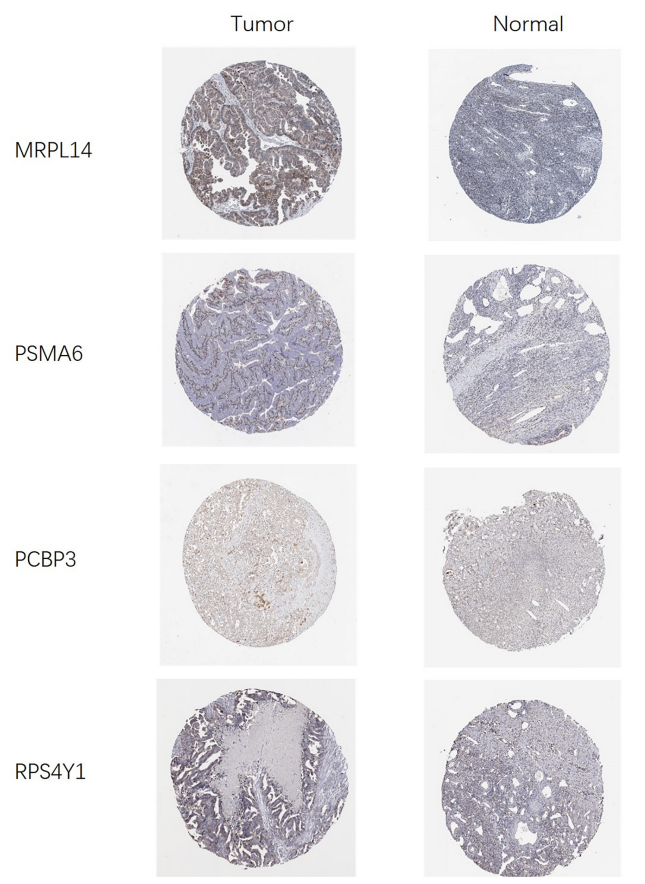


FIGURE 11 | Validation of prognostic RBP expression in OSC and normal ovarian tissue using the HPA database.

inhibits apoptosis of ovarian cancer cells (Lee, 2003). Toll-like receptor 3 (TLR3) was shown to play a dual role in ovarian cancer by eliminating tumor cells via upregulation of interferons and activation of natural killer cells and also by promoting cancer development (Husseinzadeh and Davenport, 2014).

Five of the differentially expressed RBPs, namely MRPL14, ZNF239, PSMA6, PCBP3, and RPS4Y1, showed prognostic value in OSC by univariate and multivariate Cox regression analyses. MPRL14 was found to be upregulated in tumor cells and its expression was positively correlated with the outcome of OSC patients. Reduced ZNF239 and PSMA6 and elevated PCBP3 and RPS4Y1 levels were associated with worse prognosis. The mitochondrial ribosomal proteins (MRPs) are the counterpart of cytoplasmic ribosomes relating to maintain mitochondrial DNA stability (O'Brien et al., 1999). The MRPL14 single nucleotide polymorphism may be related to diabetic retinopathy through steroid metabolism or insulin resistance (Lin et al., 2013). MRPL14 is highly expressed in thyroid tumor (Jacques et al., 2013), but does not reveal the relationship with prognosis. In the past 5 years, no relationship between ZNF239 and any type of tumor has been reported. The proteasome gene, PMSA6, encodes the a1 protein, which is involved in the formation of the outer rings of the 20s core proteasome and is subject to post-translational regulation (Choudhary et al., 2009; Wang et al., 2013). The location of the PSMA6 gene occurs in a region containing microsatellites that have been implicated in coronary artery disease (CAD) (Alsmadi et al., 2009), type 2 diabetes mellitus (T2DM) (Sjakste et al., 2007), Grave's disease (Sjakste et al., 2004), asthma (Zemeckiene et al., 2015), ankylosing spondylitis (Zhao et al., 2015), and myocardial infarction (Liu et al., 2009). In a lung cancer study, the expression of PSMA6 was up-regulated, and knocking out PSMA6 could induce lung cancer tumor cell apoptosis or the cell cycle to enter the arrest phase (Kakumu et al., 2017). However, in our study, the expression of PSMA6 in OSC is down-regulated, and the low expression of PSMA6 is associated with a worse OSC prognosis, which may be due to the different effects of PSMA6 expression on proteasome activity. PSMA6 has carcinogenic effects in various tissue tumors. Actually, the ubiquitination-proteasome degradation pathway has been proved to be the key to cell survival and

proliferation. Therefore, the detailed molecular mechanism of PSMA6 in OSC needs to be revealed. The poly(C) binding proteins (PCBPs), an RNA-binding protein involved in post-transcriptional regulation, whose important functions are mRNA activation, translation activation and translation silencing (Makeyev and Liebhaber, 2002). A study of pancreatic ductal carcinoma showed that the content of PCBP3 protein in postoperative tissues was significantly related to the survival time of patients, and the prognosis of the group with lower PCBP3 protein content was worse (Ger et al., 2018). This is consistent with the results of our study. Otherwise, the initiation of RPS4Y1 expression is the basis of Y chromosome activation (Zhou et al., 2019). There is currently no report on the relationship between RPS4Y1 and tumors. However, studies about these five RBP genes in ovarian cancer are rarely seen and the molecular link between these five RBPs and OSC progression has yet to be elucidated. Clinical specimen validation and follow up data of OSC patients are also wanted in the following research. The results of the ROC curve analysis indicated that the five RBPs showed moderate performance in identifying OSC patients who are at risk of progression; the nomogram model constructed to predict 1-year, 3-year, and 5-year OS in OSC patients yielded similar findings.

This study had some limitations. Firstly, the prediction model was based on TCGA data and no clinical validation or prospective clinical study was conducted; moreover, the limited clinical information in the TCGA dataset may have diminished the reliability of the Cox regression analysis. Nonetheless, our model based on five RBPs showed great potential being used to predict OSC patient prognosis, which can inform clinical decisions and lead to better outcomes.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study; these data can be found here: The Cancer Genome Atlas (https://portal.gdc.cancer.gov/).

AUTHOR CONTRIBUTIONS

YH and NS designed the study. SZ and YH performed the experiments. YH and SH analyzed the data. YH, SH, and FZ wrote the manuscript. All authors reviewed the final version of the article.

ACKNOWLEDGMENTS

Thanks for the establishment and sharing of TCGA and GTEx databases. We also thank the people who helped us fix the code bugs during the statistical analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.584624/full#supplementary-material

REFERENCES

Alsmadi, O., Muiya, P., Khalak, H., Al-Saud, H., Meyer, B. F., Al-Mohanna, F., et al. (2009). Haplotypes encompassing the KIAA0391 and PSMA6 gene cluster confer a genetic link for myocardial infarction and coronary artery disease. *Ann. Hum. Genet* 73(Pt 5), 475–483. doi: 10.1111/j.1469-1809.2009.00534.x

Anastasiadou, E., Jacob, L. S., and Slack, F. J. (2018). Non-coding RNA networks in Cancer. *Nat. Rev. Cancer.* 18, 5–18. doi: 10.1038/nrc.2017.99

Balzeau, J., Menezes, M. R., Cao, S., and Hagan, J. P. (2017). The LIN28/let-7 pathway in Cancer. *Front. Genet.* 8:31. doi: 10.3389/fgene.2017.00031

Brinegar, A. E., and Cooper, T. A. (2016). Roles for RNA-binding proteins in development and disease. *Brain Res.* 1647, 1–8. doi: 10.1016/j.brainres.2016.02.050

Carotenuto, P., Pecoraro, A., Palma, G., Russo, G., and Russo, A. (2019). Therapeutic approaches targeting nucleolus in Cancer. *Cells* 8:1090. doi: 10.3390/cells8091090

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., et al. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325, 834–840. doi: 10.1126/science.1175371

Dreyfuss, G., Kim, V. N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell Biol.* 3, 195–205. doi: 10.1038/nrm760

Fernandez-Ramos, D., and Martinez-Chantar, M. L. (2015). NEDDylation in liver Cancer: the regulation of the RNA binding protein Hu antigen R. *Pancreatology* 15(4 Suppl.), S49–S54. doi: 10.1016/j.pan.2015.03.006

Ger, M., Kaupinis, A., Petrulionis, M., Kurlinkus, B., Cicenas, J., Sileikis, A., et al. (2018). Proteomic identification of FLT3 and PCBP3 as potential prognostic biomarkers for pancreatic Cancer. *Anticancer Res.* 38, 5759–5765. doi: 10.21873/anticanres.12914

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. *Nat. Rev. Genet.* 15, 829–845. doi: 10.1038/nrg3813

Grafanaki, K., Anastasakis, D., Kyriakopoulos, G., Skeparnias, I., Georgiou, S., and Stathopoulos, C. (2019). Translation regulation in skin Cancer from a tRNA point of view. *Epigenomics* 11, 215–245. doi: 10.2217/epi-2018-0176

Hernandez, J., Bechara, E., Schlesinger, D., Delgado, J., Serrano, L., and Valcarcel, J. (2016). Tumor suppressor properties of the splicing regulatory factor RBM10. *RNA Biol.* 13, 466–472. doi: 10.1080/15476286.2016.1144004

Hopkins, T. G., Mura, M., Al-Ashtal, H. A., Lahr, R. M., Abd-Latip, N., Sweeney, K., et al. (2016). The RNA-binding protein LARP1 is a post-transcriptional regulator of survival and tumorigenesis in ovarian Cancer. *Nucleic Acids Res.* 44, 1227–1246. doi: 10.1093/nar/gkv1515

Husseinzadeh, N., and Davenport, S. M. (2014). Role of toll-like receptors in cervical, endometrial and ovarian Cancers: a review. *Gynecol. Oncol.* 135, 359–363. doi: 10.1016/j.ygyno.2014.08.013

Jacques, C., Guillotin, D., Fontaine, J. F., Franc, B., Mirebeau-Prunier, D., Fleury, A., et al. (2013). DNA microarray and miRNA analyses reinforce the classification of follicular thyroid tumors. *J. Clin. Endocrinol. Metab.* 98, E981–E989. doi: 10.1210/jc.2012-4006

Kakumu, T., Sato, M., Goto, D., Kato, T., Yogo, N., Hase, T., et al. (2017). Identification of proteasomal catalytic subunit PSMA6 as a therapeutic target for lung Cancer. *Cancer Sci.* 108, 732–743. doi: 10.1111/cas.13185

Kaldawy, A., Segev, Y., Lavie, O., Auslender, R., Sopik, V., and Narod, S. A. (2016). Low-grade serous ovarian Cancer: a review. *Gynecol. Oncol.* 143, 433–438. doi: 10.1016/j.ygyno.2016.08.320

King, C. E., Cuatrecasas, M., Castells, A., Sepulveda, A. R., Lee, J. S., and Rustgi, A. K. (2011). LIN28B promotes colon Cancer progression and metastasis. *Cancer Res.* 71, 4260–4268. doi: 10.1158/0008-5472.CAN-10-4637

Kudinov, A. E., Karanicolas, J., Golemis, E. A., and Boumber, Y. (2017). Musashi RNA-binding proteins as Cancer drivers and novel therapeutic targets. *Clin. Cancer Res.* 23, 2143–2153. doi: 10.1158/1078-0432.CCR-16-2728

Lee, J. M. (2003). The role of protein elongation factor eEF1A2 in ovarian Cancer. *Reprod. Biol. Endocrinol.* 1:69. doi: 10.1186/1477-7827-1-69

Li, J., Fadare, O., Xiang, L., Kong, B., and Zheng, W. (2012). Ovarian serous carcinoma: recent concepts on its origin and carcinogenesis. *J. Hematol. Oncol.* 5:8. doi: 10.1186/1756-8722-5-8

Lin, H. J., Huang, Y. C., Lin, J. M., Wu, J. Y., Chen, L. A., and Tsai, F. J. (2013). Association of genes on chromosome 6, GRIK2, TMEM217 and TMEM63B (linked to MRPL14) with diabetic retinopathy. *Ophthalmologica* 229, 54–60. doi: 10.1159/000342616

Lin, X., Shen, J., Dan, P., He, X., Xu, C., Chen, X., et al. (2018). RNA-binding protein LIN28B inhibits apoptosis through regulation of the AKT2/FOXO3A/BIM axis in ovarian Cancer cells. *Signal Transduct. Target Ther.* 3:23. doi: 10.1038/s41392-018-0026-5

Liu, X., Wang, X., Shen, Y., Wu, L., Ruan, X., Lindpaintner, K., et al. (2009). The functional variant rs1048990 in PSMA6 is associated with susceptibility to myocardial infarction in a Chinese population. *Atherosclerosis* 206, 199–203. doi: 10.1016/j.atherosclerosis.2009.02.004

Makeyev, A. V., and Liebhaber, S. A. (2002). The poly(C)-binding proteins: a multiplicity of functions and a search for mechanisms. *RNA* 8, 265–278. doi: 10.1017/s1355838202024627

Masuda, K., and Kuwano, Y. (2019). Diverse roles of RNA-binding proteins in Cancer traits and their implications in gastrointestinal Cancers. *Wiley Interdiscip. Rev. RNA.* 10, e1520. doi: 10.1002/wrna.1520

Newman, R., McHugh, J., and Turner, M. (2016). RNA binding proteins as regulators of immune cell biology. *Clin. Exp. Immunol.* 183, 37–49. doi: 10.1111/cei.12684

O'Brien, T. W., Fiesler, S. E., Denslow, N. D., Thiede, B., Wittmann-Liebold, B., Mougey, E. B., et al. (1999). Mammalian mitochondrial ribosomal proteins (2). Amino acid sequencing, characterization, and identification of corresponding gene sequences. *J. Biol. Chem.* 274, 36043–36051. doi: 10.1074/jbc.274.51.36043

Pereira, B., Billaud, M., and Almeida, R. (2017). RNA-binding proteins in Cancer: old players and new actors. *Trends Cancer* 3, 506–528. doi: 10.1016/j.trecan.2017.05.003

Protter, D. S. W., and Parker, R. (2016). Principles and properties of stress granules. *Trends Cell Biol.* 26, 668–679. doi: 10.1016/j.tcb.2016.05.004

Scotti, M. M., and Swanson, M. S. (2016). RNA mis-splicing in disease. *Nat. Rev. Genet.* 17, 19–32. doi: 10.1038/nrg.2015.3

Sjakste, T., Eglite, J., Sochnevs, A., Marga, M., Pirags, V., Collan, Y., et al. (2004). Microsatellite genotyping of chromosome 14q13.2-14q13 in the vicinity of proteasomal gene PSMA6 and association with Graves' disease in the Latvian population. *Immunogenetics* 56, 238–243. doi: 10.1007/s00251-004-0687-9

Sjakste, T., Kalis, M., Poudziunas, I., Pirags, V., Lazdins, M., Groop, L., et al. (2007). Association of microsatellite polymorphisms of the human 14q13.2 region with type 2 diabetes mellitus in latvian and finnish populations. *Ann. Hum. Genet.* 71(Pt 6), 772–776. doi: 10.1111/j.1469-1809.2007.00372.x

Torre, L. A., Islami, F., Siegel, R. L., Ward, E. M., and Jemal, A. (2017). Global Cancer in women: burden and trends. *Cancer Epidemiol. Biomarkers Prev.* 26, 444–457. doi: 10.1158/1055-9965.EPI-16-0858

Wang, H., Jiang, M., Zhu, H., Chen, Q., Gong, P., Lin, J., et al. (2013). Quantitative assessment of the influence of PSMA6 variant (rs1048990) on coronary artery disease risk. *Mol. Biol. Rep.* 40, 1035–1041. doi: 10.1007/s11033-012-2146-2

Xiong, D. D., Qin, Y., Xu, W. Q., He, R. Q., Wu, H. Y., Wei, D. M., et al. (2018). A network pharmacology-based analysis of multi-target, multi-pathway, multi-compound treatment for ovarian serous cystadenocarcinoma. *Clin. Drug Investig.* 38, 909–925. doi: 10.1007/s40261-018-0683-8

Zemeckiene, Z., Sitkauskiene, B., Gasiuniene, E., Paramonova, N., Tamasauskiene, L., Vitkauskiene, A., et al. (2015). Evaluation of proteasomal gene polymorphisms in Lithuanian patients with asthma. *J. Asthma.* 52, 447–452. doi: 10.3109/02770903.2014.982761

Zhao, H., Wang, D., Fu, D., and Xue, L. (2015). Predicting the potential ankylosing spondylitis-related genes utilizing bioinformatics approaches. *Rheumatol. Int.* 35, 973–979. doi: 10.1007/s00296-014-3178-9

Zhou, Q., Wang, T., Leng, L., Zheng, W., Huang, J., Fang, F., et al. (2019). Single-cell RNA-seq reveals distinct dynamic behavior of sex chromosomes during early human embryogenesis. *Mol. Reprod. Dev.* 86, 871–882. doi: 10.1002/mrd.23162

Zong, F. Y., Fu, X., Wei, W. J., Luo, Y. G., Heiner, M., Cao, L. J., et al. (2014). The RNA-binding protein QKI suppresses Cancer-associated aberrant splicing. *PLoS Genet.* 10:e1004289. doi: 10.1371/journal.pgen.1004289

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 16 October 2020
doi: 10.3389/fgene.2020.580149

Molecular Characterization and Clinical Relevance of RNA Binding Proteins in Colorectal Cancer

Zhen Zhang[1,2†], Ling Wang[3†], Quan Wang[1,2], Mengmeng Zhang[1,2], Bo Wang[1,2], Kewei Jiang[1,2], Yingjiang Ye[1,2], Shan Wang[1,2] and Zhanlong Shen[1,2*]*

[1] Department of Gastroenterological Surgery, Peking University People's Hospital, Beijing, China, [2] Laboratory of Surgical Oncology, Beijing Key Laboratory of Colorectal Cancer Diagnosis and Treatment Research, Peking University People's Hospital, Beijing, China, [3] Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Hepatopancreatobiliary Surgery Department I, Peking University Cancer Hospital and Institute, Beijing, China

Abnormal expression of RNA binding proteins (RBPs) has been reported across various cancers. However, the potential role of RBPs in colorectal cancer (CRC) remains unclear. In this study, we performed a systematic bioinformatics analysis of RBPs in CRC. We downloaded CRC data from The Cancer Genome Atlas (TCGA) database. Our analysis identified 242 differentially expressed RBPs between tumor and normal tissues, including 200 upregulated and 42 downregulated RBPs. Next, we found eight RBPs (RRS1, PABPC1L, TERT, SMAD6, UPF3B, RP9, NOL3, and PTRH1) related to the prognoses of CRC patients. Among these eight prognosis-related RBPs, four RBPs (NOL3, PTRH1, UPF3B, and SMAD6) were selected to construct a prognostic risk score model. Furthermore, our results indicated that the prognostic risk score model accurately predicted the prognosis of CRC patients [area under the receiver operating characteristic curve (AUC)for 3- and 5-year overall survival (OS) and was 0.645 and 0.672, respectively]. Furthermore, we developed a nomogram based on a prognostic risk score model. The nomogram was able to demonstrate the wonderful performance in predicting 3- and 5-year OS. Additionally, we validated the clinical value of four risk genes in the prognostic risk score model and identified that these risk genes were associated with tumorigenesis, lymph node metastasis, distant metastasis, clinical stage, and prognosis. Finally, we used the TIMER and Human Protein Atlas (HPA)database to validate the expression of four risk genes at the transcriptional and translational levels, respectively, and used a clinical cohort to validate the roles of NOL3 and UPF3B in predicting the prognosis of CRC patients. In summary, our study demonstrated that RBPs have an effect on CRC tumor progression and might be potential prognostic biomarkers for CRC patients.

Keywords: colorectal cancer, RNA binding protein, prognostic model, transcriptomics, TCGA

OPEN ACCESS

Edited by:
Yongsheng Kevin Li,
Harbin Medical University, China

Reviewed by:
Punit Prasad,
Institute of Life Sciences (ILS), India
Lun Li,
Fudan University, China
Zhitong Bing,
Lanzhou University, China

****Correspondence:***
Shan Wang
shanwang60@sina.com
Zhanlong Shen
shenlong1977@163.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *04 July 2020*
Accepted: *17 September 2020*
Published: *16 October 2020*

Citation:
Zhang Z, Wang L, Wang Q, Zhang M, Wang B, Jiang K, Ye Y, Wang S and Shen Z (2020) Molecular Characterization and Clinical Relevance of RNA Binding Proteins in Colorectal Cancer. Front. Genet. 11:580149. doi: 10.3389/fgene.2020.580149

INTRODUCTION

Colorectal cancer (CRC) is one of the most common cancers of the gastrointestinal tract. It is the third leading cause of cancer-related death worldwide (Siegel et al., 2020). Although surgical and adjuvant therapies have improved, the 5-year overall survival (OS) rate of CRC patients ranges from 90 to 10% (Van Cutsem et al., 2014). The poor prognosis of CRC is primarily due to tumor distant metastasis and recurrence

(Roth et al., 2010; Kow, 2019). Therefore, understanding the mechanisms that lead to CRC initiation and progression is necessary for diagnosis, therapeutic interventions, and prognostic prediction.

RNA binding proteins (RBPs) are a class of proteins involved in splicing, modifications, transport, localization, stability, degradation, and translation of RNAs (Mitchell and Parker, 2014; Perron et al., 2018). In fact, more than 1,500 human RBPs have been validated by high-throughput screens and experiments (Gerstberger et al., 2014; Neelamraju et al., 2015). RBPs play vital roles in several essential cellular processes by interacting with their target RNAs (Moore, 2005). The target RNAs of RBPs are diverse and include microRNAs, transfer RNAs (tRNAs), small interfering RNA, small nucleolar RNAs, and small nuclear RNAs (Hentze et al., 2018). Abnormally expressed RBPs regulate the expression and function of oncogenes and tumor-suppressor genes via post-transcriptional regulatory mechanisms across various cancers. For example, aberrant hnRNPM expression promoted breast cancer metastasis by controlling CD44 splice isoform switching during epithelial–mesenchymal transition (EMT) (Xu et al., 2014). In melanomas, CPEB4 increased the translation of melanoma drivers MITF and RAB72A, which helps promote tumor proliferation (Pérez-Guijarro et al., 2016). In hepatocellular carcinoma, HuR/methyl-HuR and AUF1 modulate MAT1A and MAT2A expression through post-translational regulation of their messenger RNAs (mRNAs), thus impacting tumor progression (Vázquez-Chantada et al., 2010). In CRC, overexpression of IMP-1 increased proliferation by directly binding to and stabilizing c-Myc (Mongroo et al., 2011).

With the rapid development of high-throughput sequencing, researchers have have been able to perform a systematic functional analysis of RBPs using high-throughput bioinformatics profiling. Recent bioinformatics studies have implied that RBPs can predict the prognosis of different cancers, such as breast cancer, lung adenocarcinoma, glioma, hepatocellular carcinoma, and leukemia (Li et al., 2019a,b, 2020; Saha et al., 2019; Wang K. et al., 2019,Wang Z. et al., 2020). In this study, we downloaded CRC data from the Cancer Genome Atlas (TCGA) database. Next, we selected differentially expressed RBPs to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. Furthermore, we identified prognostic RBPs that enable the construction of a prognostic risk score model. Furthermore, we built a nomogram based on the prognostic risk score model. Finally, we explored the clinical value of these risk genes. Our study detected that several RBPs are involved in CRC, which might be used to predict the prognosis of CRC patients and inhibit tumor progression in the future.

MATERIALS AND METHODS

Data Acquisition

RNA sequencing data and the corresponding clinical data were downloaded from the TCGA database[1]. Overall, 568 CRC patients were used to analyze the differentially expressed RBPs in the TCGA database. Then, 489 CRC patients were selected for clinical analyses, as these patients had complete clinical information, including primary tumor, lymph node metastasis, distant metastasis, clinical stage, and follow-up for at least 1 month. Additionally, we downloaded RNA-seq data of the GSE29623 cohort from the Gene Expression Omnibus (GEO) database[2]. In total, 1,542 RBPs were included in our study (Gerstberger et al., 2014; Neelamraju et al., 2015).

Clinical Samples and Immunohistochemistry Staining

In total, tumor and normal tissues of 44 CRC patients were obtained from the Peking University People's Hospital. All tissues were histopathologically confirmed by pathologists. This study was granted approval by the ethics committee of Peking University People's Hospital. Immunohistochemistry (IHC)staining was performed according to prior published protocols (Zhang et al., 2019). Briefly, the staining index scores were assigned as follows: staining intensity (negative: 0; weak: 1; moderate: 2; strong: 3) and positive staining (<5%: 0; 5–25%: 1; 26–50%: 2; 51–75%: 3; > 75%: 4). The staining index scores were calculated by multiplying the staining intensity score by the positive staining score, which ranged from 0 to 12. The antibodies used in this study included anti-NOL3 (Proteintech, United States) and anti-UPF3B (Proteintech, United States).

Functional Enrichment Analysis of Differentially Expressed RNA Binding Proteins

The GO and KEGG pathway analyses were performed to analyze the biological functions of these differentially expressed RBPs. The GO terms include biological process (BP), cellular component (CC), and molecular function (MF).

Building a Prognostic Model

A univariate Cox regression analysis was performed to identify prognostic RBPs. Subsequently, a multivariable Cox regression analysis was carried out to construct a prognostic risk score model. Furthermore, the risk score was calculated as follows:

$$riskscore = \sum_{i=1}^{n} coefficieent(i)^{*} expression(i),$$

where coefficient(i) and expression(i) represent the regression coefficient and expression levels of selected genes in the prognostic risk score model, respectively. The time-dependent receiver operating characteristic (ROC) analysis was used to assess the prognostic ability of the prognostic risk score model. Then, a nomogram was built to predict the OS of CRC patients. The calibration plot and concordance index (C−index) were used to evaluate the performance of the nomogram.

[1]https://portal.gdc.cancer.gov/

[2]https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29623

External Validation of the Expression and Genetic Alterations of the Risk Genes

The TIMER database[3] and the Human Protein Atlas (HPA) database[4] were utilized to explore and validate gene expression at the transcriptional and translational level, respectively. The cBioportal for Cancer Genomics[5] was used to identify each genetic alteration, which included four CRC studies.

Statistical Analyses

Statistical analyses were performed using the R software v3.5.1. Kaplan–Meier analysis was used to construct survival curves, and the log-rank test was utilized to assess the significance of differences. *T*-test and Wilcoxon signed-rank test were used to explore quantitative variables. $P < 0.05$ represents statistical significance.

RESULTS

Identifying Differentially Expressed RNA Binding Proteins in Colorectal Cancer

Figure 1A shows detailed designs of the study. We downloaded transcriptomic files of CRC from the TCGA database encompassing 568 tumors and 44 normal samples. The Wilcoxon signed-rank test was applied to identify significantly differentially expressed RBPs. The R package Limma was used to identify differentially expressed RBPs according to the following parameters: $|\log_2 FC| > 1$ and false discovery rate (FDR) < 0.05. Among the 1,542 RBPs, we detected 242 differentially expressed RBPs, including 200 upregulated and 42 downregulated RBPs between tumor and normal tissues (**Figures 1B,C**).

Functional Enrichment Analysis of Differentially Expressed RNA Binding Proteins in Colorectal Cancer

To investigate the biological significance of RBPs in CRC, we performed GO and KEGG pathway analyses of 242 differentially expressed RBPs using the R package clusterProfiler. We displayed the top 10 significantly enriched GO terms (**Figure 2A**). Results revealed that these differentially expressed RBPs were significantly enriched in biological processes such as non-coding RNA processing, ribonucleoprotein complex biogenesis, ribosome biogenesis, and ribosomal RNA metabolism. They were mainly located in the preribosome, nucleolar part, small-subunit processome, 90S preribosome, and cytoplasmic ribonucleoprotein granule. They were found to participate in various molecular functions, including catalytic activity (acting on RNA), catalytic activity (acting on a tRNA), ribonuclease activity, single-stranded RNA binding, and poly(U) RNA binding. For KEGG pathway analysis, we found that these differentially expressed RBPs were mainly associated with ribosome biogenesis in eukaryotes, RNA degradation, RNA transport, mRNA surveillance pathway, and aminoacyl-tRNA biosynthesis (**Figure 2B**).

Identifying Prognostic RNA Binding Proteins in Colorectal Cancer

To assess the prognostic significance of RBPs in CRC, we performed a univariate Cox regression analysis of 242 differentially expressed RBPs. We detected eight RBPs (RRS1, PABPC1L, TERT, SMAD6, UPF3B, RP9, NOL3, and PTRH1) that were related to the prognosis of CRC patients ($p < 0.05$, **Figure 3A**). Also, all these eight prognostic RBPs were protective factors because CRC patients with high expression of these RBPs had poor prognosis. The expression of eight prognostic RBPs in tumor and normal tissues of CRC patients are shown in **Figure 3B**.

Constructing a Prognostic Risk Score Model in Colorectal Cancer

We constructed the optimum prognostic risk score model for prediction of OS of CRC patients by using multivariate Cox regression analysis. The identified eight prognostic RBPs were used to construct the prognostic risk score model. Among the eight prognostic RBPs, we identified UPF3B, SMAD6, NOL3, and PTRH1 as risk genes in the prognostic risk score model. Furthermore, coefficients of four risk genes are shown in **Figure 4A**. We calculated the risk scores using regression coefficient and expression levels of the risk genes according to this equation: risk score = 0.4257 * expression (UPF3B) + 0.5463 * expression (SMAD6) + 0.562 * expression (NOL3) + 2.368 * expression (PTRH1). Then, all CRC patients were divided into either high-risk or low-risk groups according to the risk scores. We demonstrated that patients in the high-risk group had a shorter OS time compared with those in the low-risk group (**Figure 4B**). We measured the prognostic ability of the risk score model through the use of an ROC analysis, which was conducted using the R package survivalROC. Also, our results indicated that the area under the ROC curve for 3- and 5-year OS was 0.645 and 0.672, respectively (**Figure 4C**). Heat map of mRNA expression indicated that all four were upregulated in the high-risk group (**Figure 4D**). Distribution of risk scores and survival status of patients are shown in **Figures 4E,F**. In addition, we found a higher percentage of deaths in the high-risk group.

We next assessed whether the four RBP-related gene models can predict the survival prognosis of CRC patients among additional CRC cohorts. We calculated the risk scores using the same formula that was used in the GSE29623 cohort. The results demonstrated that patients in the high-risk group had shorter OS time compared with those in the low-risk group (**Supplementary Figures 1A–D**), which is consistent with the TCGA CRC cohort. The results demonstrated that the four RBP-related gene models can accurately predict the prognosis of CRC patients.

Additionally, we explored the relationship between risk scores and clinical features of CRC patients. These results implied that risk scores were significantly higher in CRC patients that have deeper tumor infiltration (**Figure 4G**), distant metastasis

[3]https://cistrome.shinyapps.io/timer/

[4]http://www.proteinatlas.org

[5]http://www.cbioportal.org/

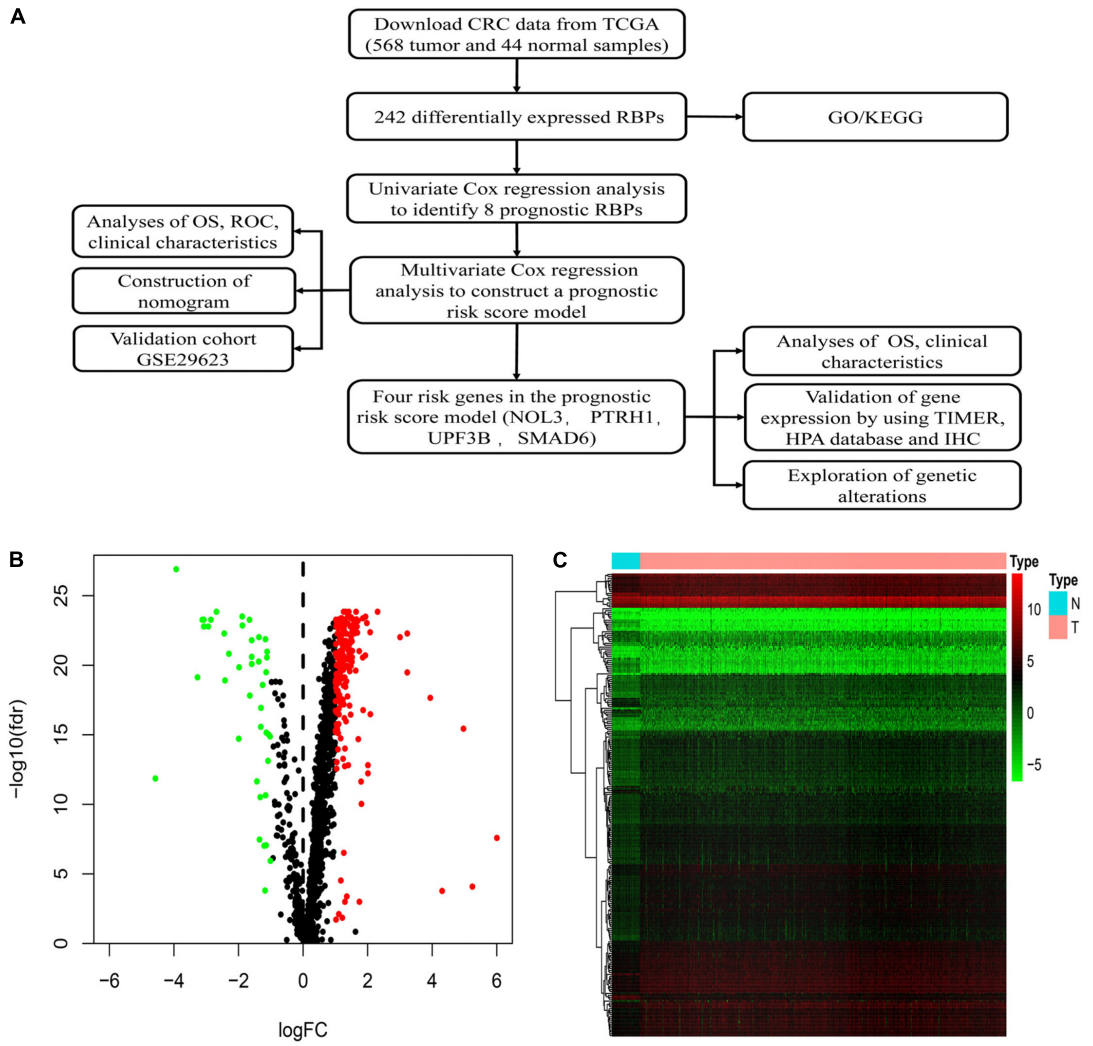


FIGURE 1 | Identification of differentially expressed RNA binding proteins (RBPs) in colorectal cancer (CRC). **(A)** Flowchart of our methods. **(B)** Volcano plots of differentially expressed RBPs in CRC. **(C)** Heatmap plots of differentially expressed RBPs in CRC.

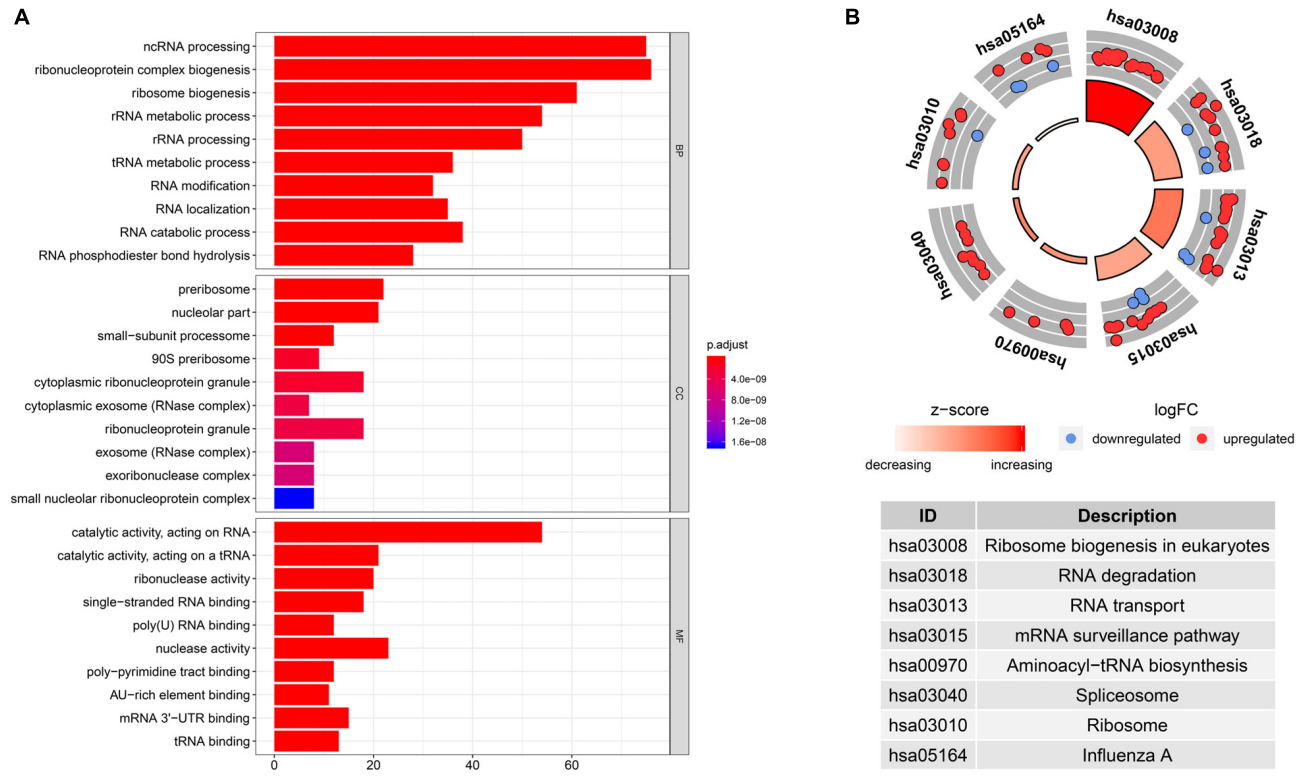
(**Figure 4H**), and lymph node metastasis (**Figure 4I**) and are at a late clinical stage (**Figure 4J**).

Building a Nomogram Based on the Prognostic Risk Score Model

Univariate and multivariate Cox regression analyses demonstrated that age, clinical stage, and the risk scores obtained from the prognostic risk score model were independent prognostic factors of CRC patients (**Figures 5A,B**). Therefore, based on these three independent prognostic factors, we constructed a nomogram that would be able to predict 3- and 5-year OS of CRC patients using the R package rms (**Figure 6A**). The calibration plots showed that the probability of predicting 3- or 5-year OS through nomogram agreed with actual observation (**Figures 6B,C**). Furthermore, the C-index for predicting OS through nomogram was 0.75 (95% CI: 0.69–0.81).

Validating Clinical Value of the Four Risk Genes in the Prognostic Risk Score Model

To further explore the clinical value of these four risk genes, we analyzed the relationship between expression of these genes and clinical features of the CRC patients. The survival curves showed that high expression of NOL3, PTRH1, SMAD6, and UPF3B is associated with poor prognosis of CRC patients (**Figures 7A–D**). We also found that NOL3 was significantly overexpressed in patients with lymph node metastasis (**Figure 7E**), distant metastasis (**Figure 7F**), and at a late clinical stage (**Figure 7G**). PTRH1 was found to be significantly overexpressed in patients with deeper tumor infiltration (**Figure 7H**). UPF3B was significantly upregulated in patients with deeper tumor infiltration (**Figure 7I**), as well as late clinical stage (**Figure 7J**). However, SMAD6 was downregulated in patients with deeper tumor infiltration (**Figure 7K**).



ID	Description
hsa03008	Ribosome biogenesis in eukaryotes
hsa03018	RNA degradation
hsa03013	RNA transport
hsa03015	mRNA surveillance pathway
hsa00970	Aminoacyl−tRNA biosynthesis
hsa03040	Spliceosome
hsa03010	Ribosome
hsa05164	Influenza A

FIGURE 2 | Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of differentially expressed RBPs in CRC. **(A)** GO terms of differentially expressed RBPs. **(B)** KEGG pathways of differentially expressed RBPs.

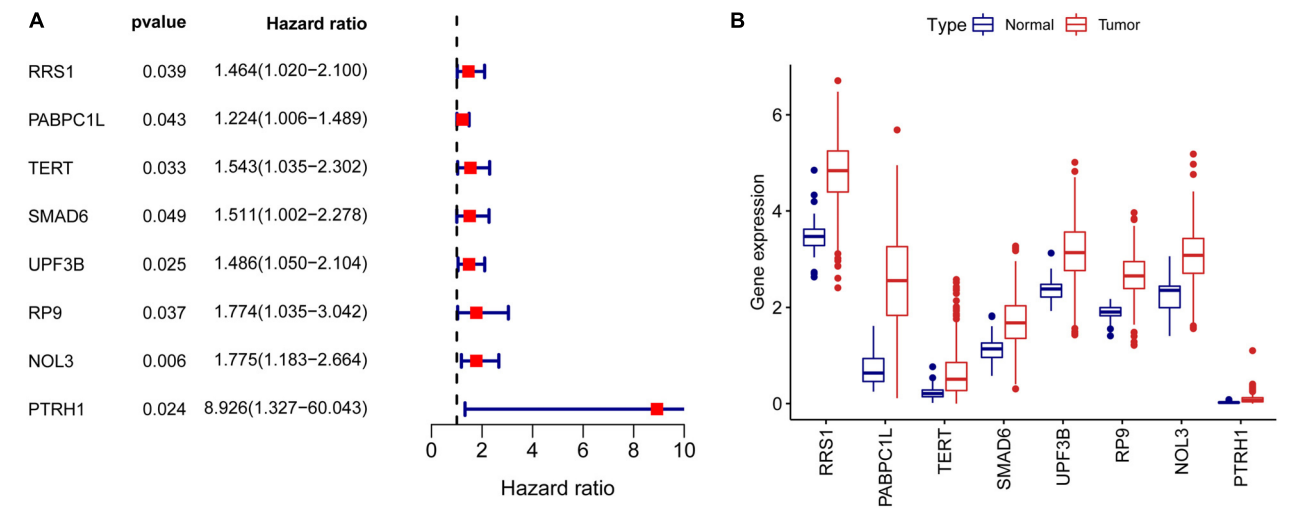


FIGURE 3 | Univariate Cox regression analysis to identify prognostic RBPs in CRC. **(A)** Forrest plot of univariate Cox regression analyses in CRC. **(B)** mRNA expression of eight prognostic RBPs in CRC.

External Validation of the Four Risk Genes

Consistent with results from the TCGA database, each of the four risk genes was found to be significantly upregulated in both colon and rectal cancer, according to the TIMER database (**Figure 8**). Interestingly, we identified that the expression of these four genes is not the same across different cancers. For instance, UPF3B is upregulated in esophageal carcinoma, stomach adenocarcinoma, hepatocellular carcinoma, and breast invasive carcinoma, whereas UPF3B was downregulated in kidney chromophobe, prostate adenocarcinoma, and thyroid carcinoma. Additionally,

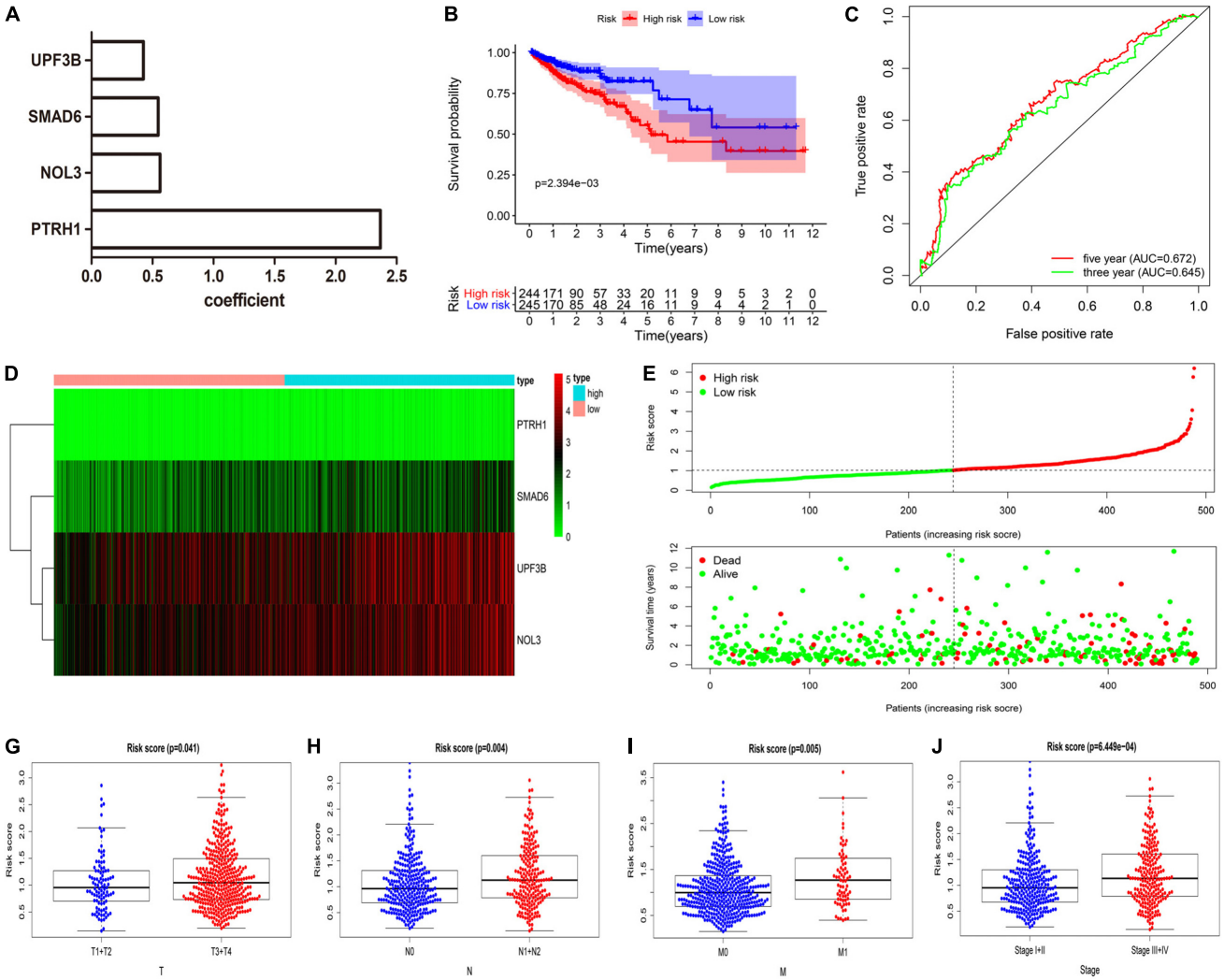


FIGURE 4 | Four RBP-related gene model to predict overall survival (OS) of CRC patients. **(A)** Coefficient of four risk genes identified by multivariate Cox regression analysis. **(B)** OS curve for high-risk and low-risk groups in the prognostic risk score model. **(C)** Time-dependent receiver operating characteristic (ROC) analysis of the prognostic risk score model.**(D–F)** Heat map of mRNA expression **(D)**, distribution of risk score **(E)**, and survival status **(F)** of patients in high-risk and low-risk groups. **(G–J)** Relationships between risk score and T **(G)**, N **(H)**, M **(I)**, and clinical stage **(J)**. (T, primary tumor; N, lymph node metastasis; M, distant metastasis).

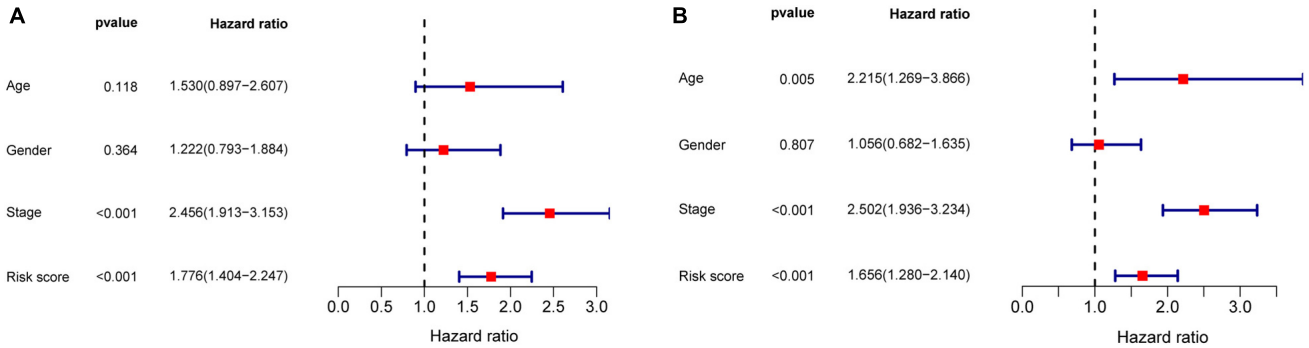


FIGURE 5 | Univariate and multivariate Cox regression analyses in CRC. **(A)** Forrest plot of the univariate Cox regression analyses in CRC. **(B)** Forrest plot of the multivariate Cox regression analyses in CRC.

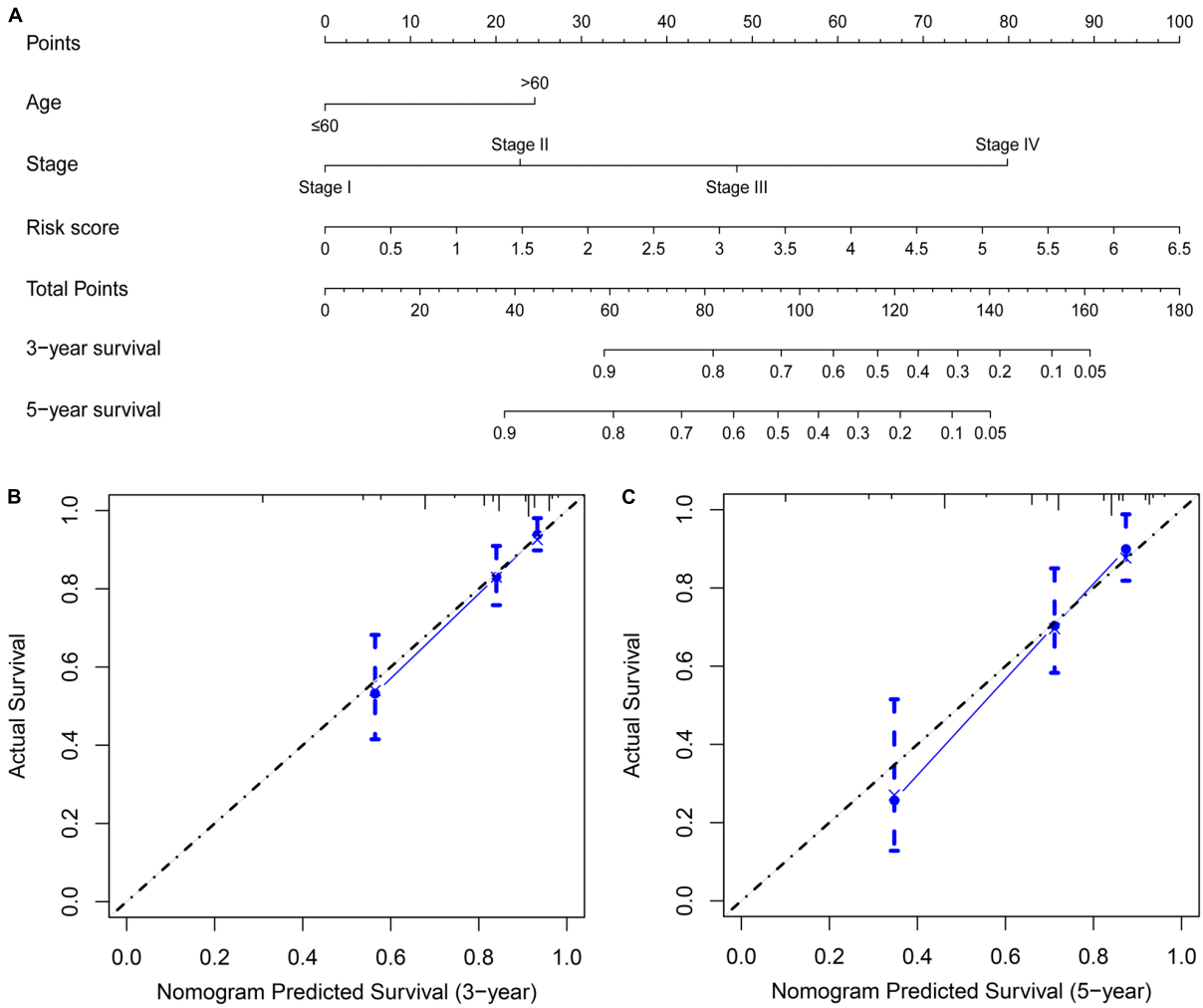


FIGURE 6 | Building a nomogram predicting OS of CRC patients. **(A)** CRC survival nomogram.The calibration plot for predicting patient survival at **(B)** 3 and **(C)** 5 years.

the expression of UPF3B was not found to be significantly different in pancreatic adenocarcinoma, kidney renal clear cell carcinoma, and kidney renal papillary cell carcinoma.

To further validate the expression of the four risk genes at the translational level, we analyzed the protein expression of these genes in the HPA database. The results indicated that NOL3 and UPF3B were overexpressed in CRC tumor tissues compared with normal tissues (**Figure 9A**). PTRH1 expression was not significantly different between the CRC tumor tissues and normal tissues (**Figure 9A**). However, information of SMAD6 levels were not found on the website. In addition, genetic alterations of the four risk genes were found to rarely occur (**Figure 9B**).

Validation of the Clinical Significance of NOL3 and UPF3B in Colorectal Cancer Patients by Immunohistochemistry

We obtained 44 pairs of CRC samples from Peking University People's Hospital to validate protein expression of the two key RBPs (NOL3 and UPF3B) in CRC. Immunohistochemical staining results displayed that NOL3 and UPF3B were upregulated in CRC tumor tissues (**Figures 10A–C**). In addition, our results showed that overexpression of NOL3 and UPF3B was associated with poor prognosis of CRC patients (**Figures 10D,E**). These results indicated that NOL3 and UPF3B play vital roles in predicting the prognosis of CRC patients.

DISCUSSION

Recently, numerous studies have focused on certain characteristics, such as autophagy and metabolic reprogramming, to identify gene signatures that are able to predict the mortality risk of cancer (Wan et al., 2019; Lin et al., 2020; Liu et al., 2020). In this study, we identified four RBP-related genes that were able to predict the OS of CRC patients. First, we detected 242 differentially expressed RBPs from a total of

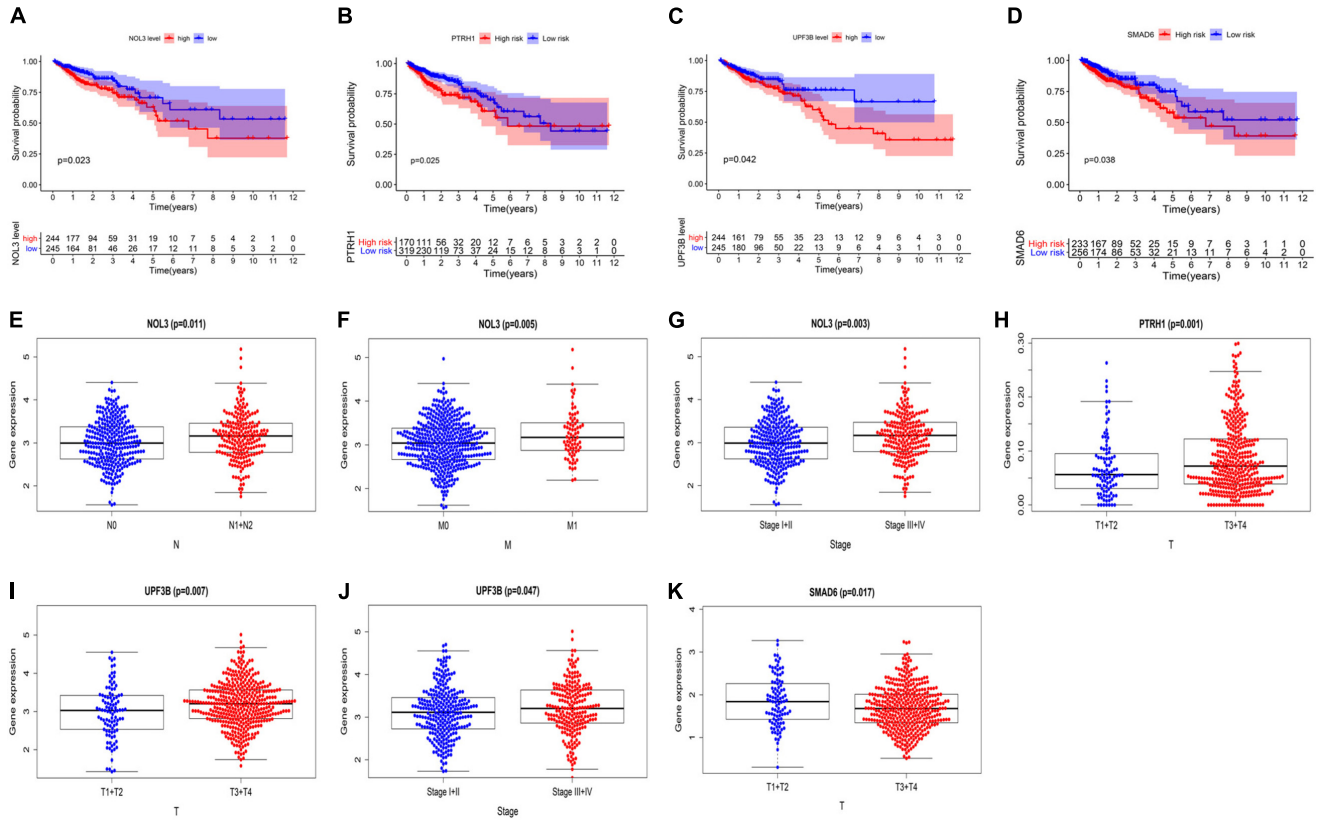


FIGURE 7 | Relationship between the four risk genes expression and clinical features of CRC patients in The Cancer Genome Atlas (TCGA). **(A–D)** Survival curves of NOL3 **(A)**, PTRH1 **(B)**, UPF3B **(C)**, and SMAD6 **(D)**. **(E–G)** Relationships between NOL3 expression and N **(E)**, M **(F)**, and clinical stage **(G)**. **(H)** Relationships between PTRH1 expression and T. **(I,J)** Relationships between UPF3B expression and T **(I)** and clinical stage **(J)**. **(K)** Relationships between SMAD6 expression and T. (T, primary tumor; N, lymph node metastasis; M, distant metastasis).

1,542 RBPs. Then, we performed univariate and multivariable Cox regression analyses and selected four prognosis-related RBPs to construct a prognostic risk score model. Also, results showed that the prognostic risk score model accurately predicted prognosis of CRC patients. Furthermore, we built a nomogram based on independent prognostic factors (including the risk score obtained from the prognostic risk score model, age, and clinical stage). Also, the nomogram performed well in predicting the 3- and 5-year OS. We also validated the clinical value of the four risk genes and found that they were associated with tumorigenesis, lymph node metastasis, distant metastasis, clinical stage, and OS. Finally, we confirmed the vital roles of NOL3 and UPF3B in predicting prognosis of CRC patients using IHC in a clinical cohort.

Based on four prognosis-related RBPs, we constructed a prognostic risk score model to predict OS of CRC patients. Some of these genes were found to be related to tumorigenesis and progression of CRC and other malignancies. NOL3 was strongly upregulated across multiple cancers. In particular, NOL3 was found to be highly expressed in AML and was associated with poor prognosis of AML patients (Carter et al., 2011, 2019; Mak et al., 2014a,b). Overexpression of NOL3 was also related to poor prognosis of nasopharyngeal carcinoma patients (Wu et al., 2013). Studies also found that NOL3 promoted tumorigenesis, metastasis, and chemoresistance in breast cancer, all of which contributed to worse patient prognosis (Medina-Ramirez et al., 2011). NOL3 was a direct target of miR-185 in gastric cancer (Li et al., 2014). Recent discoveries have also identified that upregulation of NOL3 was associated with worse prognosis among CRC patients (Mercier et al., 2008; Tóth et al., 2016), which is consistent with our results. Finally, NOL3 might be modulated by known cancer signaling proteins including Ras (Wu et al., 2010) and HIF-1 (Ao et al., 2012), and the lncRNA PCAT6 (Huang et al., 2019), in CRC.

SMAD6, a member of the SMAD family, negatively modulates the transforming growth factor-β signaling pathway (Jung et al., 2013). SMAD6 is predictive of patient survival in oral squamous cell carcinoma (Mangone et al., 2010). SMAD6 was found to be overexpressed in glioma, and its overexpression is associated with poor patient survival (Jiao et al., 2018). SMAD6 correlated with poor patient survival among non-small cell lung cancer, and its knockdown inhibited cell proliferation and increased apoptosis in the lung cancer cell line (Jeon et al., 2008). Our study indicated

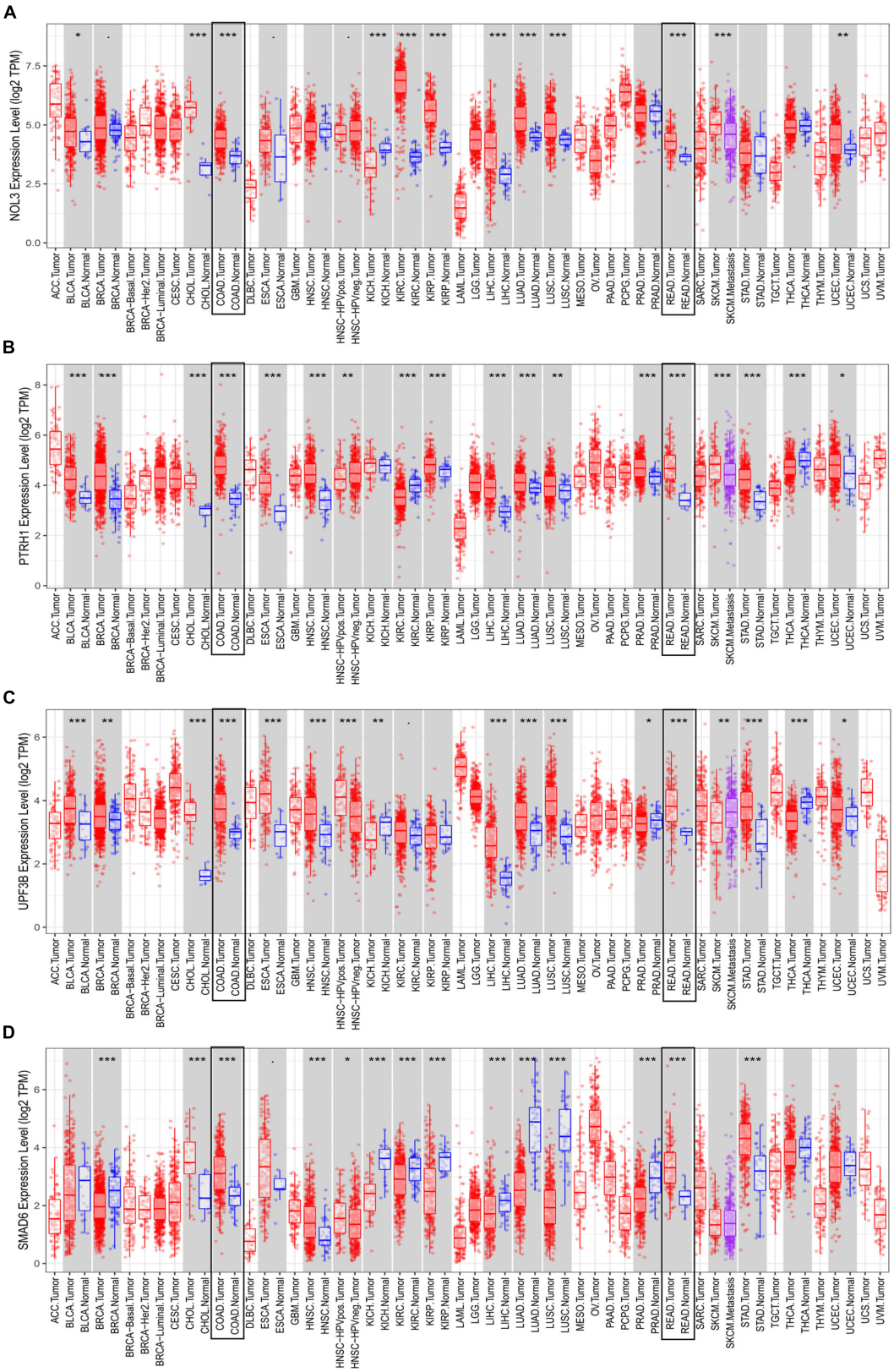


FIGURE 8 | Expression of the four risk genes in multiple cancers. Expression of **(A)** NOL3, **(B)** PTRH1, **(C)** UPF3B, and **(D)** SMAD6 in TIMER. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

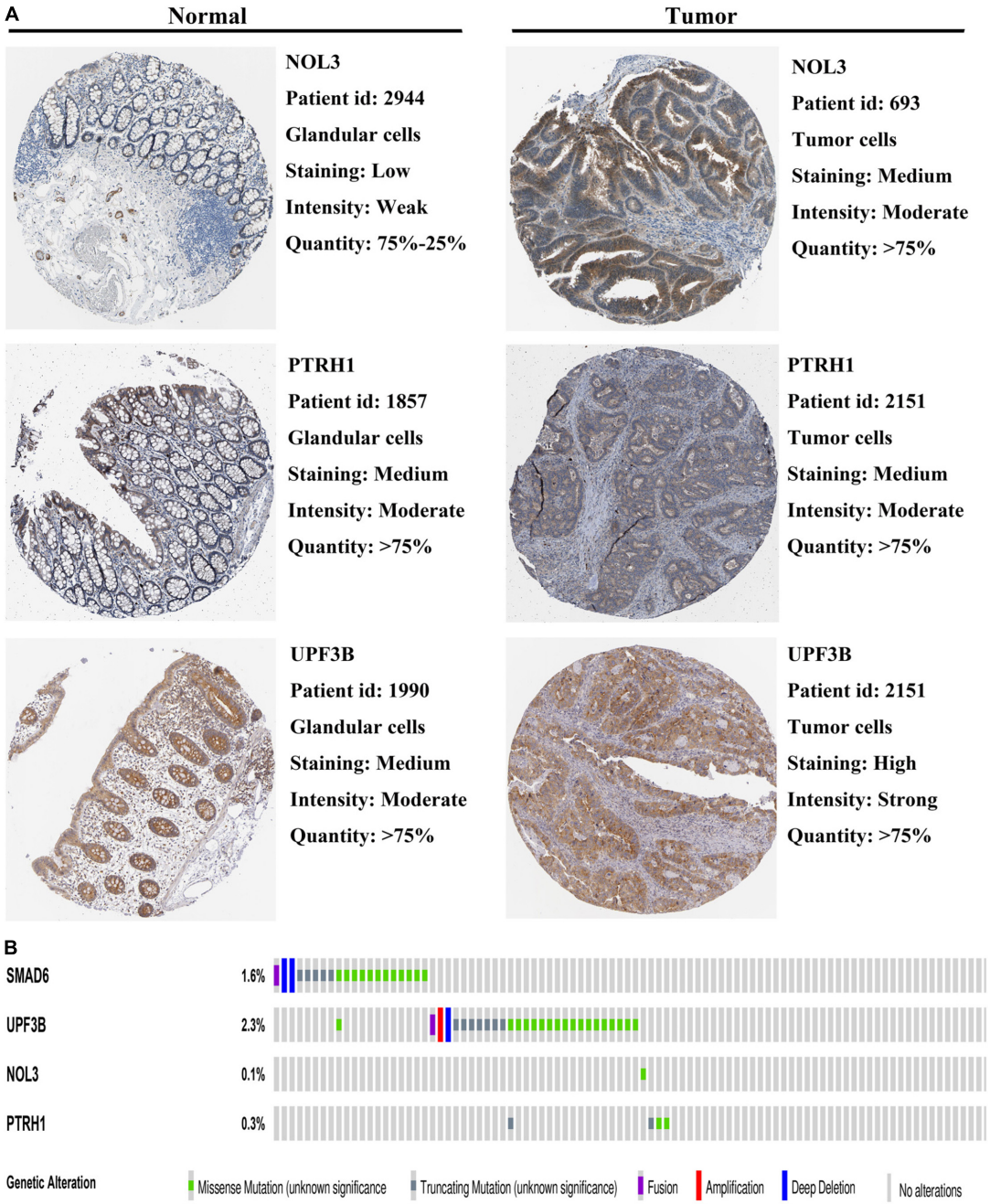


FIGURE 9 | Protein expression and genetic alterations of the four risk genes. **(A)** Representative protein level of the four risk genes in CRC tumor and normal tissues in the Human Protein Atlas database. Data of SMAD6 are not available in this database. **(B)** Genetic alterations of the four risk genes in CRC in the cBioportal for Cancer Genomics.

that overexpression of SMAD6 is related to worse patient prognosis in CRC.

Our study identified that upregulated PTRH1 and UPF3B correlated with worse prognosis of CRC patients. However, we only found a few studies about these two genes in CRC and other cancers. UPF3B, a member of the UP-frameshift proteins, mediates nonsense-mediated mRNA decay (Raimondeau et al., 2018). UP-frameshift proteins include UPF1, UPF2, UPF3A, and UPF3B (Raimondeau et al., 2018). Although UPF3B is less well-studied, we found several studies about additional UP-frameshift proteins in cancers. UPF1 regulates tumor progression via diverse mechanisms across different kinds of cancers, including CRC (Bordonaro and Lazarova, 2019), hepatocellular carcinoma (Chang et al., 2016; Zhang et al., 2017), pancreatic adenosquamous carcinoma (Liu et al., 2014), glioblastoma (Shao et al., 2019), and

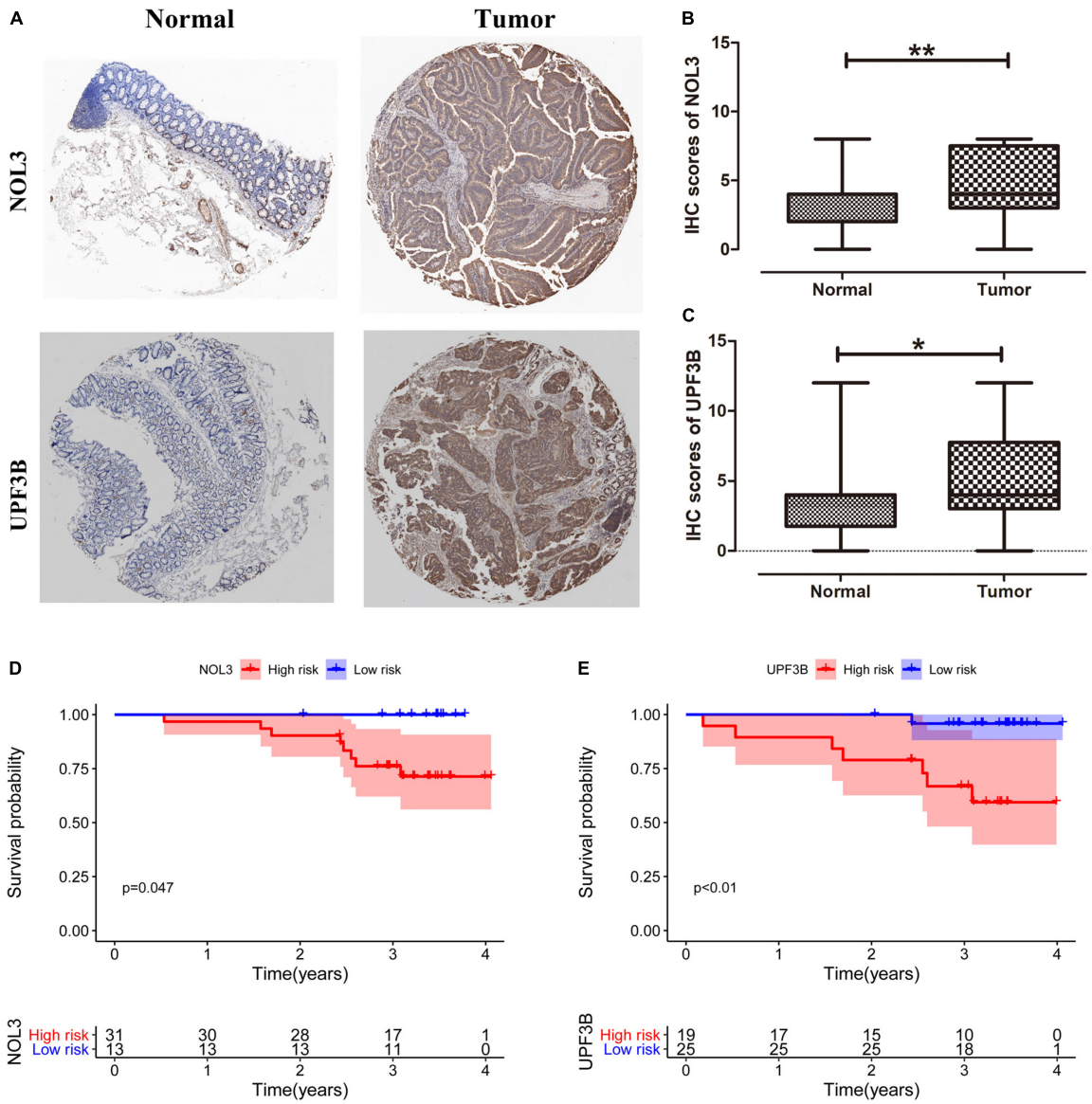


FIGURE 10 | Clinical significance of NOL3 and UPF3B in CRC patients by immunohistochemistry (IHC). **(A)** IHC staining for NOL3 and UPF3B in tumor and normal tissues of CRC patients. **(B,C)** Statistical analysis of NOL3 and UPF3B in tumor and normal tissues of CRC patients, respectively. **(D,E)** Survival curves of NOL3 and UPF3B. *$p < 0.05$, **$p < 0.01$.

endometrial carcinoma (Xing et al., 2020). UPF3A partially contributes to the effect of calcium homeostasis endoplasmic reticulum protein (CHERP) in promoting tumorigenesis in CRC (Wang Q. et al., 2019). UPF3B and other UP-frameshift proteins can be interacted (Raimondeau et al., 2018). These data, combined with results from our study, suggest that UPF3B regulates tumor progression of CRC and may represent a potential prognostic biomarker for CRC patients. However, the mechanism of its effect in CRC requires further study.

Although our study indicates that RBPs prominently contribute to the prognosis of CRC patients, several limitations need to be pointed out. First, the clinical cohort contains fewer patients, which may lead to deviation. Additionally, the mechanisms of how these RBPs regulate the progression of CRC require further exploration.

CONCLUSION

In conclusion, we performed a comprehensive bioinformatics analysis of RBPs and identified several potential prognostic RBPs in CRC. The prognostic risk score model, including four RBPs, is an independent prognostic factor for CRC. These four RBPs are involved in tumorigenesis, progression, and prognosis of CRC. RBPs represent an alternative strategy to interfere

with tumor progression and predict the prognosis of CRC patients in the future.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Peking University People's Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ZZ, LW, SW, and ZS designed the study. ZZ, LW, QW, MZ, and BW analyzed the data. ZZ and LW wrote the manuscript. KJ, YY, SW, and ZS reviewed and edited the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Key Research and Development Program of China (2017YFC1308800) and the National Natural Science Foundation of China (Grant Nos. 81672375, 81972240, and 81871962).

ACKNOWLEDGMENTS

We thank our colleagues for their helpful suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.580149/full#supplementary-material

Supplementary Figure 1 | Test of four RBPs-related gene model in the GSE29623 cohort. **(A)** Overall survival curve for high-risk and low-risk groups in the GSE29623 cohort. **(B–D)** Heat map of mRNA expression **(B)**, distribution of risk score **(C)**, and survival status **(D)** of patients in high-risk and low-risk groups in the GSE29623 cohort, respectively.

REFERENCES

Ao, J. E., Kuang, L. H., Zhou, Y., Zhao, R., and Yang, C. M. (2012). Hypoxia-inducible factor 1 regulated ARC expression mediated hypoxia induced inactivation of the intrinsic death pathway in p53 deficient human colon cancer cells. *Biochem. Biophys. Res., Commun.* 420, 913–917. doi: 10.1016/j.bbrc.2012.03.101

Bordonaro, M., and Lazarova, D. (2019). Amlexanox and UPF1 modulate Wnt signaling and apoptosis in HCT-116 colorectal cancer cells. *J. Cancer* 10, 287–292. doi: 10.7150/jca.28331

Carter, B. Z., Mak, P. Y., Wang, X., Tao, W., Ruvolo, V., Mak, D., et al. (2019). An ARC-Regulated IL1β/Cox-2/PGE2/β-Catenin/ARC circuit controls leukemia-microenvironment interactions and confers drug resistance in AML. *Cancer Res.* 79, 1165–1177. doi: 10.1158/0008-5472.Can-18-0921

Carter, B. Z., Qiu, Y. H., Zhang, N., Coombes, K. R., Mak, D. H., Thomas, D. A., et al. (2011). Expression of ARC (apoptosis repressor with caspase recruitment domain), an antiapoptotic protein, is strongly prognostic in AML. *Blood* 117, 780–787. doi: 10.1182/blood-2010-04-280503

Chang, L., Li, C., Guo, T., Wang, H., Ma, W., Yuan, Y., et al. (2016). The human RNA surveillance factor UPF1 regulates tumorigenesis by targeting Smad7 in hepatocellular carcinoma. *J. Exp. Clin. Cancer Res.* 35:8. doi: 10.1186/s13046-016-0286-2

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. *Nat. Rev. Genet.* 15, 829–845. doi: 10.1038/nrg3813

Hentze, M. W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* 19, 327–341. doi: 10.1038/nrm.2017.130

Huang, W., Su, G., Huang, X., Zou, A., Wu, J., Yang, Y., et al. (2019). Long noncoding RNA PCAT6 inhibits colon cancer cell apoptosis by regulating anti-apoptotic protein ARC expression via EZH2. *Cell Cycle* 18, 69–83. doi: 10.1080/15384101.2018.1558872

Jeon, H. S., Dracheva, T., Yang, S. H., Meerzaman, D., Fukuoka, J., Shakoori, A., et al. (2008). SMAD6 contributes to patient survival in non-small cell lung cancer and its knockdown reestablishes TGF-beta homeostasis in lung cancer cells. *Cancer Res.* 68, 9686–9692. doi: 10.1158/0008-5472.Can-08-1083

Jiao, J., Zhang, R., Li, Z., Yin, Y., Fang, X., Ding, X., et al. (2018). Nuclear Smad6 promotes gliomagenesis by negatively regulating PIAS3-mediated STAT3 inhibition. *Nat. Commun.* 9:2504. doi: 10.1038/s41467-018-04936-9

Jung, S. M., Lee, J. H., Park, J., Oh, Y. S., Lee, S. K., Park, J. S., et al. (2013). Smad6 inhibits non-canonical TGF-β1 signalling by recruiting the deubiquitinase A20 to TRAF6. *Nat. Commun.* 4:2562. doi: 10.1038/ncomms3562

Kow, A. W. C. (2019). Hepatic metastasis from colorectal cancer. *J. Gastrointest. Oncol.* 10, 1274–1298. doi: 10.21037/jgo.2019.08.06

Li, Q., Wang, J. X., He, Y. Q., Feng, C., Zhang, X. J., Sheng, J. Q., et al. (2014). MicroRNA-185 regulates chemotherapeutic sensitivity in gastric cancer by targeting apoptosis repressor with caspase recruitment domain. *Cell Death Dis.* 5:e1197. doi: 10.1038/cddis.2014.148

Li, W., Gao, L. N., Song, P. P., and You, C. G. (2020). Development and validation of a RNA binding protein-associated prognostic model for lung adenocarcinoma. *Aging* 12, 3558–3573. doi: 10.18632/aging.102828

Li, Y., McGrail, D. J., Xu, J., Li, J., Liu, N. N., Sun, M., et al. (2019a). MERIT: systematic analysis and characterization of mutational effect on RNA interactome topology. *Hepatology* 70, 532–546. doi: 10.1002/hep.30242

Li, Y., Xiao, J., Bai, J., Tian, Y., Qu, Y., Chen, X., et al. (2019b). Molecular characterization and clinical relevance of m(6)A regulators across 33 cancer types. *Mol. Cancer* 18:137. doi: 10.1186/s12943-019-1066-3

Lin, Q. G., Liu, W., Mo, Y. Z., Han, J., Guo, Z. X., Zheng, W., et al. (2020). Development of prognostic index based on autophagy-related genes analysis in breast cancer. *Aging* 12, 1366–1376. doi: 10.18632/aging.102687

Liu, C., Karam, R., Zhou, Y., Su, F., Ji, Y., Li, G., et al. (2014). The UPF1 RNA surveillance gene is commonly mutated in pancreatic adenosquamous carcinoma. *Nat. Med.* 20, 596–598. doi: 10.1038/nm.3548

Liu, G. M., Xie, W. X., Zhang, C. Y., and Xu, J. W. (2020). Identification of a four-gene metabolic signature predicting overall survival for hepatocellular carcinoma. *J. Cell Physiol.* 235, 1624–1636. doi: 10.1002/jcp.29081

Mak, P. Y., Mak, D. H., Mu, H., Shi, Y., Ruvolo, P., Ruvolo, V., et al. (2014a). Apoptosis repressor with caspase recruitment domain is regulated by MAPK/PI3K and confers drug resistance and survival advantage to AML. *Apoptosis* 19, 698–707. doi: 10.1007/s10495-013-0954-z

Mak, P. Y., Mak, D. H., Ruvolo, V., Jacamo, R., Kornblau, S. M., Kantarjian, H., et al. (2014b). Apoptosis repressor with caspase recruitment domain modulates second mitochondrial-derived activator of caspases mimetic-induced cell death through BIRC2/MAP3K14 signalling in acute myeloid leukaemia. *Br. J. Haematol.* 167, 376–384. doi: 10.1111/bjh.13054

Mangone, F. R., Walder, F., Maistro, S., Pasini, F. S., Lehn, C. N., Carvalho, M. B., et al. (2010). Smad2 and Smad6 as predictors of overall survival in oral

squamous cell carcinoma patients. *Mol. Cancer* 9:106. doi: 10.1186/1476-4598-9-106

Medina-Ramirez, C. M., Goswami, S., Smirnova, T., Bamira, D., Benson, B., Ferrick, N., et al. (2011). Apoptosis inhibitor ARC promotes breast tumorigenesis, metastasis, and chemoresistance. *Cancer Res.* 71, 7705–7715. doi: 10.1158/0008-5472.Can-11-2192

Mercier, I., Vuolo, M., Jasmin, J. F., Medina, C. M., Williams, M., Mariadason, J. M., et al. (2008). ARC (apoptosis repressor with caspase recruitment domain) is a novel marker of human colon cancer. *Cell Cycle* 7, 1640–1647. doi: 10.4161/cc.7.11.5979

Mitchell, S. F., and Parker, R. (2014). Principles and properties of eukaryotic mRNPs. *Mol. Cell* 54, 547–558. doi: 10.1016/j.molcel.2014.04.033

Mongroo, P. S., Noubissi, F. K., Cuatrecasas, M., Kalabis, J., King, C. E., Johnstone, C. N., et al. (2011). IMP-1 displays cross-talk with K-Ras and modulates colon cancer cell survival through the novel proapoptotic protein CYFIP2. *Cancer Res.* 71, 2172–2182. doi: 10.1158/0008-5472.Can-10-3295

Moore, M. J. (2005). From birth to death: the complex lives of eukaryotic mRNAs. *Science* 309, 1514–1518. doi: 10.1126/science.1111443

Neelamraju, Y., Hashemikhabir, S., and Janga, S. C. (2015). The human RBPome: from genes and proteins to human disease. *J. Proteomics* 127(Pt A), 61–70. doi: 10.1016/j.jprot.2015.04.031

Pérez-Guijarro, E., Karras, P., Cifdaloz, M., Martínez-Herranz, R., Cañón, E., Graña, O., et al. (2016). Lineage-specific roles of the cytoplasmic polyadenylation factor CPEB4 in the regulation of melanoma drivers. *Nat. Commun.* 7:13418. doi: 10.1038/ncomms13418

Perron, G., Jandaghi, P., Solanki, S., Safisamghabadi, M., Storoz, C., Karimzadeh, M., et al. (2018). A general framework for interrogation of mRNA stability programs identifies RNA-binding proteins that govern cancer transcriptomes. *Cell Rep.* 23, 1639–1650. doi: 10.1016/j.celrep.2018.04.031

Raimondeau, E., Bufton, J. C., and Schaffitzel, C. (2018). New insights into the interplay between the translation machinery and nonsense-mediated mRNA decay factors. *Biochem. Soc. Trans.* 46, 503–512. doi: 10.1042/bst20170427

Roth, A. D., Tejpar, S., Delorenzi, M., Yan, P., Fiocca, R., Klingbiel, D., et al. (2010). Prognostic role of KRAS and BRAF in stage II and III resected colon cancer: results of the translational study on the PETACC-3, EORTC 40993, SAKK 60-00 trial. *J. Clin. Oncol.* 28, 466–474. doi: 10.1200/jco.2009.23.3452

Saha, S., Murmu, K. C., Biswas, M., Chakraborty, S., Basu, J., Madhulika, S., et al. (2019). Transcriptomic analysis identifies RNA binding proteins as putative regulators of myelopoiesis and leukemia. *Front. Oncol.* 9:692. doi: 10.3389/fonc.2019.00692

Shao, L., He, Q., Liu, Y., Liu, X., Zheng, J., Ma, J., et al. (2019). UPF1 regulates the malignant biological behaviors of glioblastoma cells via enhancing the stability of Linc-00313. *Cell Death Dis.* 10:629. doi: 10.1038/s41419-019-1845-1

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. *CA Cancer J. Clin.* 70, 7–30. doi: 10.3322/caac.21590

Tóth, C., Meinrath, J., Herpel, E., Derix, J., Fries, J., Buettner, R., et al. (2016). Expression of the apoptosis repressor with caspase recruitment domain (ARC) in liver metastasis of colorectal cancer and its correlation with DNA mismatch repair proteins and p53. *J. Cancer Res. Clin. Oncol.* 142, 927–935. doi: 10.1007/s00432-015-2102-3

Van Cutsem, E., Cervantes, A., Nordlinger, B., and Arnold, D. (2014). Metastatic colorectal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* 25(Suppl. 3), iii1–iii9. doi: 10.1093/annonc/mdu260

Vázquez-Chantada, M., Fernández-Ramos, D., Embade, N., Martínez-Lopez, N., Varela-Rey, M., Woodhoo, A., et al. (2010). HuR/methyl-HuR and AUF1 regulate the MAT expressed during liver proliferation, differentiation, and carcinogenesis. *Gastroenterology* 138, 1943–1953. doi: 10.1053/j.gastro.2010.01.032

Wan, B., Liu, B., Yu, G., Huang, Y., and Lv, C. (2019). Differentially expressed autophagy-related genes are potential prognostic and diagnostic biomarkers in clear-cell renal cell carcinoma. *Aging* 11, 9025–9042. doi: 10.18632/aging.102368

Wang, K., Li, L., Fu, L., Yuan, Y., Dai, H., Zhu, T., et al. (2019). Integrated bioinformatics analysis the function of RNA binding proteins (RBPs) and their prognostic value in breast cancer. *Front, Pharmacol.* 10:140. doi: 10.3389/fphar.2019.00140

Wang, Q., Wang, Y., Liu, Y., Zhang, C., Luo, Y., Guo, R., et al. (2019). U2-related proteins CHERP and SR140 contribute to colorectal tumorigenesis via alternative splicing regulation. *Int. J. Cancer* 145, 2728–2739. doi: 10.1002/ijc.32331

Wang, Z., Tang, W., Yuan, J., Qiang, B., Han, W., and Peng, X. (2020). Integrated analysis of RNA-Binding proteins in glioma. *Cancers* 12:892. doi: 10.3390/cancers12040892

Wu, L., Nam, Y. J., Kung, G., Crow, M. T., and Kitsis, R. N. (2010). Induction of the apoptosis inhibitor ARC by Ras in human cancers. *J. Biol. Chem.* 285, 19235–19245. doi: 10.1074/jbc.M110.114892

Wu, P., Tang, Y., He, J., Qi, L., Jiang, W., and Zhao, S. (2013). ARC is highly expressed in nasopharyngeal carcinoma and confers X-radiation and cisplatin resistance. *Oncol. Rep.* 30, 1807–1813. doi: 10.3892/or.2013.2622

Xing, T. R., Chen, P., Wu, J. M., Gao, L. L., Yang, W., Cheng, Y., et al. (2020). UPF1 participates in the progression of endometrial cancer by inhibiting the expression of lncRNA PVT1. *Onco Targets Ther.* 13, 2103–2114. doi: 10.2147/ott.S233149

Xu, Y., Gao, X. D., Lee, J. H., Huang, H., Tan, H., Ahn, J., et al. (2014). Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. *Genes Dev.* 28, 1191–1203. doi: 10.1101/gad.241968.114

Zhang, H., You, Y., and Zhu, Z. (2017). The human RNA surveillance factor Up-frameshift 1 inhibits hepatic cancer progression by targeting MRP2/ABCC2. *Biomed. Pharmacother.* 92, 365–372. doi: 10.1016/j.biopha.2017.05.090

Zhang, Z., Zheng, X., Li, J., Duan, J., Cui, L., Yang, L., et al. (2019). Overexpression of UBR5 promotes tumor growth in gallbladder cancer via PTEN/PI3K/Akt signal pathway. *J. Cell Biochem.* 120, 11517–11524. doi: 10.1002/jcb.28431

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 07 January 2021
doi: 10.3389/fgene.2020.617872

Transcriptome Analyses Identify an RNA Binding Protein Related Prognostic Model for Clear Cell Renal Cell Carcinoma

*Yue Wu[1,2], Xian Wei[1,2], Huan Feng[1,2], Bintao Hu[1,2], Bo Liu[3], Yang Luan[1,2], Yajun Ruan[1,2], Xiaming Liu[1,2], Zhuo Liu[1,2], Shaogang Wang[1,2], Jihong Liu[1,2] and Tao Wang[1,2]**

[1] Department of Urology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2] Institute of Urology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [3] Department of Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China*

RNA binding proteins (RBPs) play a key role in post-transcriptional gene regulation. They have been shown to be dysfunctional in a variety of cancers and are closely related to the occurrence and progression of cancers. However, the biological function and clinical significance of RBPs in clear cell renal carcinoma (ccRCC) are unclear. In our current study, we downloaded the transcriptome data of ccRCC patients from The Cancer Genome Atlas (TCGA) database and identified differential expression of RBPs between tumor tissue and normal kidney tissue. Then the biological function and clinical value of these RBPs were explored by using a variety of bioinformatics techniques. We identified a total of 40 differentially expressed RBPs, including 10 down-regulated RBPs and 30 up-regulated RBPs. Eight RBPs (*APOBEC3G, AUH, DAZL, EIF4A1, IGF2BP3, NR0B1, RPL36A,* and *TRMT1*) and nine RBPs (*APOBEC3G, AUH, DDX47, IGF2BP3, MOV10L1, NANOS1, PIH1D3, TDRD9*, and *TRMT1*) were identified as prognostic related to overall survival (OS) and disease-free survival (DFS), respectively, and prognostic models for OS and DFS were constructed based on these RBPs. Further analysis showed that OS and DFS were worse in high-risk group than in the low-risk group. The area under the receiver operator characteristic curve of the model for OS was 0.702 at 3 years and 0.726 at 5 years in TCGA cohort and 0.783 at 3 years and 0.795 at 5 years in E-MTAB-1980 cohort, showing good predictive performance. Both models have been shown to independently predict the prognosis of ccRCC patients. We also established a nomogram based on these prognostic RBPs for OS and performed internal validation in the TCGA cohort, showing an accurate prediction of ccRCC prognosis. Stratified analysis showed a significant correlation between the prognostic model for OS and ccRCC progression.

Keywords: clear cell renal cell carcinoma, RNA binding proteins, prognostic model, survival analysis, bioinformatics

OPEN ACCESS

Edited by:
Peter G. Zaphiropoulos, Karolinska Institutet (KI), Sweden

Reviewed by:
Cheng Wenjun, Nanjing Medical University, China
Guodong Yang, Yangtze University, China

***Correspondence:**
Tao Wang
twang@hust.edu.cn;
tjhwt@126.com

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *15 October 2020*
Accepted: *07 December 2020*
Published: *07 January 2021*

Citation:
Wu Y, Wei X, Feng H, Hu B, Liu B, Luan Y, Ruan Y, Liu X, Liu Z, Wang S, Liu J and Wang T (2021) Transcriptome Analyses Identify an RNA Binding Protein Related Prognostic Model for Clear Cell Renal Cell Carcinoma. Front. Genet. 11:617872. doi: 10.3389/fgene.2020.617872

Abbreviations: RCC, renal cell carcinoma; ccRCC, clear cell renal cell carcinoma; TCGA, The Cancer Genome Atlas; RBPs, RNA binding proteins; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; FC, fold change; OS, overall survival; DFS, disease-free survival; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic; AUC, area under the receiver operating characteristic curve; FDR, false discovery rate.

INTRODUCTION

Renal cell carcinoma (RCC) accounts for 2.4% of all malignancies, with an estimated 400,000 new cases and 175,000 deaths worldwide each year (Bray et al., 2018; Siegel et al., 2018). the clear cell renal cell carcinoma (ccRCC) is the most common subtype of RCC, accounting for approximately 70–80% and presents a high risk of heterogeneity and metastasis (Rini et al., 2009; Ljungberg et al., 2019). Although surgical resection can effectively resolve the early stage of ccRCC, 30% of patients still have recurrence or metastasis after surgery, and the late stage of ccRCC has a high mortality rate due to insensitivity to traditional radiotherapy or chemotherapy (Battaglia and Lucarelli, 2015; Moch et al., 2016; Tamma et al., 2019). Therefore, further understanding of the molecular mechanisms of ccRCC and the discovery of more effective molecular biomarkers are essential for early screening, diagnosis, monitoring for metastasis, recurrence, and quality of life in patients.

Post-transcriptional regulation of RNA is an important aspect of gene expression regulation. RNA binding proteins (RBPs) are a class of proteins widely expressed in cells, which form ribonucleoprotein (RNP) complexes through binding at different sites or random interaction with target RNA, thus strictly regulating RNA metabolism (Iadevaia and Gerber, 2015; Hentze et al., 2018). Currently, there are 1,542 RBP coding genes, accounting for 7.5% of all human protein-coding genes, which have been verified by experiments (Gerstberger et al., 2014). These RBPs regulate a variety of biological processes including RNA processing, splicing, mRNA stability, output, localization, and translation, thus maintaining the physiological balance of the cell (Masuda and Kuwano, 2019). Given this, it comes as no surprise that RBPs dysfunction has been linked to a variety of human diseases. Ribosomal diseases caused by ribosomal protein and rRNA biogenic factor defects, such as Diamond–Blackfan anemia and Shwachman–Diamond syndrome, affect the same tissues and exhibit similar pathology precisely because RBPs bind to the same type of RNA (Narla and Ebert, 2010). Mutations in mRBPs or their targets in neurons lead to abnormal aggregation of proteins or RNA, resulting in a variety of neurodegenerative and neuromuscular diseases (Scheper et al., 2007). However, the role of RBPs in tumor genesis and development is rare.

Some studies have shown that RBPs are abnormally expressed in tumor tissues compared with normal tissues and are associated with patient prognosis (Patry et al., 2003; Busà et al., 2007; Ortiz-Zapater et al., 2011). In lung cancer, QKI inhibits tumor cell proliferation by competing with the splicing factor SF1 (Zong et al., 2014). In melanoma, CPEB4 promotes tumor cell proliferation by regulating polyadenylation and promoting the translation of melanoma drivers (Pérez-Guijarro et al., 2016). Knockdown SAM68 in breast cancer cells inhibited tumor cell proliferation by upregulation of cell cycle inhibitors P21 and CDKN1B/P27 (Song et al., 2010). However, in the field of ccRCC, existing studies only described the effect of RBPs on the overall survival (OS) of ccRCC patients (Hua et al., 2020; Zhu et al., 2020), and few RBPs models can be used to predict the prognosis of ccRCC patients. The development of new RBPs models has gradually become an effective method to explore new therapeutic targets. Therefore, in our current study, we systematically and deeply analyzed the molecular biological function and clinical significance of RBPs in ccRCC to promote our understanding of ccRCC progress, and established risk score models for OS and disease-free survival (DFS), which may provide new biomarkers for disease diagnosis and treatment prognosis.

MATERIALS AND METHODS

Preprocessing Data and Identifying Differential Expression RBPs

Transcriptome data of 72 normal renal tissue specimens and 539 ccRCC specimens were downloaded from The Cancer Genome Atlas database (TCGA[1]). We then used the edgeR package[2] to preprocess the raw data and identify the differentially expressed RBPs based on $|\log_2$ fold change (FC)$| > 1.0$ and false discovery rate (FDR) < 0.05. We also downloaded the E-MTAB-1980 dataset from the ArrayExpress database[3] and downloaded the transcriptome data of 436 ccRCC patients containing DFS information from the cBioportal database[4].

Function and Pathway Enrichment Analysis

We used the WEB-based Gene Set Analysis Toolkit (WebGestalt[5]) online analysis tool to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of these differentially expressed RBPs (Liao et al., 2019). The GO terms including biological process, cellular component, and molecular function. All analysis results were screened according to the criteria of $P < 0.05$ and gene number > 5.

Selection of Prognostic Related RBPs

To identify RBPs with important prognostic significance, we first performed univariate Cox regression analysis of all these differentially expressed RBPs. The least absolute shrinkage and selection operator (LASSO) regression analysis was then used for further screening. Finally, multivariate Cox regression analysis was used to further screen out RBPs with important prognostic value. A $P < 0.05$ was considered significant.

Construction and Evaluation of Prognostic Model for OS

We constructed a multivariate Cox proportional hazards regression model to predict the prognosis of ccRCC patients based on these prognostic related RBPs. The risk score for each patient in the model was calculated using the following formula:

$$Risk\ score = \sum_{i=1}^{n} Expi\beta i,$$

[1]https://portal.gdc.cancer.gov/

[2]http://www.bioconductor.org/packages/release/bioc/html/edgeR.html

[3]https://www.ebi.ac.uk/arrayexpress/

[4]https://www.cbioportal.org/datasets

[5]http://www.webgestalt.org/

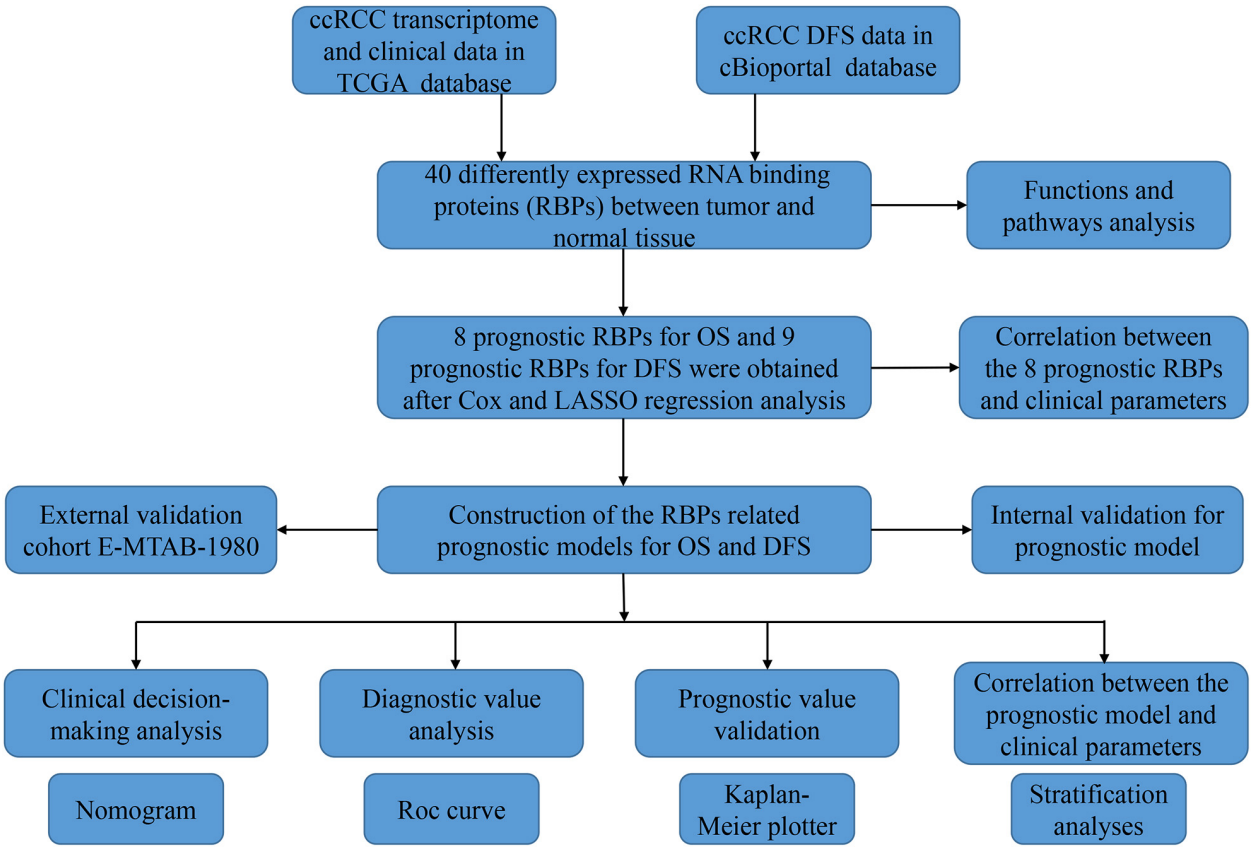


FIGURE 1 | The research diagram for analyzing RBPs in ccRCC.

In this formula, *Exp* represents the expression value of each gene, and β represents the corresponding regression coefficient. We then divided ccRCC patients from the TCGA cohort into low-risk and high-risk subgroups based on the median risk score, and compared OS between the two groups to initially assess the predictive power of the model. In addition, we used the Survival ROC R package to establish the ROC curve to assess the prognostic efficacy of the model and used the rms R package to draw the nomogram to predict OS. Finally, we divided the 539 samples in the TCGA cohort into the training group and the validation group as internal validation and the E-MTAB-1980 cohort with 101 sample information as external validation to evaluate the stability and predictive efficacy of the model.

Correlation Between Prognostic Model for OS, Prognostic RBPs and Clinical Parameters

To explore the clinical significance of the prognostic model in different clinical parameters, we stratified the patients according to the different clinical parameters and performed survival analysis. We also explored the relationship between these eight prognostic RBPs and clinical parameters. A $P < 0.05$ was considered significant.

Gene Set Enrichment Analysis

We divided the patients into low-risk and high-risk groups based on the median risk score of the prognostic model, and then performed gene set enrichment analysis (GSEA) by using GSEA_4.0.3 software[6]. A $P < 0.05$ and FDR < 0.25 were considered to be significant differences.

Express Level and Prognostic Significance Verification of Prognostic Related RBPs

We used The Human Protein Atlas (HPA[7]) online database to verify the protein expression levels of these prognostic related RBPs. And the Kaplan–Meier plotter[8] online tool was used to assess the prognostic significance of these prognostic related RBPs in ccRCC patients.

Construction and Evaluation of Prognostic Model for DFS

Since DFS is also important for the prognosis of tumor patients, we constructed a prognostic model for DFS. We downloaded

[6]https://www.gsea-msigdb.org/gsea/downloads.jsp
[7]https://www.proteinatlas.org/
[8]https://kmplot.com/analysis/

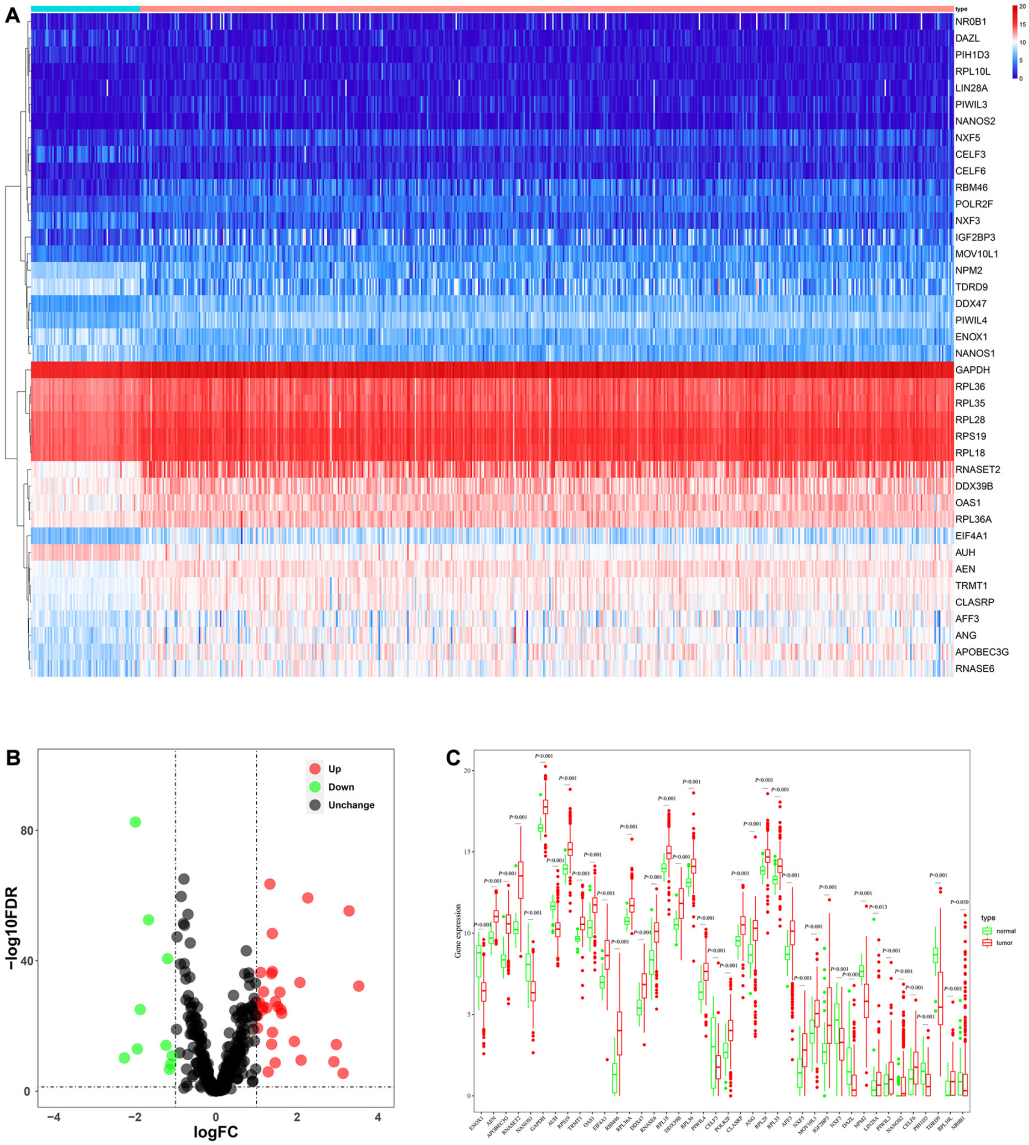


FIGURE 2 | The expression and distribution of differentially expressed RBPs in ccRCC. **(A)** Heatmap of the 40 differentially expressed RBPs; **(B)** Volcano plot of 455 RBPs **(C)** visualization of the expression levels of the 40 differentially expressed RBPs.

transcriptomic data from the cBioportal database for 436 ccRCC patients with DFS information. Then the prognostic RBPs were screened by Cox regression analysis and LASSO regression analysis and a prognostic model for DFS was constructed.

Statistical Analysis

R software (Version 4.0.0) was used for statistical analysis. The differentially expressed genes in tumor tissues and normal tissues were analyzed by "edgeR" package. Cox regression analysis was used to screen for genes associated with prognosis. The OS and DFS of patients were analyzed by Kaplan–Meier method and log-rank test. The "survival ROC" package was used to analyze the ROC curve. The "rms" package was used to draw the nomogram. The Student's *t*-test or non-parametric Mann–Whitney rank sum test was used to compare the correlation between risk score, prognostic genes, and clinicopathological variables. $P < 0.05$ was considered statistically significant.

RESULTS

Screening Differentially Expressed RBPs in ccRCC

The analysis process of this study was shown in **Figure 1**. Transcriptome data of ccRCC patients were downloaded from the TCGA database, including 72 normal renal tissue samples and 539 tumor tissue samples (**Supplementary Table S1**). The edger R package was used to process the data and identify the

differentially expressed RBPs. Of the 1542 RBPs (Gerstberger et al., 2014), 40 met our criteria (| log2 FC| > 1.0, FDR < 0.05), including 10 down-regulated RBPs and 30 up-regulated RBPs. **Figure 2** showed the expression and distribution of these differentially expressed RBPs.

Function and Pathway Enrichment Analysis of These Differentially Expressed RBPs

We performed GO and KEGG enrichment analyses for these differentially expressed RBPs using the WebGestalt online analysis tool to investigate the biological functions and molecular mechanisms of these genes. The analysis results were shown in **Table 1**. The biological processes analysis showed that these RBPs were significantly enriched in RNA catabolic process, posttranscriptional regulation of gene expression, translational initiation, regulation of cellular amide metabolic process, protein localization to endoplasmic reticulum, meiotic cell cycle, gene silencing, cellular process involved in reproduction in multicellular organism, transposition, and regulation of mRNA metabolic process. The cellular component showed that these RBPs were significantly enriched in polysome, ribosome, ribonucleoprotein granule, cytosolic part, and rough endoplasmic reticulum. In terms of molecular function, these RBPs were significantly enriched in mRNA binding, catalytic activity, acting on RNA, structural constituent of ribosome, helicase activity, nuclease activity, translation regulator activity, snRNA binding, double-stranded RNA binding, ATPase activity, and nucleotidyltransferase activity. Moreover, KEGG analysis showed that these RBPs were mainly enriched in ribosome, RNA transport, influenza A, mRNA surveillance pathway, herpes simplex infection, and ribosome biogenesis in eukaryotes.

TABLE 1 | KEGG pathway and GO enrichment analysis of differentially expressed RNA binding proteins.

	GO term	*P*-value
Biological processes	RNA catabolic process	2.81e-12
	Posttranscriptional regulation of gene expression	2.34e-10
	Translational initiation	3.72e-8
	Regulation of cellular amide metabolic process	5.31e-8
	Protein localization to endoplasmic reticulum	0.000002
	Meiotic cell cycle	0.000004
	Gene silencing	0.000016
	Cellular process involved in reproduction in multicellular organism	0.000034
	Transposition	0.000057
	Regulation of mRNA metabolic process	0.000073
Cellular component	Polysome	3.38e-10
	Ribosome	4.88e-8
	Ribonucleoprotein granule	6.13e-7
	Cytosolic part	0.000001
	Rough endoplasmic reticulum	0.019853
Molecular function	mRNA binding	4.25e-10
	Catalytic activity, acting on RNA	2.15e-8
	Structural constituent of ribosome	1.84e-7
	Helicase activity	0.000052
	Nuclease activity	0.000228
	Translation regulator activity	0.000561
	snRNA binding	0.005137
	Double-stranded RNA binding	0.018541
	ATPase activity	0.031468
	Nucleotidyltransferase activity	0.049070
KEGG pathway	Ribosome	1.03e-8
	RNA transport	0.000494
	Influenza A	0.000553
	mRNA surveillance pathway	0.001136
	Herpes simplex infection	0.008472
	Ribosome biogenesis in eukaryotes	0.014958

GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Prognostic Related RBPs Selection

We performed a univariate Cox regression analysis on all these differentially expressed RBPs and obtained 25 prognostic related RBPs (**Figure 3**). We further performed LASSO regression analysis on these 25 genes to screen the RBPs with prognostic significance, and obtained 9 RBPs including *APOBEC3G, AUH, DAZL, DDX47, EIF4A1, IGF2BP3, NR0B1, RPL36A, and TRMT1* (**Supplementary Figure S2**). And multivariate Cox regression analysis showed that 8 of the 9 RBPs, namely, *APOBEC3G, AUH, DAZL, EIF4A1, IGF2BP3, NR0B1, RPL36A, and TRMT1* independently predicted prognosis of ccRCC patients.

Prognostic Related Risk Score Model for OS Construction and Evaluation

We used these eight genes screened from multivariate Cox regression analysis to establish a prognostic model for OS (**Table 2**). Each ccRCC patient's risk score was calculated according to the following formula:

$$Risk\ score = (0.0951 \times \text{Exp APOBEC3G}) + (-0.1621 \times \text{Exp AUH}) + (0.0945 \times \text{Exp DAZL}) + (0.1571 \times \text{Exp EIF4A1}) + (0.1190 \times \text{Exp IGF2BP3}) + (0.0998 \times \text{Exp NR0B1}) + (0.1722 \times \text{Exp RPL36A}) + (0.2380 \times \text{Exp TRMT1})$$

Based on the median risk score, 539 ccRCC patients in the TCGA cohort were divided into low-risk and high-risk subgroups for survival analysis to assess the predictive power of the model. Survival analysis showed that patients in the high-risk group had lower OS than those in the low-risk group (P = 9.556e-13, **Figure 4A**). We then performed the time-dependent receiver operating characteristic (ROC) analysis to further evaluate the predictive performance of the eight RBPs signature, and the area under the ROC curve (AUC) of the model was 0.729 at 1 year, 0.702 at 3 years, and 0.726 at 5 years (**Figure 4B**). **Figure 4C** showed the survival status of each patient in the TCGA cohort assessed by risk score. Subsequently, to evaluate the applicability and stability of the prognostic model for OS, these 539 ccRCC patients in the TCGA cohort were randomly divided into a training data set and a validation data set. We then used the same formula to calculate the risk score of each patient to assess the predictive performance of the model. The results showed that patients in the high-risk group in the training data set had worse OS than those in the low-risk group (P = 1.908e-05, **Figure 5A**).

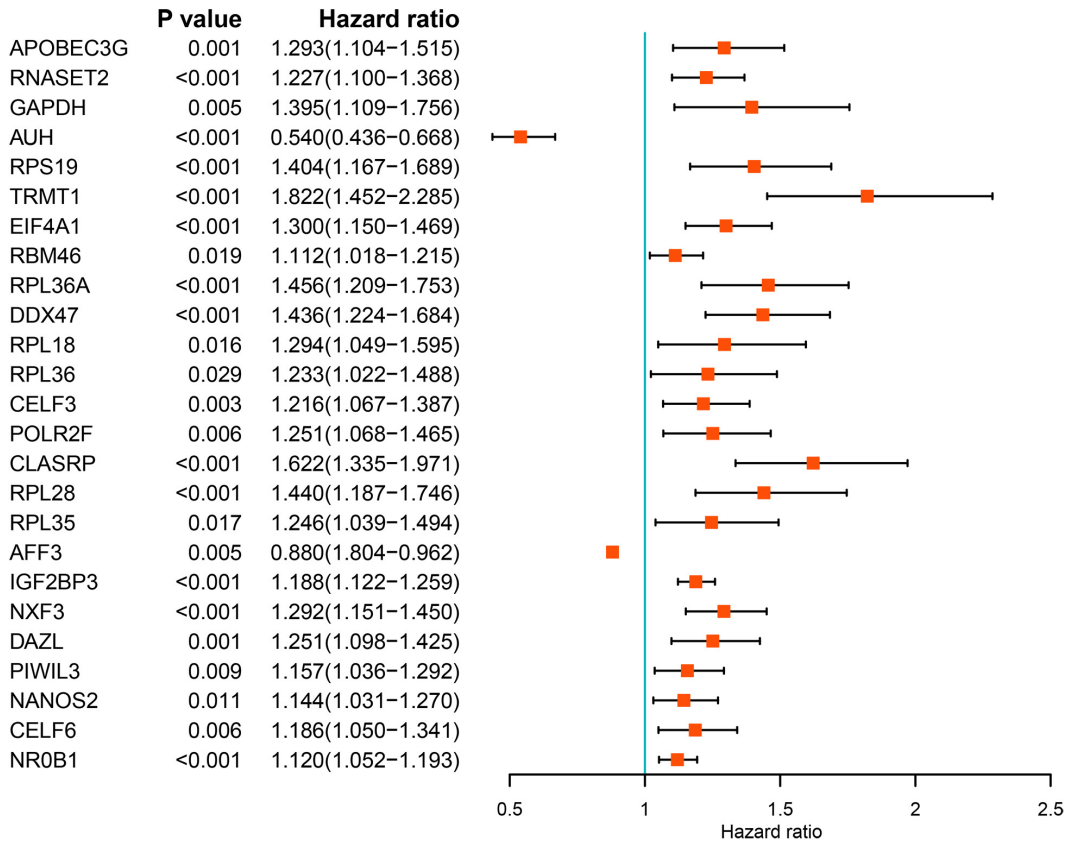


FIGURE 3 | Univariate Cox regression analysis of differentially expressed RBPs.

We found that the AUC was 0.750 at 1 year, 0.697 at 3 years, and 0.759 at 5 years (**Figure 5B**). And patients in the validation data set had similar results (**Figures 5C,D**). In addition, to assess whether the model has similar predictive power in other ccRCC patient cohorts, the same risk score formula was used for the E-MTAB-1980 dataset. Survival analysis also showed that patients in the high-risk group had lower OS than those in the low-risk group (P = 0.00033, **Figure 6A**), and the AUC of the model was 0.788 at 1 year, 0.783 at 3 years, and 0.795 at 5 years (**Figure 6B**). And **Figure 6C** showed the survival status of each patient in the E-MTAB-1980 cohort assessed by risk score. These results showed that the signature of these eight RBPs has good predictive performance and stability.

Next, we used the risk score in the prognostic model as the label to explore the functional differences between the two subgroups by conducting GSEA. The results showed that ribosome was mainly enriched in the high-risk group (**Figure 7A**), indicating that the regulation of RBPs was mainly involved in high-risk ccRCC patients. In addition, we performed a univariate Cox regression analysis for different clinical characteristics of ccRCC patients to evaluate their respective predictive significance. The results showed that age, tumor grade, tumor stage, primary tumor location, regional lymph node invasion, distant metastasis, and risk score were all associated with the OS of ccRCC patients (**Figure 7B**). However, multiple regression analysis showed that only age (P < 0.001), tumor grade (P = 0.020), tumor stage (P < 0.001), and risk score (P < 0.001) were independent prognostic factors related to OS of ccRCC patients (**Figure 7C**). Moreover, to establish a quantitative prognostic approach for ccRCC patients, we drew a nomogram based on the risk score and other clinical variables (**Figure 7D**). By drawing a vertical line between each prognosis axis and the total point axis, we can predict the survival probability of ccRCC patients at 1, 3, and 5 years. We also constructed calibration curves to evaluate the predictive performance of

TABLE 2 | Multivariate Cox regression analysis to identify prognosis-related RNA binding proteins.

Gene	Coef	Exp(coef)	se(coef)	z	*Pr* (> | z|)
APOBEC3G	0.0951	1.0998	0.0844	1.1264	0.2600
AUH	−0.1621	0.8504	0.1318	−1.2299	0.2187
DAZL	0.0945	1.0991	0.0695	1.3597	0.1739
EIF4A1	0.1571	1.1701	0.0710	2.2138	0.0268
IGF2BP3	0.1190	1.1264	0.0346	3.4376	0.0006
NR0B1	0.0998	1.1050	0.0366	2.7241	0.0064
RPL36A	0.1722	1.1879	0.1493	1.1532	0.2488
TRMT1	0.2380	1.2687	0.1632	1.4583	0.1448

Coef, coefficient.

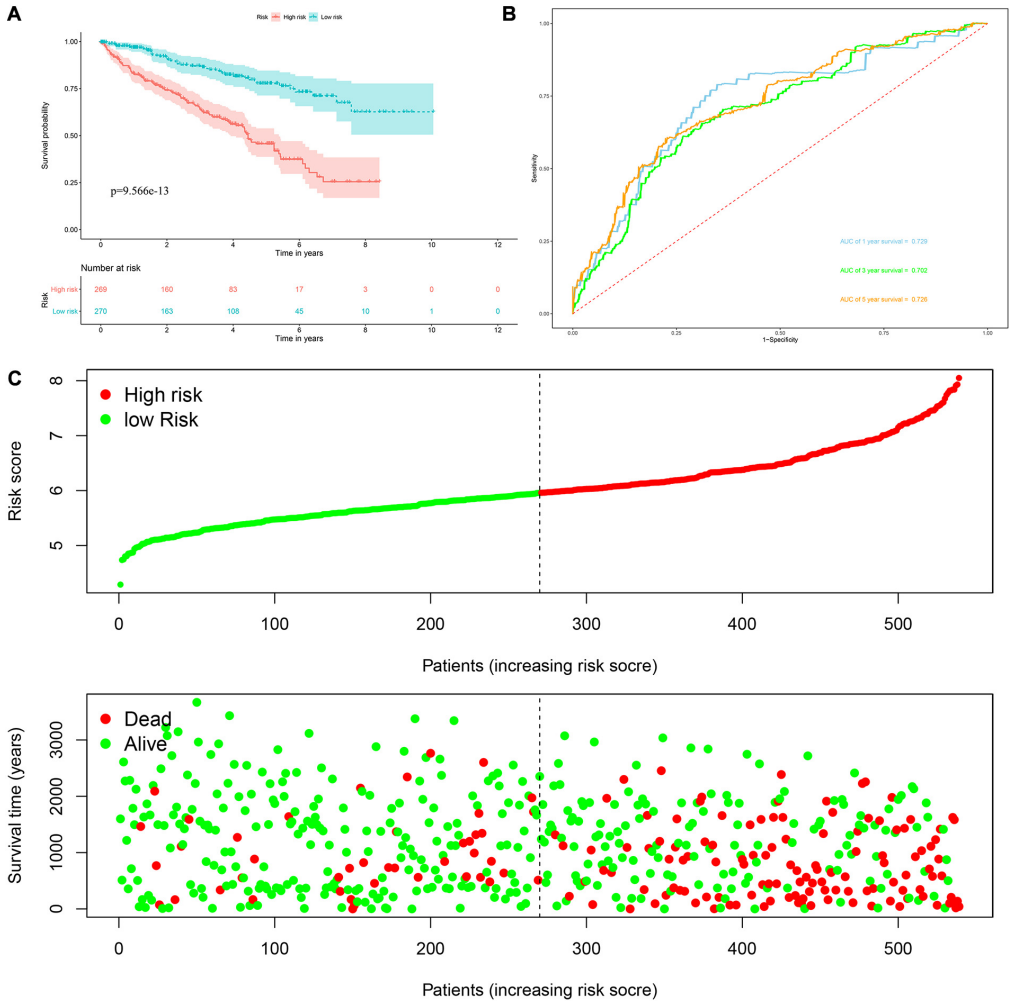


FIGURE 4 | Risk score analysis of the eight RBPs prognostic model for OS in the TCGA cohort. **(A)** Kaplan–Meier OS curve analysis; **(B)** time dependent ROC curve analysis; **(C)** survival status of each patient.

the nomogram, and the results showed that there was high consistency between the predicted results and the actual results (**Figures 7E–G**). And we used the TCGA and E-MTAB-1980 cohorts to verify the accuracy and stability of nomogram to expand its clinical application and availability. Survival analysis showed that nomogram could better distinguish ccRCC patients with low survival rates in TCGA and E-MTAB-1980 cohorts ($P < 0.001$ and $P = 2.32e\text{-}05$, **Figures 7H,J**). Based on the nomogram, the AUC in the TCGA cohort was 0.867 at 1 year, 0.806 at 3 years and 0.778 at 5 years (**Figure 7I**), and the AUC in the E-MTAB-1980 cohort was 0.910 at 1 year, 0.917 at 3 years, and 0.892 at 5 years (**Figure 7K**).

Prognostic Value of the Prognostic Model for OS Stratified by Clinical Parameters

To explore the clinical significance of the signature based on these eight RBPs in the ccRCC patients stratified by different clinical parameters, we stratified ccRCC patients from TCGA database according to age, gender, grade, stage, T stage, M stage, and N stage. Kaplan–Meier survival curve analysis showed that the OS was significantly shorter for the ccRCC patients in the high-risk group compared to the low-risk group ccRCC patients (**Figure 8**). These results indicate that the signature of these eight RBPs can predict the prognosis of ccRCC patients without considering clinical parameters.

Relationship Between Prognostic Model for OS and Clinical Parameters

We analyzed the correlation between the prognostic model based on these eight RBPs and clinical parameters to explore whether the prognostic model might influence the progression of ccRCC. The results showed no significant correlation between age and prognostic model (**Figure 9A**). However, the risk score of females was significantly lower than that of male (**Figure 9B**), the risk score of G1-2 was significantly lower

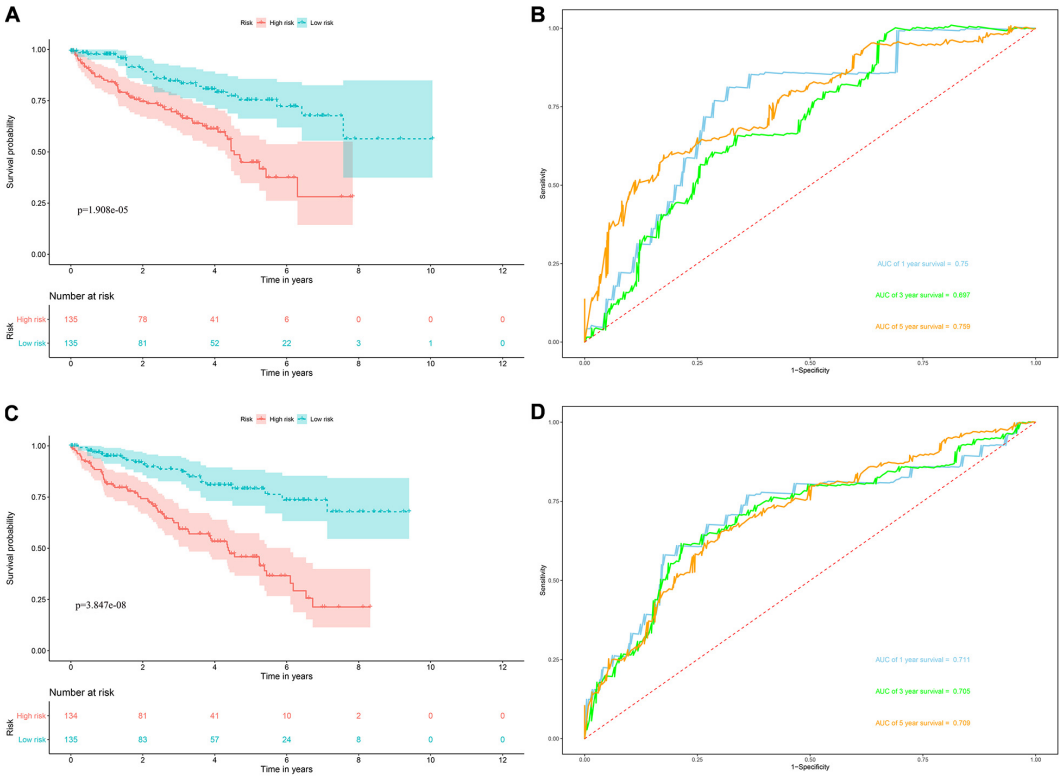


FIGURE 5 | Risk score analysis of the eight RBPs prognostic model for OS in the training and validation data set. **(A)** Kaplan–Meier OS curve analysis in the training data set; **(B)** time dependent ROC curve analysis in the training data set; **(C)** Kaplan–Meier OS curve analysis in the validation data set; **(D)** time dependent ROC curve analysis in the validation data set.

than that of G3-4 (**Figure 9C**), the risk score of stage I-II was significantly lower than that of stage III-IV (**Figure 9D**), the risk score of T1-2 was significantly lower than that of T3-4 (**Figure 9E**), the risk score of M0 was significantly lower than that of M1 (**Figure 9F**) (The N1 in the N stage is very small and cannot be analyzed). These results showed that prognostic model for OS was significantly correlated with ccRCC tumor progression.

Correlation Between Prognostic RBPs and Clinical Parameters

Based on the above results, we also analyzed the relationship between prognostic RBPs for OS and clinical parameters to further investigate the role of prognostic RBPs in ccRCC. The results showed that AUH, EIF4A1, IGF2BP3, and RPL36A were significantly correlated with gender; APOBEC3G, AUH, IGF2BP3, RPL36A, and TRMT1 were significantly correlated with grade; APOBEC3G, AUH, DAZL, IGF2BP3, RPL36A, and TRMT1 were significantly correlated with stage; APOBEC3G, AUH, DAZL, IGF2BP3, NR0B1, RPL36A, and TRMT1 were significantly correlated with T stage; APOBEC3G, AUH, IGF2BP3, RPL36A, and TRMT1 were significantly correlated with M stage. However, there was no significant correlation between NR0B1 and these clinical parameters (**Table 3**).

Express Level and Prognostic Significance Verification of Prognostic Related RBPs

To assess the prognostic significance of these prognostic related RBPs in ccRCC patients, we used the Kaplan–Meier plotter online tool to confirm the relationship between these genes and OS. The results showed that all the eight RBPs were related to the OS in ccRCC patients (**Figure 10**). Subsequently, we used the HPA online database to verify the protein expression levels of these prognostic related RBPs, the results showed that APOBEC3G, EIF4A1, and TRMT1 were significantly increased in ccRCC tissue compared with normal renal tissue (**Figures 11A,D,G**). And AUH, DAZL, IGF2BP3, and RPL36A were significantly reduced in ccRCC tissue compared with normal renal tissue (**Figures 11B,C,E,F**). However, the protein expression level of NR0B1 was not available on the HPA online database.

Construction of a Prognostic Model for DFS

In view of the important influence of DFS on the prognosis of ccRCC, we also constructed a prognostic model for DFS. The expression data of 436 ccRCC patients and the corresponding DFS information were download from the cBioportal database. We then identified 9 prognostic RBPs including *APOBEC3G,*

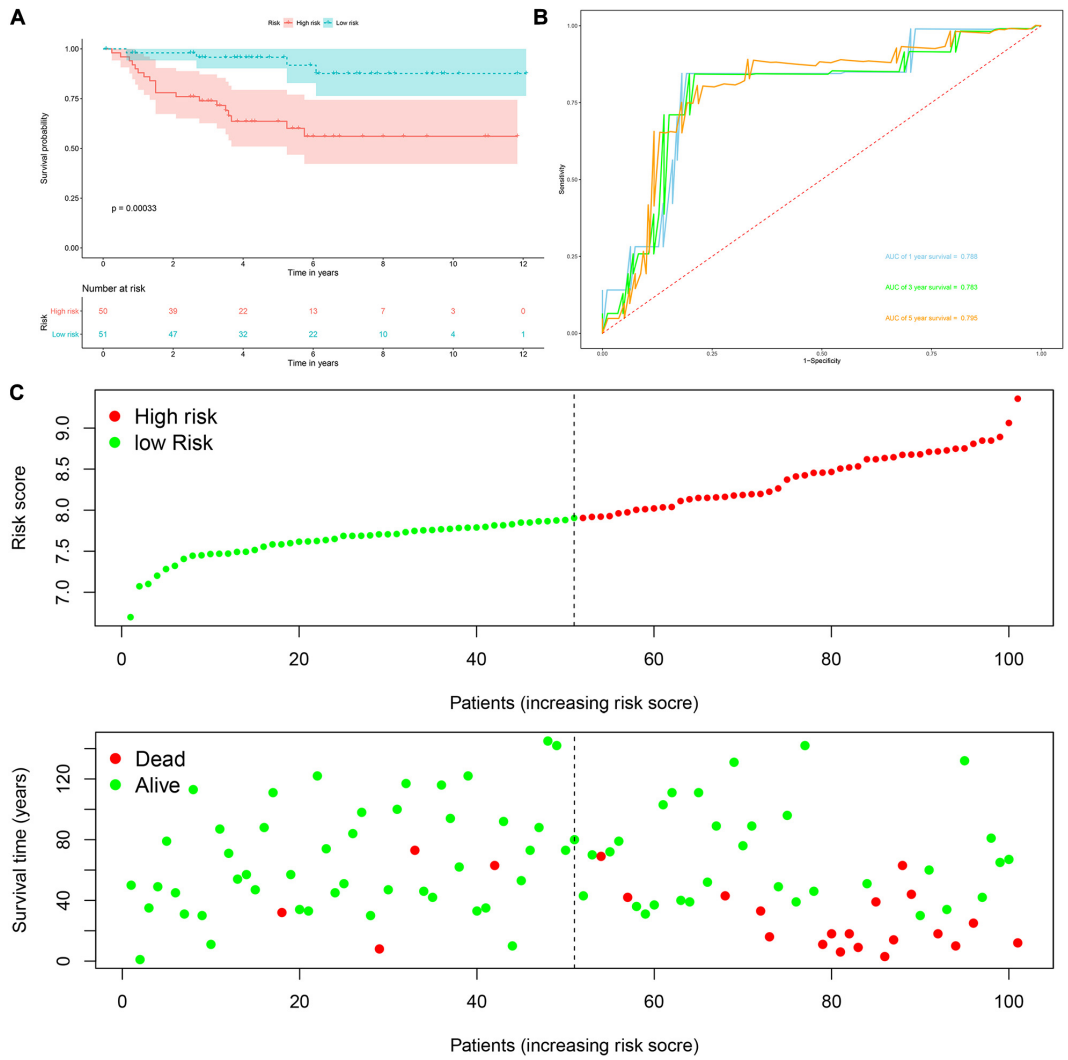


FIGURE 6 | Risk score analysis of the eight RBPs prognostic model for OS in the E-MTAB-1980 cohort. **(A)** Kaplan–Meier OS curve analysis; **(B)** time dependent ROC curve analysis; **(C)** survival status of each patient.

AUH, DDX47, IGF2BP3, MOV10L1, NANOS1, PIH1D3, TDRD9, and *TRMT1* by univariate Cox regression analysis, LASSO regression analysis and multivariate Cox regression analysis. We then constructed a prognostic model for DFS based on these nine prognostic RBPs and calculated each patient's risk score based on the following formula: Risk score = (0.0852 × Exp APOBEC3G) + (−0.3683 × Exp AUH) + (0.4195 × Exp DDX47) + (0.1445 × Exp IGF2BP3) + (−0.2077 × Exp MOV10L1) + (0.4206 × Exp NANOS1) + (0.7675 × Exp PIH1D3) + (−0.1011 × Exp TDRD9) + (0.2895 × Exp TRMT1). Based on the median risk score, these 436 ccRCC patients were divided into high-risk and low-risk groups for survival analysis to assess the predictive performance of the prognostic model. The results showed that patients in the high-risk group had worse DFS than those in the low-risk group (P = 1.110e-16, **Figure 12A**). We found that the AUC for DFS was 0.729 at 1 year, 0.764 at 3 years, and 0.782 at 5 years (**Figure 12D**). These results showed that the RBPs associated prognostic model for DFS has good predictive performance.

In addition, we randomly divided the whole dataset into a training data set (n = 218) and a validation data set (n = 218) to assess the applicability and stability of the prognostic model for DFS. We used the same formula to calculate each patient's risk score. Survival analysis showed that patients in the high-risk group in the training data set had worse DFS than those in the low-risk group (P = 1.127e-10, **Figure 12B**). The AUC for DFS was 0.718 at 1 year, 0.763 at 3 years, and 0.813 at 5 years (**Figure 12E**). Patients in the validation data set had similar results (**Figures 12C,F**).

Moreover, the prognostic value of the prognostic model for DFS and different clinical parameters were evaluated by Cox regression analysis. The results indicated that the tumor grade, tumor stage, primary tumor location, distant metastasis, and risk score of ccRCC patients were significantly correlated with DFS

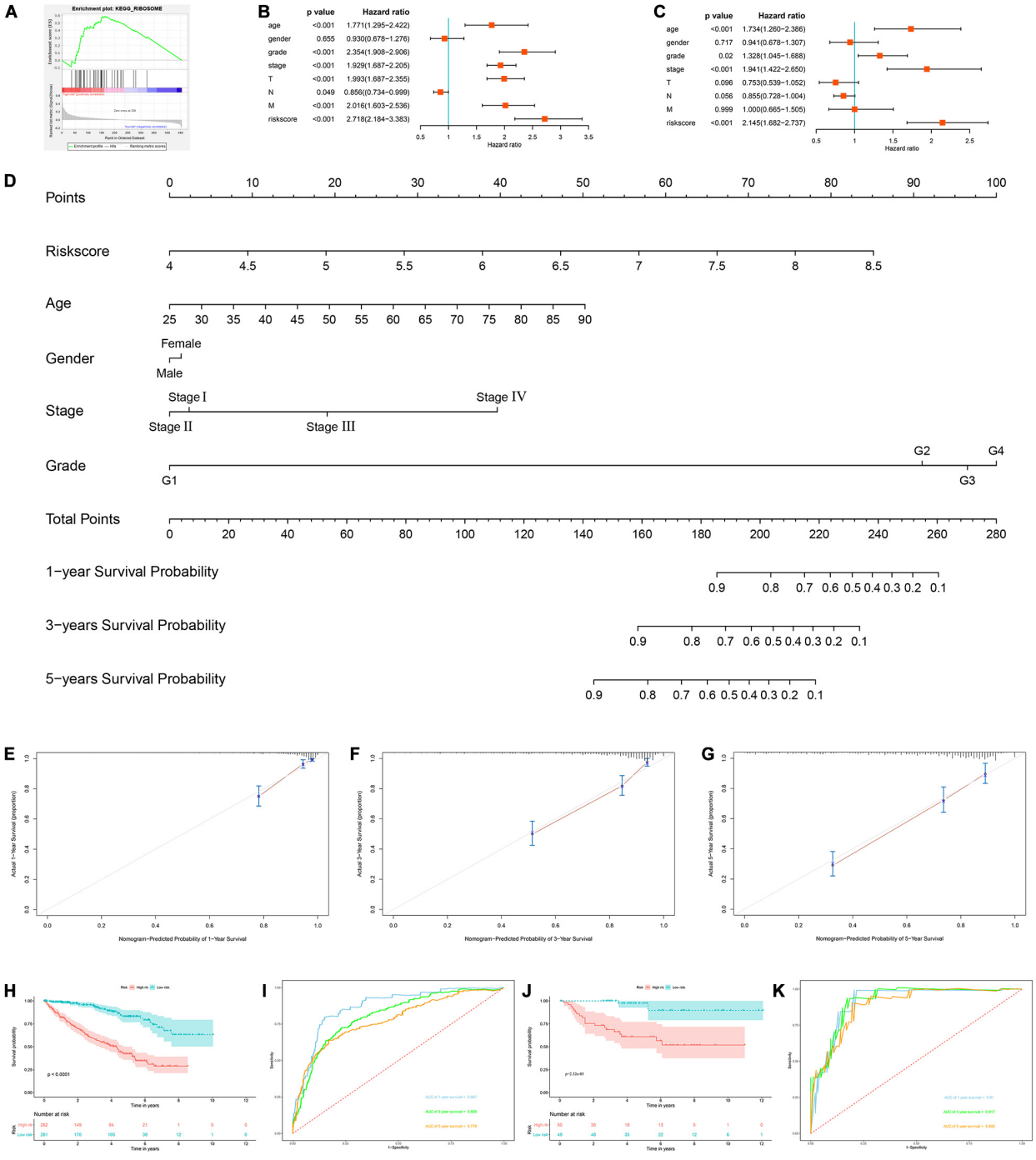


FIGURE 7 | Construction of a nomogram and assessment the prognostic significance of different clinical characteristics in ccRCC patients. **(A)** Gene set enrichment analysis comparing the high-risk and low-risk groups based on the TCGA cohort; **(B)** univariate Cox regression analysis of correlations between risk score for OS and clinical parameters; **(C)** multivariate Cox regression analysis of correlations between risk score for OS and clinical parameters; **(D)** nomogram for predicting the 1- year, 3-year, and 5-year OS of ccRCC patients; **(E–G)** calibration curves of the nomogram to predict OS at 1, 3, and 5 years; **(H)** Kaplan–Meier OS curve analysis in the TCGA cohort based the nomogram; **(I)** time dependent ROC curve analysis in the TCGA cohort based the nomogram; **(J)** Kaplan–Meier OS curve analysis in the E-MTAB-1980 cohort based the nomogram; **(K)** Time dependent ROC curve analysis in the E-MTAB-1980 cohort based the nomogram.

($P < 0.001$, **Figure 12G**). However, multiple regression analysis revealed that tumor grade, tumor stage, and risk score were independent prognostic factors associated with DFS ($P < 0.001$, **Figure 12H**). These results suggested that the RBPs associated prognostic model for DFS was also a good predictor of ccRCC patient outcomes.

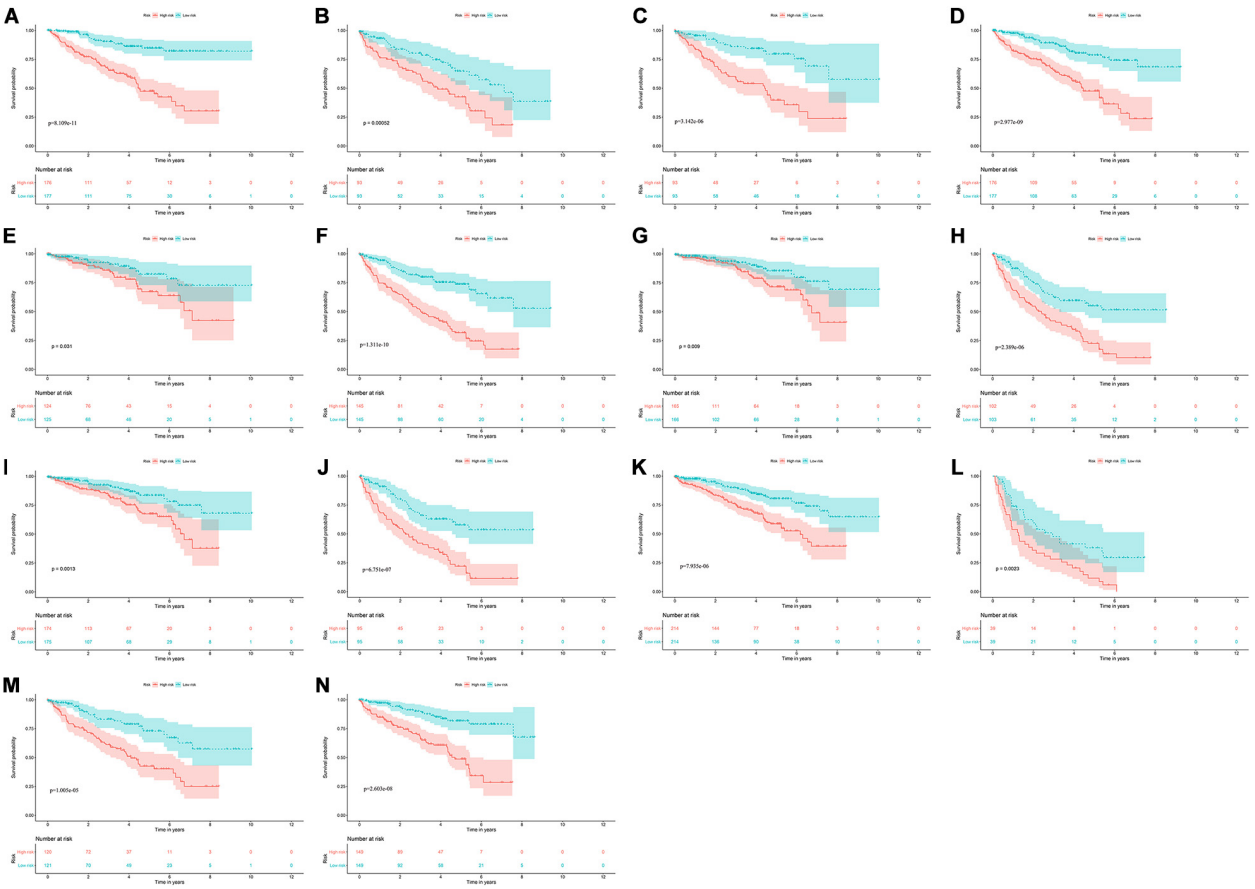


FIGURE 8 | Kaplan–Meier survival curves analysis stratified by different clinical parameters. **(A)** Age ≤ 65; **(B)** Age > 65; **(C)** Female; **(D)** Male; **(E)** Grade1-2; **(F)** Grade3-4; **(G)** Stage I-II; **(H)** Stage III-IV; **(I)** T stage1-2; **(J)** T stage3-4; **(K)** M stage0; **(L)** M stage1; **(M)** N stage0; **(N)** N stage1-X.

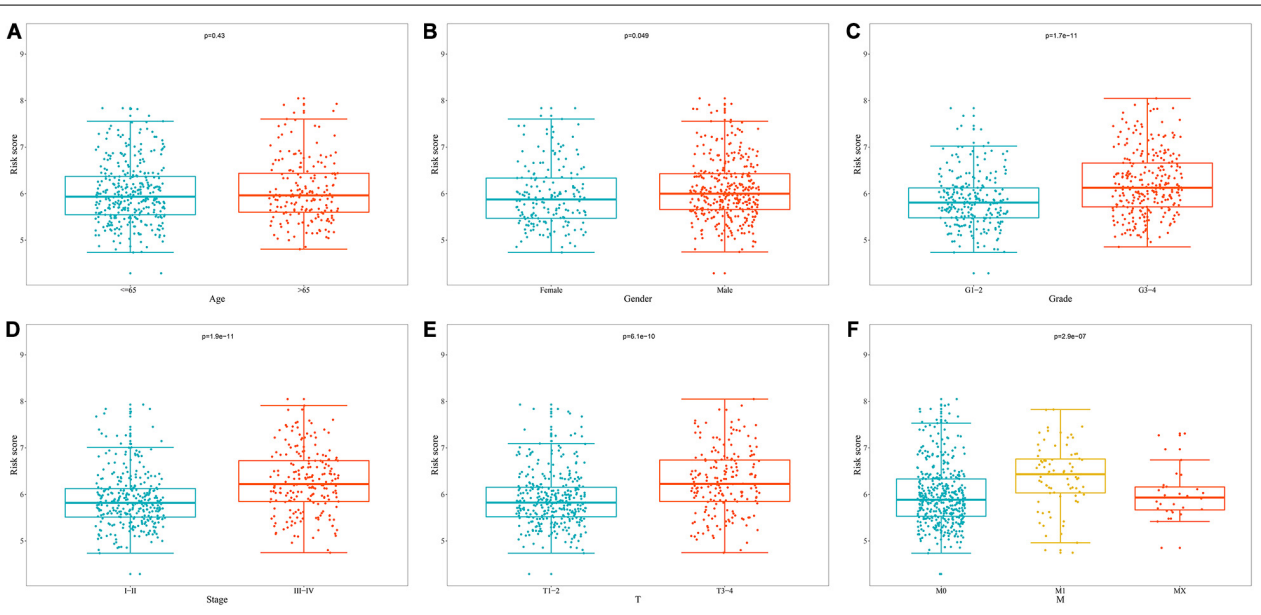


FIGURE 9 | Relationship between prognostic model for OS and clinical parameters. **(A)** Age; **(B)** Gender; **(C)** Grade; **(D)** Stage; **(E)** T stage; **(F)** M stage.

TABLE 3 | The relationship between prognostic related RNA binding proteins and clinicopathologic parameters.

Gene		Gender		Grade		Stage		T stage		M stage	
		Female	Male	G1-2	G3-4	I-II	III-IV	T1-T2	T3-T4	M0	M1
N		186	353	249	282	331	205	349	190	428	78
APOBEC3G	*t*-value	1.432		5.900		5.688		5.095		4.057	
	P-value	0.153		<0.001		<0.001		<0.001		<0.001	
AUH	*t*-value	2.799		NA*		6.545		5.595		4.589	
	P-value	0.005		<0.001		<0.001		<0.001		<0.001	
DAZL	*t*-value	0.181		NA*		NA*		NA*		0.817	
	P-value	0.857		0.256		0.047		0.049		0.415	
EIF4A1	*t*-value	2.652		0.947		1.545		1.783		0.845	
	P-value	0.008		0.344		0.123		0.075		0.398	
IGF2BP3	*t*-value	2.566		6.141		NA*		NA*		NA*	
	P-value	0.011		<0.001		<0.001		<0.001		<0.001	
NR0B1	*t*-value	0.951		0.355		NA*		NA*		NA*	
	P-value	0.342		0.723		0.087		0.030		0.569	
RPL36A	*t*-value	NA*		3.359		4.685		3.871		2.464	
	P-value	<0.001		0.001		<0.001		<0.001		0.014	
TRMT1	*t*-value	0.526		3.356		2.443		2.059		2.225	
	P-value	0.599		<0.001		0.015		0.040		0.027	

*NA, not available. *Non-parametric Mann–Whitney rank sum test.*

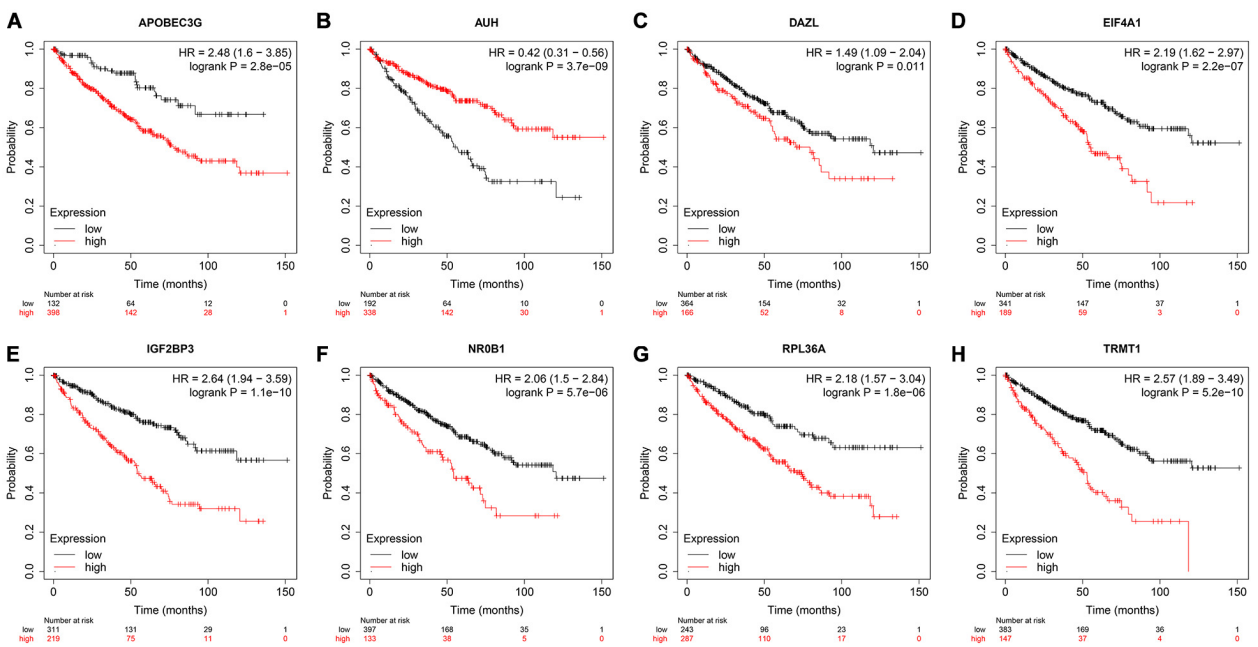


FIGURE 10 | Validation the prognostic value of the prognostic RBPs for OS in ccRCC by Kaplan–Meier plotter. **(A)** APOBEC3G; **(B)** AUH; **(C)** DAZL; **(D)** EIF4A1; **(E)** IGF2BP3; **(F)** NR0B1; **(G)** RPL36A; **(H)** TRMT1.

DISCUSSION

Malignant tumor is a kind of complex heterogeneous diseases, apart from the classic view that affect cancer or tumor suppressor gene signal channel change decision, It has also been found to be associated with post-transcriptional hijacking by tumor cells, enabling them to rapidly and stably regulate protein expression levels in response to intracellular and extracellular signaling changes to adapt to local microenvironments (Pereira et al., 2017). RBPs are a key player in post-transcriptional events, participating in almost all post-transcriptional regulation, controlling intracellular transcript metabolism and function, and maintaining homeostasis. Multiple studies have reported that RBPs are dysregulated in cancers and regulate cancers

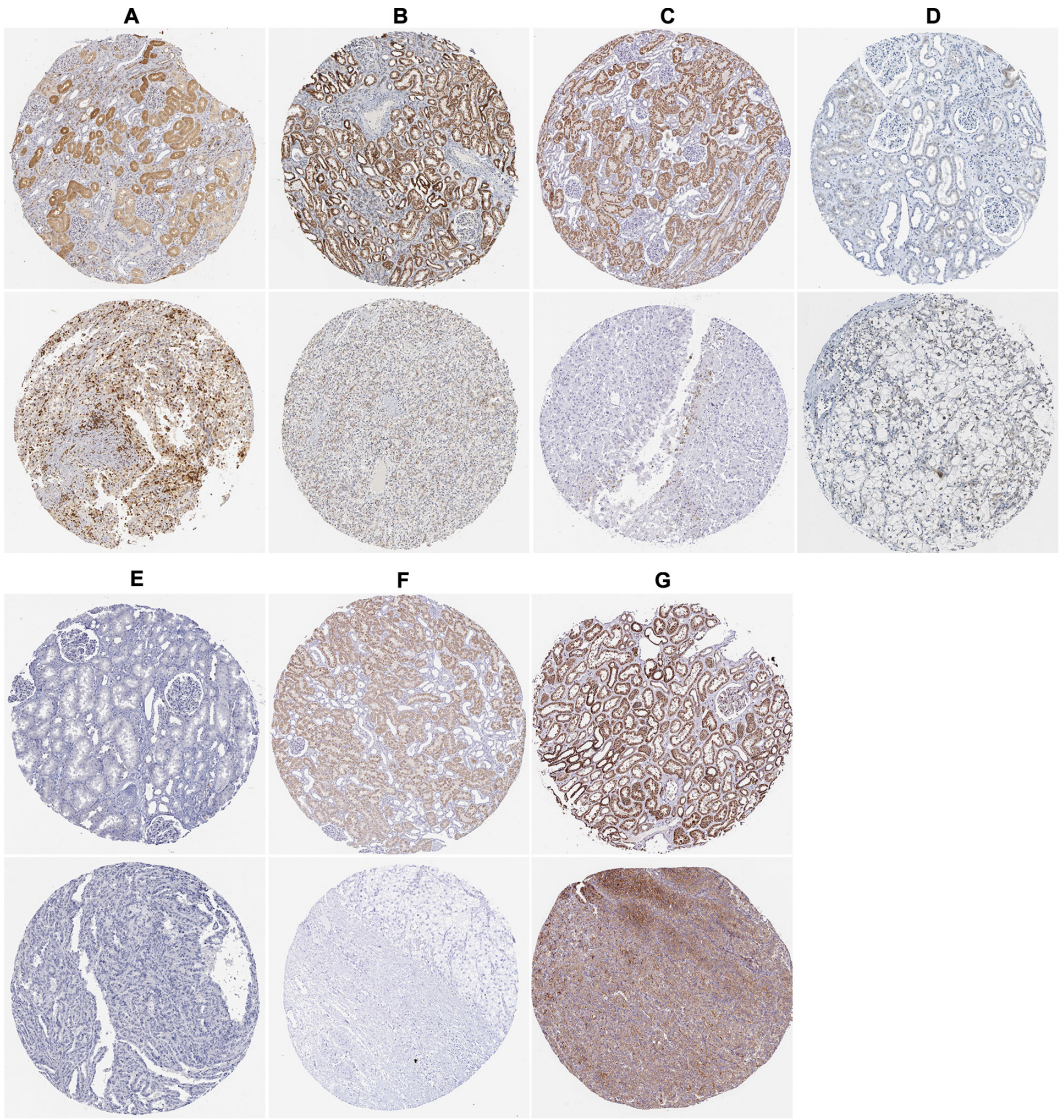


FIGURE 11 | The expression status of the prognostic RBPs proteins in ccRCC and normal renal tissues in the HPA database. **(A)** APOBEC3G; **(B)** AUH; **(C)** DAZL; **(D)** EIF4A1; **(E)** IGF2BP3; **(F)** RPL36A; **(G)** TRMT1.

progression through a variety of mechanisms, including transcriptional and posttranscriptional regulation, genomic change, and posttranslational modification (Patry et al., 2003; Busà et al., 2007; Ortiz-Zapater et al., 2011; Song et al., 2010; Zong et al., 2014; Pérez-Guijarro et al., 2016). However, the expression pattern and role of RBPs in ccRCC are rarely reported. In this study, we systematically analyzed the transcriptome data of ccRCC patients from TCGA database, and identified differential expression RBPs between tumor tissue and normal kidney tissue. We then performed functional enrichment analysis to evaluate their biological function, and performed univariate Cox regression analysis, LASSO regression analysis and multivariate Cox regression analysis to screen prognostic related RBPs and constructed a prognostic risk score model for OS based on these RBPs. In addition, we also built a prognostic model for DFS to predict ccRCC prognosis based on prognostic related RBPs.

The biological function and pathway enrichment analysis of differentially expressed RBPs showed that these genes were significantly enriched in posttranscriptional regulation of gene expression, translational initiation, transposition, protein localization to endoplasmic reticulum, RNA catabolic process, regulation of cellular amide metabolic process, regulation of mRNA metabolic process, gene silencing, ribonucleoprotein granule, mRNA binding, ribosome, polysome, catalytic activity, acting on RNA, translation regulator activity, nuclease activity, double-stranded RNA binding, nucleotidyltransferase activity, RNA transport, and mRNA surveillance pathway, which involved RNA processing, splicing, localization, RNA metabolism and subsequent translation regulation. Previous studies have shown

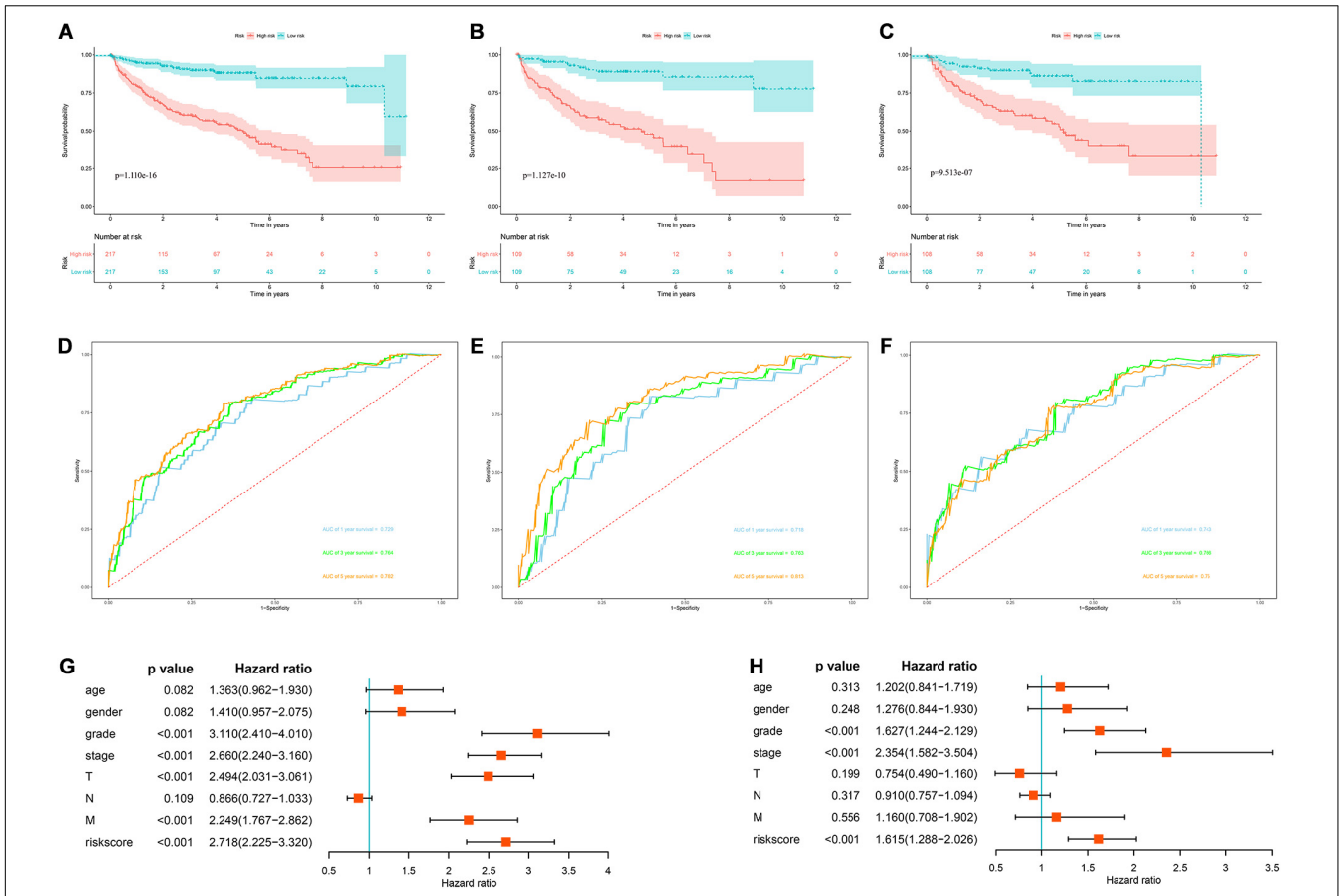


FIGURE 12 | The RBPs-related prognostic model for DFS. **(A)** Kaplan–Meier DFS curve analysis in the cBioportal cohort; **(B)** Kaplan–Meier DFS curve analysis in the training data set; **(C)** Kaplan–Meier DFS curve analysis in the validation data set **(D)** time dependent ROC curve analysis in the cBioportal cohort; **(E)** time dependent ROC curve analysis in the training data set; **(F)** time dependent ROC curve analysis in the validation data set; **(G)** univariate Cox regression analysis of correlations between risk score for DFS and clinical parameters; **(H)** multivariate Cox regression analysis of correlations between risk score for DFS and clinical parameters.

that multiple RBPs regulatory mechanisms have been identified in cancers, including transcriptional and posttranscriptional regulation, genomic change, and posttranslational modification (Patry et al., 2003; Trabucchi et al., 2009; Ishii et al., 2014; Preca et al., 2015). In lung adenocarcinoma, splice regulator RBM10 inhibits tumor cell proliferation and Notch signaling activity (Bechara et al., 2013). Cancer transcription factor MYC up-regulates the mRNA expression of hnRNPA1 and hnRNPA2 in gliomas, which promotes the synthesis of pyruvate kinase M subtype 2 (PKM2) and participates in glycolytic transformation (Clower et al., 2010). IMP1 has been reported to be elevated in multiple tumors, and reduced IMP1 expression can impair the normal transmission and local translation of adhesive and motif-related target mRNAs (Gu et al., 2009). EIF4E is a key factor in mRNA cycling and translation, and it has been found that EIF4E is overexpressed in a variety of tumors and is associated with poor prognosis (Ruggero et al., 2004). These results suggest that RBPs may influence the occurrence and progression of tumors by regulating multiple biological processes including RNA processing, RNA metabolism, RNA transport, translation regulation and mRNA surveillance pathway.

In addition, we performed univariate Cox regression analysis, LASSO regression analysis and multivariate Cox regression analysis on these differentially expressed RBPs, and 8 prognostic related RBPs including *APOBEC3G, AUH, DAZL, EIF4A1, IGF2BP3, NR0B1, RPL36A,* and *TRMT1* were selected. APOBEC3G, a member of the Apolipoprotein B mRNA editing enzyme-catalyzed polypeptide (APOBEC) family, was found to be overexpressed in renal carcinoma tissues and cell lines (Komohara et al., 2007), consistent with our results. Olson et al. (2018) found that this family is the source of somatic mutations in tumor cells that drive tumor evolution and may be associated with tumor cell recurrence, metastasis, and treatment resistance. AUH was found to be under-expressed in RCC and significantly associated with poorer survival in patients (Zhang et al., 2019), which is similar to our results. The DAZL mutation was found to be associated with testicular cancer (Ruark et al., 2013). The main function of EIF4A1 is to release mRNA structure in combination with other translation factors (Qi et al., 2013). EIF4A1 has been reported to be associated with malignant phenotypes of tumor cells, tumor-specific survival, and susceptibility to therapeutic drugs

(Nagel et al., 2010; Liang et al., 2014). Wei et al. (2019) found that miR-1284 inhibited the progression of gastric cancer by targeting EIF4A1. IGF2BP3 has been found to be overexpressed in a variety of tumors including lung (Wang et al., 2003), colon (Li et al., 2009), and liver cancers (Jeng et al., 2008). Accumulating studies have shown that IGF2BP3 is a promising prognostic factor for a variety of cancers including gastric cancer and RCC (Kim et al., 2014; Tschirdewahn et al., 2019). NR0B1 is a member of the orphan receptor family and is normally expressed mainly in the adrenal cortex, ovaries and support cells (Ikeda et al., 1996). Studies have found that NR0B1 is abnormally expressed in endometrial cancer, prostate cancer, lung cancer and other cancers, and plays an important role (Saito et al., 2005; Seo et al., 2007; Nakamura et al., 2009). Oda et al. (2009) found that NR0B1 mainly affects tumor cell invasion, colony formation and tumorigenic activity, and is related to the malignant potential of lung adenocarcinoma. RPL36A mainly encodes ribosomal protein L36a. Kim et al. (2004) found that overexpression of RPL36A in hepatocellular carcinoma was associated with enhanced cell proliferation, and RPL36A may be a potential target for anticancer therapy for hepatocellular carcinoma. Alshabi et al. (2019) also found that high expression of RPL36A was associated with the tumorigenesis of glioblastoma multiform. Nagel et al. (2010) found that TRMT1 was involved in the activation of LYL1 in leukemia cells and thus affected the differentiation of lymphocytes. GSEA analysis results showed that the regulation of RBPs was mainly concentrated in patients in the high-risk group, indicating that RBPs mainly regulates and affects patients in the high-risk group. However, the exact molecular mechanisms are unknown, and further exploration of possible mechanisms may be valuable. Subsequently, we constructed a prognostic model for OS based on these 8 RBPs to predict the prognosis of ccRCC patients. Survival analysis and ROC curve analysis showed that the model has good predictive performance. We then plotted a nomogram to establish a quantitative assessment method to predict the survival probability of ccRCC patients. According to our prognostic model for OS, patients with poor prognosis can be screened out, which may be conducive to timely adjustment of treatment regimens and individualized treatment.

Further analysis showed that the prognostic model for OS could independently predict the prognosis of ccRCC patients and was associated with the progression of ccRCC tumors. And the results of Kaplan-Meier Plotter online tool analysis showed that all 8 prognostic RBPs were related to OS in ccRCC patients. Moreover, we constructed an RBPS-related prognostic model for DFS, showing that this prognostic model can also independently predict the prognosis of ccRCC patients.

Overall, our study provides new insights into the occurrence and progress of ccRCC. In addition, the prognostic models for OS and DFS based on prognostic RBPs have good predictive performance, which are helpful to improve the clinical treatment decision and monitor the prognosis of patients. However, there are limitations in our study. First, our study is mainly based on a single bioomics information, and different characteristics of different platforms may lead to patient heterogeneity. Second, the model construction and validation of this study were designed by retrospective analysis, and the model still needs to be validated through a prospective clinical cohort. Moreover, the lack of clinical prognostic information in the study analysis may reduce the reliability of statistics. Finally, the prognostic models for OS and DFS based on prognostic RBPs showed good predictive performance. However, the exact molecular mechanisms of these prognostic RBPs involved in the occurrence, progression, and prognosis of renal cancer are still unclear, and the possible molecular mechanism and biological function need to be further explored.

CONCLUSION

We systematically analyzed the biological function and prognostic value of RBPs in ccRCC by using a variety of bioinformatics techniques. These RBPs may be involved in the pathogenesis, progression and metastasis of tumors. For the first time, we established prognostic risk score models for OS and DFS based on prognostic RBPs, and revealed they are independent prognostic factors related to OS and DFS in ccRCC patients. Our results are helpful to understand the molecular mechanism of ccRCC from a new perspective and to develop new prognostic markers or therapeutic targets.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov/.

AUTHOR CONTRIBUTIONS

YW designed the study and performed the data analysis. XW, HF, BH, and BL performed the data analysis. YL, YR, XL, ZL, SW, and JL performed the data analysis and revised the manuscript. TW designed the study and revised the manuscript. All the authors read and approved the final manuscript.

FUNDING

This study was supported by the Medical Youth Top Talent Program of Hubei Provincial.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.617872/full#supplementary-material

REFERENCES

Alshabi, A. M., Vastrad, B., Shaikh, I. A., and Vastrad, C. (2019). Identification of crucial candidate genes and pathways in Glioblastoma multiform by bioinformatics analysis. *Biomolecules* 9:201. doi: 10.3390/biom9050201

Battaglia, M., and Lucarelli, G. (2015). The role of renal surgery in the era of targeted therapy: the urologist's perspective. *Urologia* 82, 137–138. doi: 10.5301/uro.5000105

Bechara, E. G., Sebestyén, E., Bernardis, I., Eyras, E., and Valcárcel, J. (2013). RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. *Mol. Cell* 52, 720–733. doi: 10.1016/j.molcel.2013.11.010

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Busà, R., Paronetto, M. P., Farini, D., Pierantozzi, E., Botti, F., Angelini, D. F., et al. (2007). The RNA-binding protein Sam68 contributes to proliferation and survival of human prostate cancer cells. *Oncogene* 26, 4372–4382. doi: 10.1038/sj.onc.1210224

Clower, C. V., Chatterjee, D., Wang, Z., Cantley, L. C., Vander Heiden, M. G., and Krainer, A. R. (2010). The alternative splicing repressors hnRNP A1/A2 and PTB influence pyruvate kinase isoform expression and cell metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1894–1899. doi: 10.1073/pnas.0914845107

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. *Nat. Rev. Genet.* 15, 829–845. doi: 10.1038/nrg3813

Gu, W., Pan, F., and Singer, R. H. (2009). Blocking beta-catenin binding to the ZBP1 promoter represses ZBP1 expression, leading to increased proliferation and migration of metastatic breast-cancer cells. *J. Cell Sci.* 122(Pt 11), 1895–1905. doi: 10.1242/jcs.045278

Hentze, M. W., Castello, A., Schwarzl, T., and Preiss, T. (2018). A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* 19, 327–341. doi: 10.1038/nrm.2017.130

Hua, X., Chen, J., Ge, S., Xiao, H., Zhang, L., and Liang, C. (2020). Integrated analysis of the functions of RNA binding proteins in clear cell renal cell carcinoma. *Genomics* doi: 10.1016/j.ygeno.2020.10.016 [Epub ahead of print].

Iadevaia, V., and Gerber, A. P. (2015). Combinatorial control of mRNA fates by RNA-binding proteins and non-coding RNAs. *Biomolecules* 5, 2207–2222. doi: 10.3390/biom5042207

Ikeda, Y., Swain, A., Weber, T. J., Hentges, K. E., Zanaria, E., Lalli, E., et al. (1996). Steroidogenic factor 1 and Dax-1 colocalization in multiple cell lineages: potential links in endocrine development. *Mol. Endocrinol.* 10, 1261–1272. doi: 10.1210/mend.10.10.9121493

Ishii, H., Saitoh, M., Sakamoto, K., Kondo, T., Katoh, R., Tanaka, S., et al. (2014). Epithelial splicing regulatory proteins 1 (ESRP1) and 2 (ESRP2) suppress cancer cell motility via different mechanisms. *J. Biol. Chem.* 289, 27386–27399. doi: 10.1074/jbc.M114.589432

Jeng, Y. M., Chang, C. C., Hu, F. C., Chou, H. Y., Kao, H. L., Wang, T. H., et al. (2008). RNA-binding protein insulin-like growth factor II mRNA-binding protein 3 expression promotes tumor invasion and predicts early recurrence and poor prognosis in hepatocellular carcinoma. *Hepatology* 48, 1118–1127. doi: 10.1002/hep.22459

Kim, H. J., Kim, G. E., Lee, J. S., Lee, J. H., Nam, J. H., and Choi, C. (2014). Insulin-like growth factor-II mRNA-binding protein 3 expression in effusion cytology: a marker for metastatic adenocarcinoma cells and a potential prognostic indicator in gastric adenocarcinoma. *Acta Cytol.* 58, 167–173. doi: 10.1159/000357199

Kim, J. H., You, K. R., Kim, I. H., Cho, B. H., Kim, C. Y., and Kim, D. G. (2004). Over-expression of the ribosomal protein L36a gene is associated with cellular proliferation in hepatocellular carcinoma. *Hepatology* 39, 129–138. doi: 10.1002/hep.20017

Komohara, Y., Suekane, S., Noguchi, M., Matsuoka, K., Yamada, A., and Itoh, K. (2007). Expression of APOBEC3G in kidney cells. *Tissue Antigens* 69, 95–98. doi: 10.1111/j.1399-0039.2006.00725.x

Li, D., Yan, D., Tang, H., Zhou, C., Fan, J., Li, S., et al. (2009). IMP3 is a novel prognostic marker that correlates with colon cancer progression and pathogenesis. *Ann. Surg. Oncol.* 16, 3499–3506. doi: 10.1245/s10434-009-0648-5

Liang, S., Zhou, Y., Chen, Y., Ke, G., Wen, H., and Wu, X. (2014). Decreased expression of EIF4A1 after preoperative brachytherapy predicts better tumor specific survival in cervical cancer. *Int. J. Gynecol. Cancer* 24, 908–915. doi: 10.1097/IGC.0000000000000152

Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* 47, W199–W205. doi: 10.1093/nar/gkz401

Ljungberg, B., Albiges, L., Abu-Ghanem, Y., Bensalah, K., Dabestani, S., Fernández-Pello, S., et al. (2019). European association of urology guidelines on renal cell Carcinoma: the 2019 update. *Eur. Urol.* 75, 799–810. doi: 10.1016/j.eururo.2019.02.011

Masuda, K., and Kuwano, Y. (2019). Diverse roles of RNA-binding proteins in cancer traits and their implications in gastrointestinal cancers. *Wiley Interdiscip. Rev. RNA* 10:e1520. doi: 10.1002/wrna.1520

Moch, H., Cubilla, A. L., Humphrey, P. A., Reuter, V. E., and Ulbright, T. M. (2016). The 2016 WHO classification of tumours of the urinary system and male genital organs-part a: renal, penile, and testicular tumours. *Eur. Urol.* 70, 93–105. doi: 10.1016/j.eururo.2016.02.029

Nagel, S., Venturini, L., Meyer, C., Kaufmann, M., Scherr, M., Drexler, H. G., et al. (2010). Multiple mechanisms induce ectopic expression of LYL1 in subsets of T-ALL cell lines. *Leuk Res.* 34, 521–528. doi: 10.1016/j.leukres.2009.06.020

Nakamura, Y., Suzuki, T., Arai, Y., and Sasano, H. (2009). Nuclear receptor DAX1 in human prostate cancer: a novel independent biological modulator. *Endocr. J.* 56, 39–44. doi: 10.1507/endocrj.k08e-177

Narla, A., and Ebert, B. L. (2010). Ribosomopathies: human disorders of ribosome dysfunction. *Blood* 115, 3196–3205. doi: 10.1182/blood-2009-10-178129

Oda, T., Tian, T., Inoue, M., Ikeda, J., Qiu, Y., Okumura, M., et al. (2009). Tumorgenic role of orphan nuclear receptor NR0B1 in lung adenocarcinoma. *Am. J. Pathol.* 175, 1235–1245. doi: 10.2353/ajpath.2009.090010

Olson, M. E., Harris, R. S., and Harki, D. A. (2018). APOBEC Enzymes as Targets for Virus and Cancer Therapy. *Cell Chem. Biol.* 25, 36–49. doi: 10.1016/j.chembiol.2017.10.007

Ortiz-Zapater, E., Pineda, D., Martínez-Bosch, N., Fernández-Miranda, G., Iglesias, M., Alameda, F., et al. (2011). Key contribution of CPEB4-mediated translational control to cancer progression. *Nat. Med.* 18, 83–90. doi: 10.1038/nm.2540

Patry, C., Bouchard, L., Labrecque, P., Gendron, D., Lemieux, B., Toutant, J., et al. (2003). Small interfering RNA-mediated reduction in heterogeneous nuclear ribonucleoparticule A1/A2 proteins induces apoptosis in human cancer cells but not in normal mortal cell lines. *Cancer Res.* 63, 7679–7688.

Pereira, B., Billaud, M., and Almeida, R. (2017). RNA-binding proteins in cancer: old players and new actors. *Trends Cancer* 3, 506–528. doi: 10.1016/j.trecan.2017.05.003

Pérez-Guijarro, E., Karras, P., Cifdaloz, M., Martínez-Herranz, R., Cañón, E., Graña, O., et al. (2016). Lineage-specific roles of the cytoplasmic polyadenylation factor CPEB4 in the regulation of melanoma drivers. *Nat. Commun.* 7:13418. doi: 10.1038/ncomms13418

Preca, B. T., Bajdak, K., Mock, K., Sundararajan, V., Pfannstiel, J., Maurer, J., et al. (2015). A self-enforcing CD44s/ZEB1 feedback loop maintains EMT and stemness properties in cancer cells. *Int. J. Cancer* 137, 2566–2577. doi: 10.1002/ijc.29642

Qi, M., Qi, Y., Ma, Y., He, R., Ji, Y., Sun, Z., et al. (2013). Over-expression of human cytomegalovirus miR-US25-2-3p downregulates eIF4A1 and inhibits HCMV replication. *FEBS Lett.* 587, 2266–2271. doi: 10.1016/j.febslet.2013.05.057

Rini, B. I., Campbell, S. C., and Escudier, B. (2009). Renal cell carcinoma. *Lancet* 373, 1119–1132. doi: 10.1016/S0140-6736(09)60229-4

Ruark, E., Seal, S., McDonald, H., Zhang, F., Elliot, A., Lau, K., et al. (2013). Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14. *Nat. Genet.* 45, 686–689. doi: 10.1038/ng.2635

Ruggero, D., Montanaro, L., Ma, L., Xu, W., Londei, P., Cordon-Cardo, C., et al. (2004). The translation factor eIF-4E promotes tumor formation and cooperates with c-Myc in lymphomagenesis. *Nat. Med.* 10, 484–486. doi: 10.1038/nm1042

Saito, S., Ito, K., Suzuki, T., Utsunomiya, H., Akahira, J., Sugihashi, Y., et al. (2005). Orphan nuclear receptor DAX-1 in human endometrium and its disorders. *Cancer Sci.* 96, 645–652. doi: 10.1111/j.1349-7006.2005.00101.x

Scheper, G. C., van der Knaap, M. S., and Proud, C. G. (2007). Translation matters: protein synthesis defects in inherited disease. *Nat. Rev. Genet.* 8, 711–723. doi: 10.1038/nrg2142

Seo, D. C., Sung, J. M., Cho, H. J., Yi, H., Seo, K. H., Choi, I. S., et al. (2007). Gene expression profiling of cancer stem cell in human lung adenocarcinoma A549 cells. *Mol. Cancer* 6:75. doi: 10.1186/1476-4598-6-75

Siegel, R. L., Miller, K. D., and Jemal, A. (2018). Cancer statistics, 2018. *CA Cancer J. Clin.* 68, 7–30. doi: 10.3322/caac.21442

Song, L., Wang, L., Li, Y., Xiong, H., Wu, J., Li, J., et al. (2010). Sam68 up-regulation correlates with, and its down-regulation inhibits, proliferation and tumourigenicity of breast cancer cells. *J. Pathol.* 222, 227–237. doi: 10.1002/path.2751

Tamma, R., Rutigliano, M., Lucarelli, G., Annese, T., Ruggieri, S., Cascardi, E., et al. (2019). Microvascular density, macrophages, and mast cells in human clear cell renal carcinoma with and without bevacizumab treatment. *Urol. Oncol.* 37, .e11–.e355. doi: 10.1016/j.urolonc.2019.01.025

Trabucchi, M., Briata, P., Garcia-Mayoral, M., Haase, A. D., Filipowicz, W., Ramos, A., et al. (2009). The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* 459, 1010–1014. doi: 10.1038/nature08025

Tschirdewahn, S., Panic, A., Püllen, L., Harke, N. N., Hadaschik, B., Riesz, P., et al. (2019). Circulating and tissue IMP3 levels are correlated with poor survival in renal cell carcinoma. *Int. J. Cancer* 145, 531–539. doi: 10.1002/ijc.32124

Wang, T., Fan, L., Watanabe, Y., McNeill, P. D., Moulton, G. G., Bangur, C., et al. (2003). L523S, an RNA-binding protein as a potential therapeutic target for lung cancer. *Br. J. Cancer* 88, 887–894. doi: 10.1038/sj.bjc.6600806

Wei, W., Cao, W., Zhan, Z., Yan, L., Xie, Y., and Xiao, Q. (2019). MiR-1284 suppresses gastric cancer progression by targeting EIF4A1. *Oncol. Targets Ther.* 12, 3965–3976. doi: 10.2147/OTT.S191015

Zhang, B., Wu, Q., Wang, Z., Xu, R., Hu, X., Sun, Y., et al. (2019). The promising novel biomarkers and candidate small molecule drugs in kidney renal clear cell carcinoma: evidence from bioinformatics analysis of high-throughput data. *Mol. Genet. Genomic Med.* 7:e607. doi: 10.1002/mgg3.607

Zhu, Z., He, A., Lin, L., Xu, C., Cai, T., and Lin, J. (2020). Biological functions and prognostic value of RNA binding proteins in clear cell renal cell Carcinoma. *J. Cancer* 11, 6591–6600. doi: 10.7150/jca.49175

Zong, F. Y., Fu, X., Wei, W. J., Luo, Y. G., Heiner, M., Cao, L. J., et al. (2014). The RNA-binding protein QKI suppresses cancer-associated aberrant splicing. *PLoS Genet.* 10:e1004289. doi: 10.1371/journal.pgen.1004289

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership