RESEARCH ARTICLE

# Integrated analysis of RNA-binding proteins in thyroid cancer

Jing Zhen[1‡], Zhe Song[2‡], WenJie Su[3], Qing-Cui Zeng[4], JiaCen Li[3]*, Qin Sun[1]*

1 Department of Geriatric Endocrinology, Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, Chengdu, P.R. China, **2** Medical Center of Vascular Surgery and Thyroid, Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, Chengdu, P.R. China, **3** Department of Anesthesiology, Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, Chengdu, P.R. China, **4** Department of Geriatric Intensive Care Unit, Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, Chengdu, P.R. China

‡ JZ and ZS are regarded as co-first authors.
* sunqin@med.uestc.edu.cn (QS); LiJJcd@163.com (JCL)
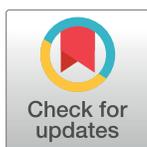
Abstract

Recently, the incidence of thyroid cancer (THCA) has been on the rise. RNA binding proteins (RBPs) and their abnormal expression are closely related to the emergence and pathogenesis of tumor diseases. In this study, we obtained gene expression data and corresponding clinical information from the TCGA database. A total of 162 aberrantly expressed RBPs were obtained, comprising 92 up-regulated and 70 down-regulated RBPs. Then, we performed a functional enrichment analysis and constructed a PPI network. Through univariate Cox regression analysis of key genes and found that *NOLC1* ($p = 0.036$), *RPS27L* ($p = 0.011$), *TDRD9* ($p = 0.016$), *TDRD6* ($p = 0.002$), *IFIT2* ($p = 0.037$), and *IFIT3* ($p = 0.02$) were significantly related to the prognosis. Through the online website Kaplan-Meier plotter and multivariate Cox analysis, we identified 2 RBP-coding genes (*RPS27L* and *IFIT3*) to construct a predictive model in the entire TCGA dataset and then validate in two subsets. In-depth analysis revealed that the data gave by this model, the patient's high-risk score is very closely related to the overall survival rate difference ($p = 0.038$). Further, we investigated the correlation between the model and the clinic, and the results indicated that the high-risk was in the male group ($p = 0.011$) and the T3-4 group ($p = 0.046$) was associated with a poor prognosis. On the whole, the conclusions of our research this time can make it possible to find more insights into the research on the pathogenesis of THCA, this could be beneficial for individualized treatment and medical decision making.

1. Introduction

The thyroid gland is an endocrine organ located in the neck of human body and secretes two hormones thyroxine and triiodothyronine that are essential for the normal functioning of all the cells as well as regulating human metabolism, controlling weight, blood pressure, heart rate, and body temperature [1–4]. Thyroid cancer arises from the parafollicular or follicular cells and considered as one of the most common malignant tumor of endocrine system. Over

the past 30 years, the incidence of thyroid cancer has been on the rise, each year it comprises for 3.4% of all tumor cases diagnosed globally [5–10]. The most famous types of thyroid tumors are papillary thyroid cancer (PTC) and follicular thyroid cancer with a better prognosis, and their survival rate can exceed 90% in the long-term development process. At present, PTC patients usually cured with conventional treatments including surgical treatment, radioactive iodine therapy or thyroid stimulating hormone suppression therapy [11–13]. Although most PTCs remain inert and survival rate of the patients is maximum, the recurrence and metastasis of tumors hinder the clinical treatment and resulted in poor prognosis [14–20]. For patients with locally advanced or distant metastatic PTC, the existing treatment methods are not enough and there is a need for a novel approach to diagnose and treat PTC.

RNA-binding proteins (RBPs) can interact with many types of RNA and have a very high appearance rate in various cells [21–25]. A total of 1542 RBPs were identified in human cells through high-throughput screening, accounting for 7.5% of all protein-coding genes [26–30]. RBPs refers to a kind of pleiotropic protein. RBPs transcribed after interacting with the target RNA, and finally regulate the level of gene expression. Previous pieces of research work revealed that RNA-binding proteins are involved in RNA metabolism and play a vital role in regulating RNA stability, alternative splicing, modification, localization, and translation [31–34]. Therefore, abnormalities in the expression of RBPs can cause several disorders including muscle atrophy, metabolic disorders, cancer, and germ cell development [35–41]. Dysregulated RBPs have been reported to mediate cell proliferation, apoptosis, angiogenesis, senescence, and metastasis in various cancer cells, the critical biological mechanisms of which include miRNAs regulation, alternative splicing, alternative Polyadenylation, RNA localization, RNA stability, and translational regulation [42–44]. Besides, some of these RBPs dysregulations are significantly associated with the prognosis of specific cancer patients. On the other hand, based on RBPs' functional role in cancer, several therapeutics targeting RBPs including small-molecule drugs, inhibitors, small peptides, and antisense oligonucleotides have been developed and applied in clinical trials [45–47]. However, till now RBPs have not been well explored in thyroid cancer. In the emergence and development of THCA, the role RBPs is still unclear. Although the treatment of THCA has achieved great results in the last ten years, due to the high incidence of tumor-specific deaths, it is still worth considering the prognosis of patients and the need for new treatment methods. Thus, the analysis of RBPs in THCA can provide new insights as potential biomarkers for treatment and pathogenesis of thyroid tumors.

In the current investigation, we collected the THCA data from the TCGA database and conducted various experiments to assess the potential molecular functions and clinical value of RBPs in THCA. In addition, we also analyzed many differentially expressed RBPs that were closely related to THCA. Meanwhile, we carried out enrichment analysis on the GO and KEGG pathways, in order to demonstrate the possible mechanism of RBP in THCA. We also established a model related to the prognosis and tested the accuracy of the model through survival analysis, ROC analysis, univariate Cox analysis, and multivariate Cox analysis.

2. Materials and methods

2.1 Datasets

We downloaded the THCA transcriptome expression profile and corresponding clinical information from the TCGA database (https://portal.gdc.cancer.gov/repository). The expression data was HTSeq-FPKM type, containing 510 THCA tissues and 58 adjacent nontumorous tissue samples.

2.2 Identification of DEGs and functional annotations

We collected 1542 genes related to RBP from the literature [21]. Then we extract RBPs-related genes from the TCGA-THCA data set, and carried out standardized correction on the FPKM data through R software for differential analysis. Next, identify the specific conditions of DEGs through the "LIMMA" package by comparison between data of 58 adjacent nontumorous tissue samples and these of 510 THCA tissues. Based on $p < 0.05$ and $|\log_2$ fold change (FC)$| \geq 1$, we can find RBPs with different expressions. Then through R software, two kinds of enrichment analysis of RBPs with very obvious differences were carried out, namely the Kyoto Encyclopedia of Genome Genomics Database (KEGG) enrichment analysis and Gene Ontology (GO) enrichment analysis ($p<0.05$).

2.3 Protein-protein interaction (PPI) network construction and module analysis

To predict the PPI protein network of differentially expressed TCGA-THCA RBPs and to analyze the degree of interaction between proteins, we used the online website STRING (https://string-db.org/) [48–51]. Next, to make the PPI network visualized, we used Cytoscape software (v3.7.2) and then MCODE plug-in included in the Cytoscape software to expand the very critical sections of the PPI network survey. We defined advanced options as K-Core = 2, Node Score cutoff = 0.2, and degree cutoff = 2.

2.4 Prognosis-related RBPs selection

We used univariate Cox regression to examine the expression levels of DEGs in the TCGA dataset ($p<0.05$ as statistical significance), and then evaluated the prognostic value of RBPs and survival time, and finally determined the RBPs related to overall survival time. Also, the Kaplan Meyer plotter (https://kmplot.com/analysis/) was used to further verify the prognostic value of various RBPs [52]. The RBPs with $p < 0.05$ were considered to be true prognostic RBPs.

2.5 Prognostic model construction and evaluation

From the prognostic-related RBPs genes obtained, we formulated a risk signature based on multivariate Cox proportional hazard regression analysis on the entire TCGA data (training set). The risk score of each sample in this model can be calculated by the following formula, which was:

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

The β in this formula refers to the regression coefficient, and *Exp* indicates the gene expression value. Moreover, the model built by the "survival" package can estimate the performance of the prognostic model by the area under the (AUC) of the receiver operating characteristic (ROC) curve with optimal cut-off value. After the construction of the prognostic model, the entire TCGA data were randomly divided into two subsets for cross-validation. Finally, we also used the "rms" package to estimate the probability of overall survival (OS) occurrence and also drew a nomogram to predict 3- and 5-year OS for THCA patients. The calibration curve of the nomogram compared the predicted OS with the observed OS.

2.6 Analysis of key RBPs

To further identify the independent prognostic parameters in the model, we obtained the immunohistochemical results of thyroid cancer through the human protein atlas (HPA;

cBioportal for Cancer Genomics was explored to investigate the genetic alterations of the prognostic genes in the gene signature. Besides, we use the data information in the TIMER database (https://cistrome.shinyapps.io/timer/) [56, 57], the correlation between RBP and tumor immune infiltration (including CD8+ T cells, neutrophils, dendritic cells, B cells, CD4+ T cells, and macrophages) was studied in depth.

2.7 Statistical analysis

In this study, we used R software (version 3.6.3) to explore the whole process. We used the R software packages "ggsignif", "ggpubr" and "ggplot2" to make box plots and quantitative statistical studies of differential expression. In R, we performed multivariate and univariate Cox analysis. The ROC analysis was calculated by the "survivalROC" R package. $p < 0.05$ and considered statistically significant.

3. Result

3.1 Identification of differentially expressed RBPs and functional enrichment analysis in THCA

The intersection of the collected 1542 RBPs data sets and THCA data sets shows that there were 1492 RBPs related genes in THCA (Fig 1A). Then we did a difference analysis of these 1492 THCA RBPs between adjacent normal tissues vs thyroid cancer samples and 162 genes were identified with significant differences (S1 Table). Among them were 70 down-regulated genes and 92 up-regulated genes as displayed in Fig 1B. To assess the molecular mechanism and potential role of RBPs, we divided them into two parts according to their expression ability, and then by using R software we accomplished GO and KEGG enrichment analysis on these two groups, respectively (Fig 1C and 1D). Mostly, RBPs work through multiple RNA-related functions in this result (the list is shown in S2 Table).

3.2 PPI network construction and key gene screening

To better understand these DEGs and seek for the potential key RBPs as biomarker candidates in the development of thyroid cancer, we constructed a PPI network among DEGs through the online website STRING, and visualized it through Cytoscape software. As depicted in Fig 2A, this PPI network has a total of 408 edges and 133 nodes. Also, we calculated the number of interactions between each node and visualized the first 30 nodes as candidate key RBPs for further examination, as presented in Fig 2B. To identify the genes related to the prognosis of thyroid cancer, we utilized univariate Cox regression analysis and found that *NOLC1* ($p = 0.036$), *RPS27L* ($p = 0.011$), *TDRD9* ($p = 0.016$), *TDRD6* ($p = 0.002$), *IFIT2* ($p = 0.037$), and *IFIT3* ($p = 0.02$) were prominently involved in the prognosis of thyroid cancer (Fig 2C; Table 1). Additionally, we determined the prognostic ability of these 6 RBP-coding genes through the Kaplan-Meier plotter (Fig 3). Among them, there was no data for *RPS27L* at the Kaplan-Meier plotter database. These outcomes revealed that these 6 genes were significantly related to prognosis.

3.3 Prognosis-related risk score model construction and validation

Through Multivariate Cox analysis on the entire TCGA training set, we identified 2 RBP-coding genes (*RPS27L* and *IFIT3*) to construct a predictive model. To predict the model and to carry out survival analysis, we divided THCA patients into two groups by using an optimal cut-off value of risk scores in the training set, namely the high-risk group and low-risk group. The Kaplan–Meier survival curve demonstrated that the high-risk subgroup was significantly
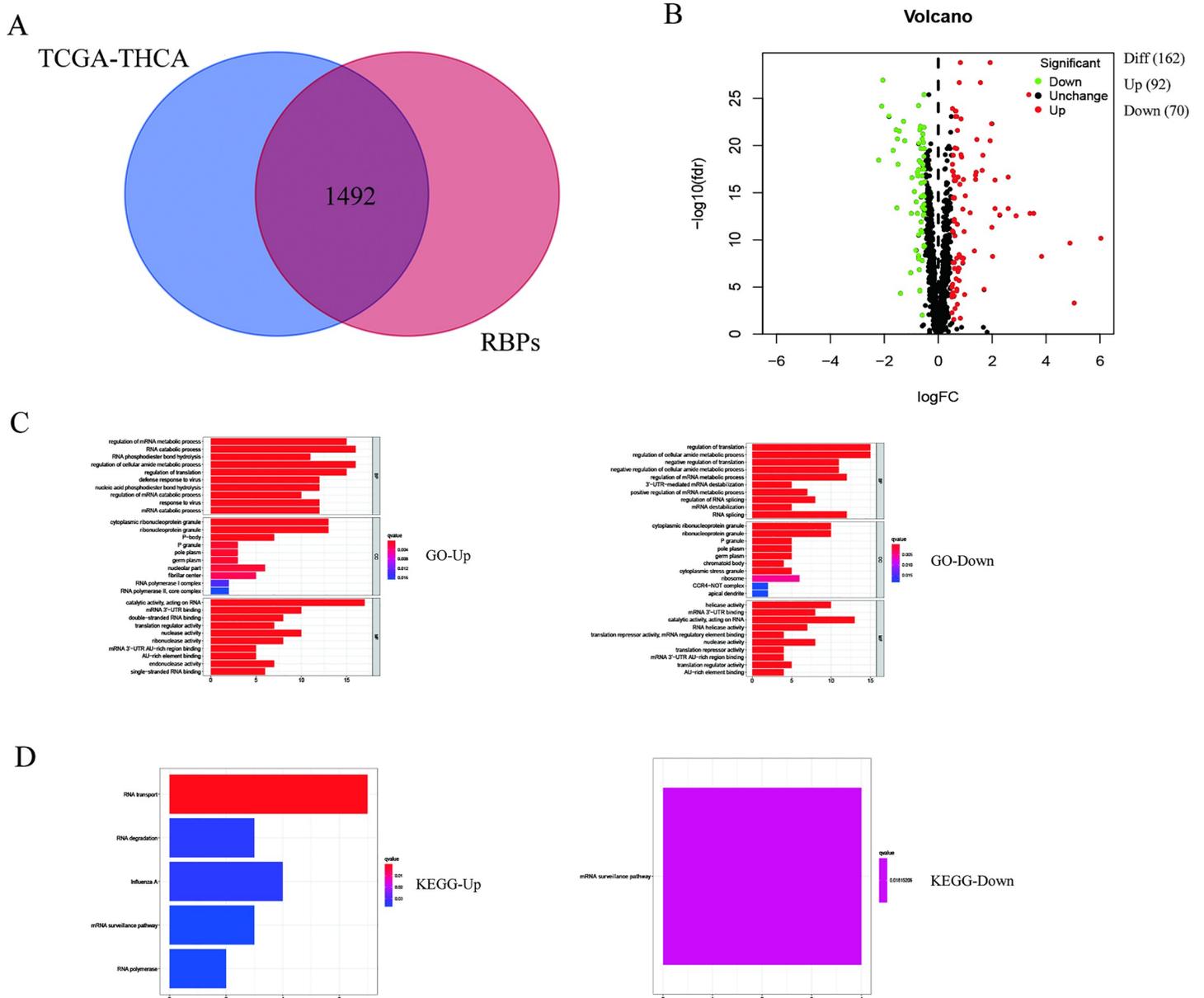
Fig 1. Identification of differentially expressed RBPs-related genes and functional enrichment analysis in THCA. (A) Venn diagrams of RBPs in THCA. Overlapping area means the 1492 genes are in both TCGA-THGA gene list (blue) and RBPs gene list (red); (B) Differentially expressed genes (DEGs) analysis of DEGs; (C) GO enrichment analysis of DEGs; (D) KEGG enrichment analysis of DEGs, p < 0.05.

related to poor survival (p = 0.038; Fig 4A). By completing the time-related ROC investigation, we can more accurately judge the prognostic value of the 2-gene model (AUC = 0.707; Fig 4D). Further, we studied the risk score distribution (Fig 4G), distribution of patients with survival status (Fig 4J) and expression heat map (Fig 4M) with the 2-RBP gene biomarker within the high-risk and low-risk subgroups in the training set.

To test the predictive power of the model, we randomly allocated two subsets from the entire TCGA data and evaluated the risk scores and optimal critical values using the same formula. The high-risk group and the low-risk group in each validation subset were divided
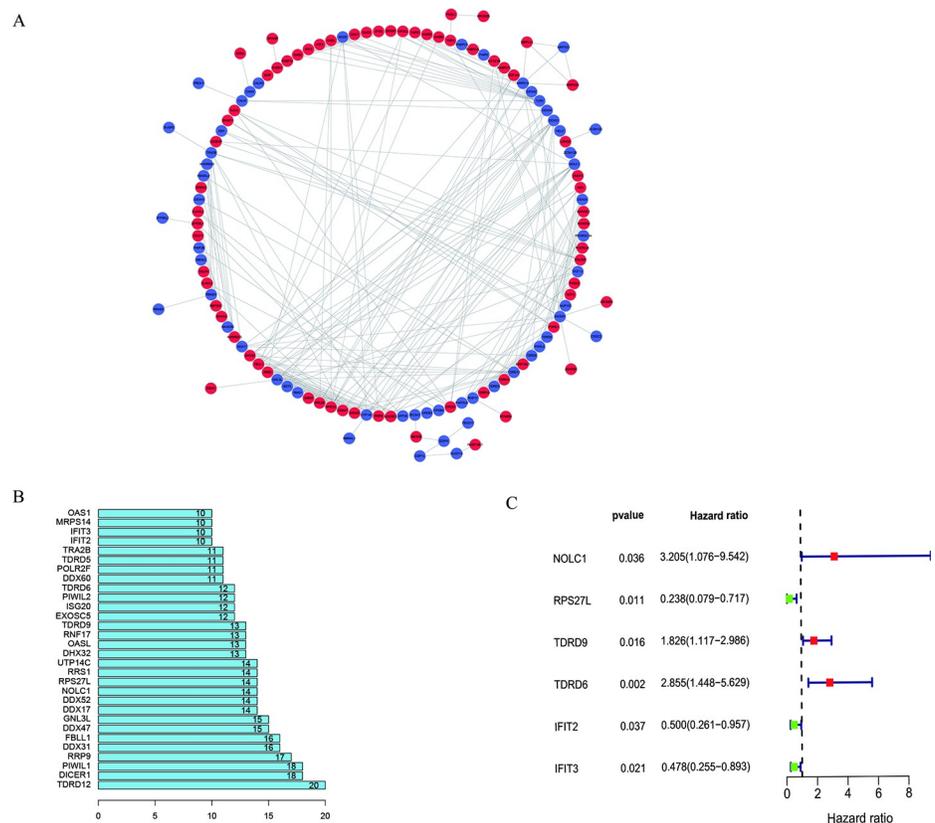
Fig 2. PPI network and key prognostic gene. (A) PPI network for RBPs. blue: down-regulated genes, red: up-regulated genes; (B) Top 30 nodes in PPI network with the number of interactions; (C) Prognostic-related genes in TCHA.

accordingly and high-risk group presented poor prognosis (Fig 4B and 4C). The AUCs in the validation subsets were 0.881 and 0.642, respectively (Fig 4E and 4F). Risk score distribution and survival status distribution are exposed in Fig 4H, 4I, 4K and 4L. The heatmaps disclose the expression of the 2 genes relatively increased along with the rise of risk scores (Fig 4N and 4O). Together, these findings implied that the 2-gene model had a good performance on the prediction of OS in THCA patients.

To construct a quantitative model for THCA prognosis, we combined the two-RBP marker to build a nomogram plot (Fig 5A) for visualization of the performance of the 2-gene prognostic model by the multivariate Cox regression in the entire TCGA data. Calibration plots also revealed that the model had a good performance in estimation of 3-year or 5-year survival of

Table 1. Univariate Cox regression analysis results.

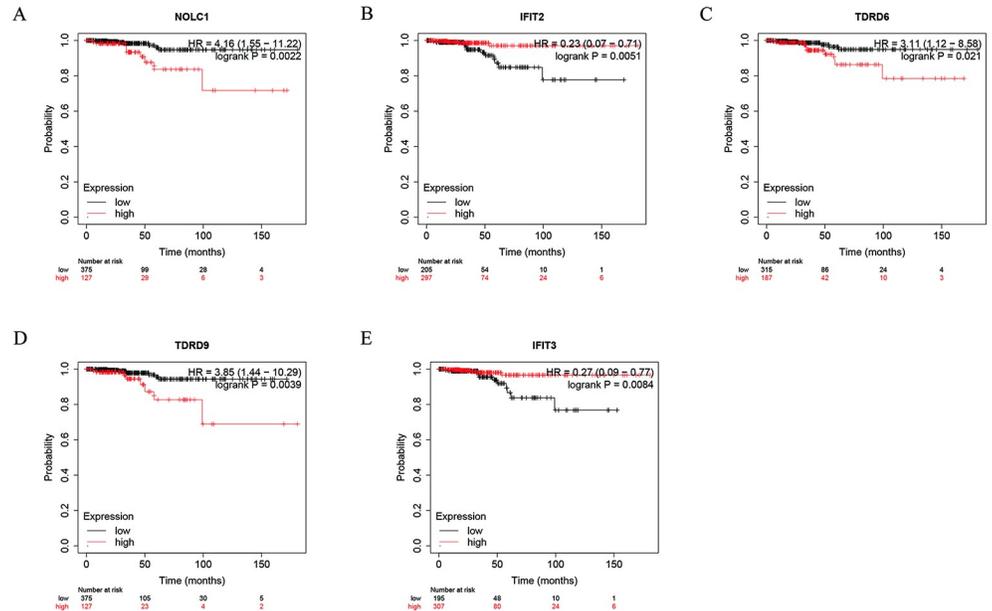| ID	HR	HR.95L	HR.95H	p value
NOLC1	3.20	1.07	9.54	0.036
RPS27L	0.23	0.078	0.71	0.010
TDRD9	1.82	1.11	2.98	0.016
TDRD6	2.85	1.44	5.62	0.0024
IFIT2	0.49	0.26	0.95	0.036
IFIT3	0.47	0.25	0.89	0.020

Fig 3. The Kaplan-Meier plotter analysis. (A) NOLC1; (B) IFIT2; (C) TDRD6; (D) TDRD9; (E) IFIT3.

THCA patients (Fig 5B and 5C) with a harmonious consistency (C-index = 0.67) between the predicted and observed survival.

3.4 Clinical correlation analysis

To assess the correlation between THCA patients and clinics, we found that age ($p < 0.001$), stage ($p = 0.003$), T ($p = 0.032$), and risk score ($p = 0.019$) were all related to prognosis through a univariate Cox analysis (Fig 6A). Multivariate Cox analysis exhibited that age ($p < 0.001$) and riskScore ($p = 0.021$) were related to prognosis (Table 2; Fig 6B). Besides, we used the R software packages "survival" and "survminer" to evaluate, and the results indicated that stage (G3-4), age (> 65), and T (T3-4) were all related to poor prognosis (Fig 7B–7D). Further, we verified and divided THCA into high-risk and low-risk subgroups. The outcomes of the study disclosed that the high-risk was from the male group ($p = 0.011$), and the T3-4 group ($p = 0.046$) was associated with a worse prognosis (Fig 7E and 7L).

3.5 Validate and analyze the RBP-coding genes in the model

We examined the expression of RBP-encoding genes in THCA through R software, and the results displayed that *RPS27L* and *IFIT3* were highly expressed in TCGA-THCA patients (Fig 8A, $p^{***}< 0.001$). It was verified using the HPA database, and the findings exhibited that positive *IFIT3* in THCA was much stronger compared with normal tissues (Fig 8B). However, there was no immunohistochemistry result for RPS27L in the HPA database. Later mutation analysis was carried out on *RPS27L* and *IFIT3* through the cBioPortal online tool, and the results proved that the mutation frequency of both was very low (Fig 8C). Also, through the TIME online tool, we found that *IFIT3* has a positive correlation with Macrophage (r = 0.448, $p^{***}< 0.001$), Neutrophil (r = 0.648, $p^{***}< 0.001$), CD4 + T cell (r = 0.38, $p^{**} < 0.01$), B cell (r = 0.523, $p^{**} < 0.01$), Dendritic cell (r = 0.693, $p^{***} < 0.001$) and CD8 + T cell (r = 0.215, $p^{**} < 0.01$) immune infiltration levels of THCA (Fig 8D). Through TIME online tool, the results indicated that both *RPS27L* and *IFIT3* were highly expressed in tumors (Fig 9).
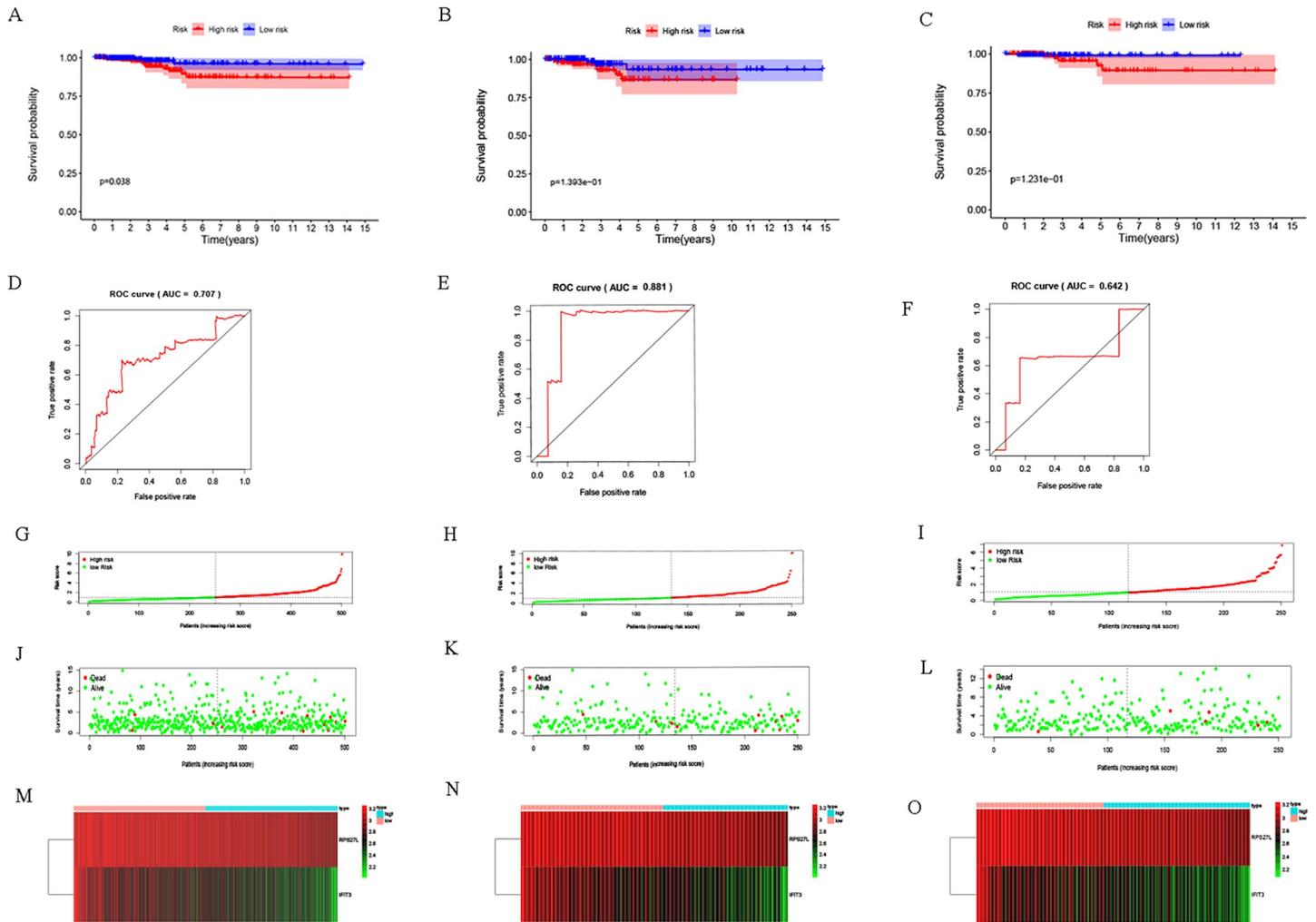
Fig 4. Risk score analysis of the prognostic model. (A) The Kaplan–Meier survival curves in the training set; (B-C) The Kaplan–Meier survival curves in two validation sets; (D) ROC analysis in the training set; (E-F) ROC analysis in the two validation sets; (G) Distribution of risk scores in the training set; (H-I) Distribution of risk scores in the validation sets; (J) Distribution of survival status of patients in different groups in the training set; (K-L) Distribution of survival status of patients in different groups in the two validation sets; (M) Heat map of 2 genes' expression along with risk scores in the training set; (N-O) Heat map of 2 genes' expression along with risk scores in the two validations sets.

https://doi.org/10.1371/journal.pone.0247836.g004

4. Discussion

Thyroid tumors are the most common endocrine tumors. Many researchers reported that abnormal expression of RBPs can cause several illness including cancer. In previous studies, it can be found that the RNA binding proteins PTRF and FNDC3B have been identified as potential prognostic biomarkers for glioblastoma [58]. Three RBP genes (*NOVA1*, *EZH2*, and *RBM24*) were identified as central genes related to the prognosis of head and neck squamous cell carcinoma (HNSCC) [59]. RBP has also been found to play a very important role in lung cancer [60]. Therefore, some therapeutic drugs have been developed to target RBPs to treat some specific cancer [45–47]. Thus, we conducted this study to further explore the role of RBP in thyroid cancer.

We collected 58 adjacent non-tumor tissue samples and 510 THCA tissues from the TGCA database. Extraction of RBP-related genes and analysis of differential expression was carried
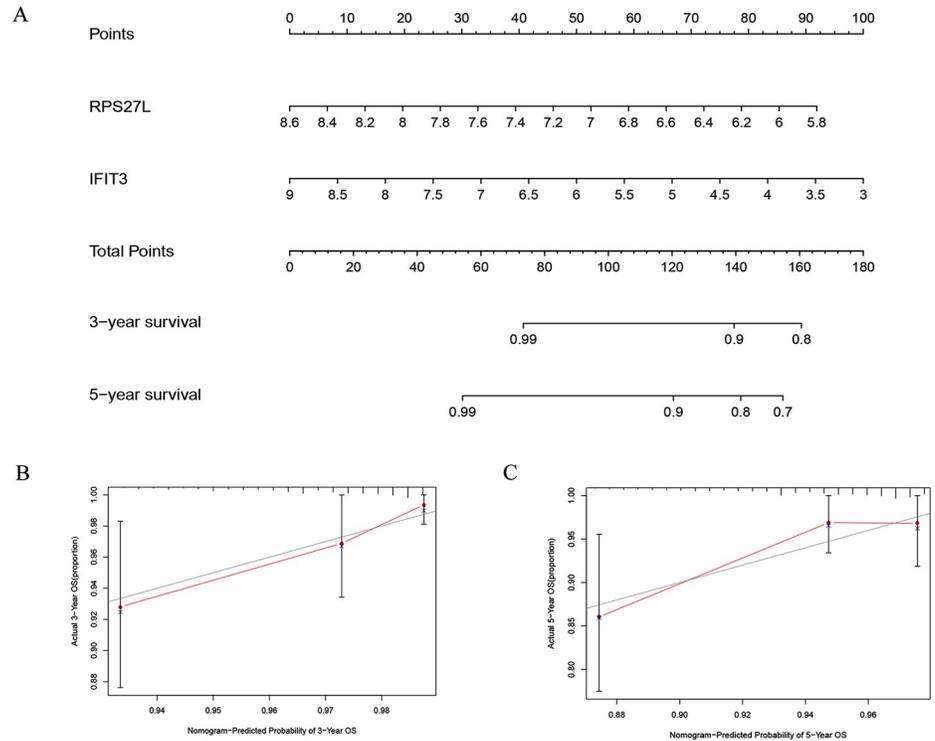
Fig 5. Nomogram and calibration plots of RBPs. (A) Nomogram to predict 3- and 5-year OS in the TCGA cohort; (B, C) Calibration plots of the nomogram to predict OS at 3 and 5 years.

https://doi.org/10.1371/journal.pone.0247836.g005

out, and 162 significantly different RBP-related genes were identified. the data demonstrated that there were 70 down-regulated genes and 92 up-regulated genes. Further we explored the KEGG and GO enrichment investigations for up-regulated genes and noticed that RNA catabolism process, the cellular amide metabolism process, virus defense reaction, cell regulation, amide metabolism process regulation, nucleic acid phosphodiester bond hydrolysis, translation regulation, RNA phosphodiester bond hydrolysis, catabolism regulation process, response to virus, mRNA catabolic process, P-body, ribonucleoprotein granule, cytoplasmic ribonucleoprotein granule, cytoplasmic stress granule, P granule, pole plasm, germplasm, nucleolar part, RNA polymerase II, core complex, fibrillar center, catalytic activity, acting on RNA, nuclease activity, translation regulator activity, double-stranded RNA binding, mRNA 3'-UTR binding,
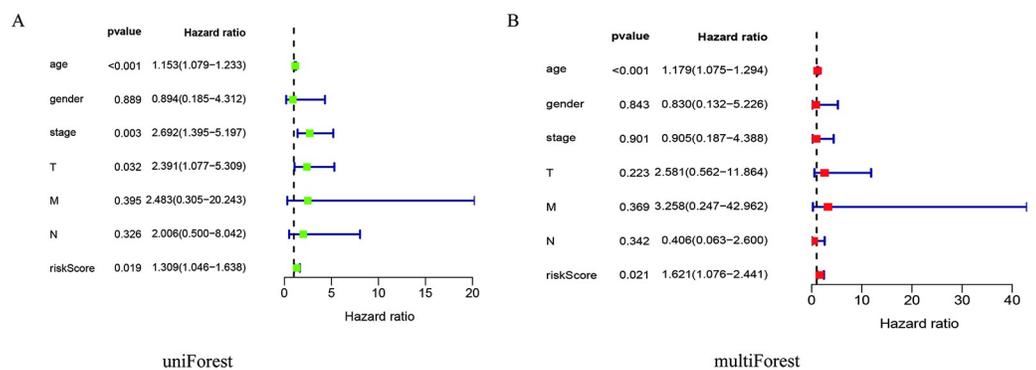

Fig 6. Univariate and multivariate Cox analysis. (A) Univariate Cox analysis; (B) multivariate Cox analysis.

https://doi.org/10.1371/journal.pone.0247836.g006

Table 2. The prognostic effect of different clinical parameters.

	Univariate analysis				Multivariate analysis			
	HR	HR.95L	HR.95H	p value	HR	HR.95L	HR.95H	p value
age	1.15	1.08	1.23	2.81E-05	1.17	1.07	1.29	0.00046
gender	0.89	0.18	4.31	0.88	0.83	0.13	5.22	0.84
stage	2.69	1.39	5.19	0.003	0.90	0.18	4.38	0.90
T	2.39	1.08	5.30	0.032	2.58	0.56	11.86	0.22
M	2.48	0.30	20.24	0.39	3.25	0.24	42.96	0.36
N	2.01	0.50	8.04	0.32	0.40	0.063	2.60	0.34
riskScore	1.31	1.05	1.63	0.019	1.62	1.07	2.44	0.021

ribonuclease activity, mRNA 3'-UTR AU-rich region binding, AU-rich element-binding, exo-nuclease activity, single-stranded RNA binding, RNA transport, RNA degradation, Influenza A, mRNA surveillance pathway, and RNA polymerase were enriched. On the other hand, we performed same enrichment study for down-regulated genes and findings showed that mRNA metabolism process Regulation, cellular amide metabolism process regulation, translation regulation, positive regulation of mRNA metabolism, 3'-UTR-mediated mRNA instability and instability, translation negative regulation, regulation of RNA splicing, mRNA instability,
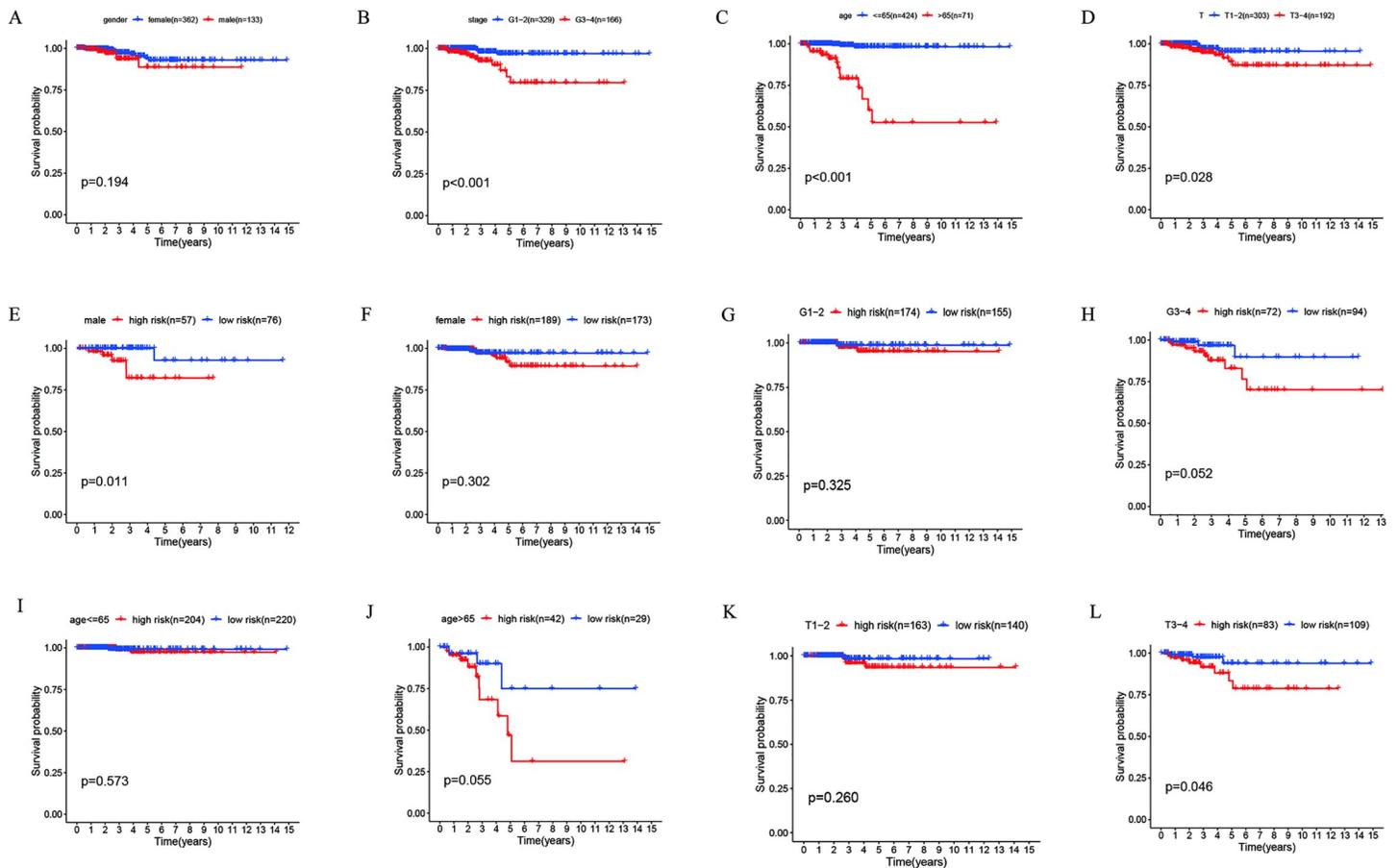

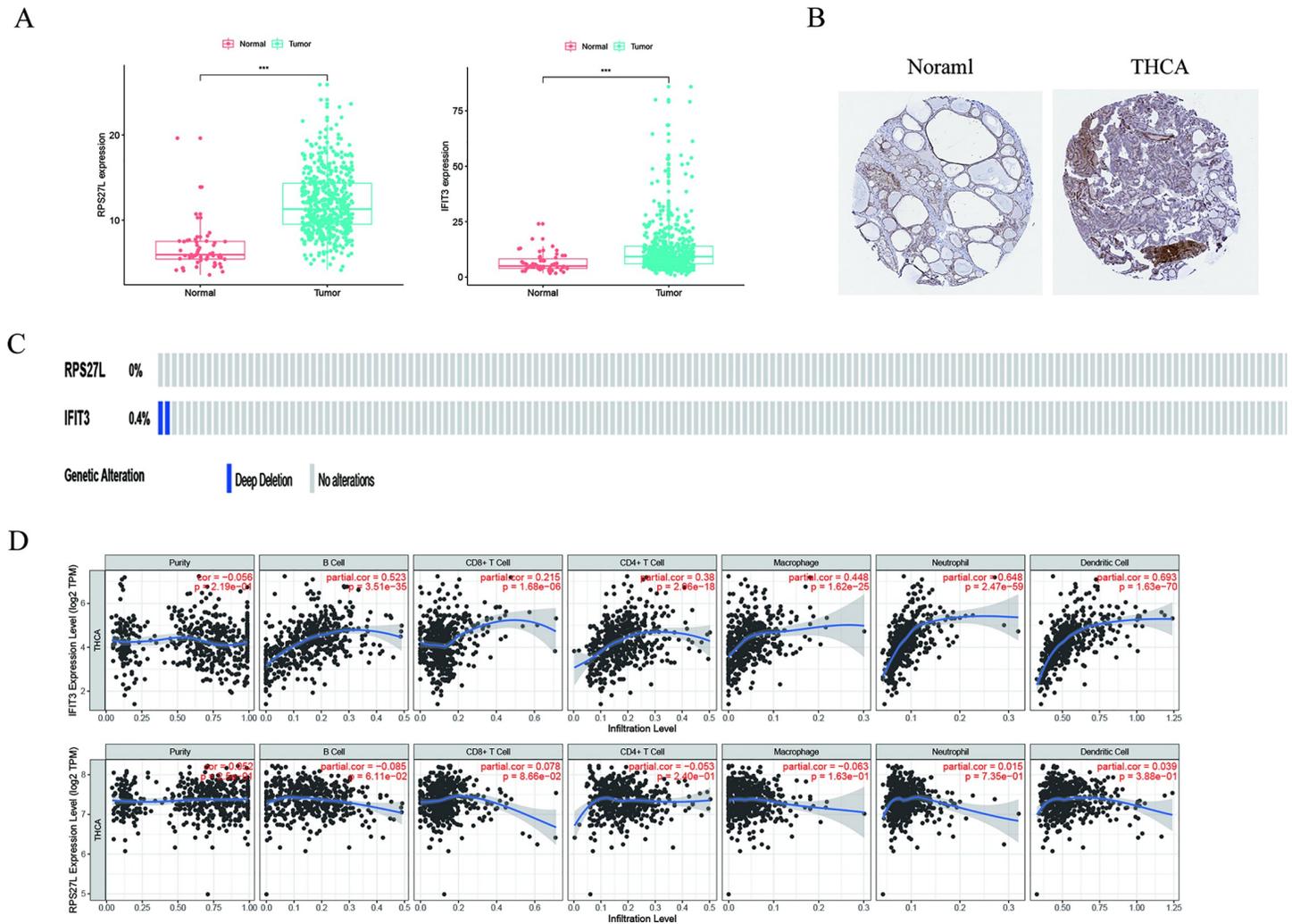**Fig 7. Clinical correlation analysis.**

Fig 8. Validate and analyze the RBP-coding genes. (A) RPS27L and IFIT3 were highly expressed in THCA (TCGA, p***<0.001); (B) IFIT3 was highly expressed in THCA (HPA); (C) RPS27L and IFIT3 mutations were very low; (D) Correlation analysis between gene expression and immune infiltrates.

https://doi.org/10.1371/journal.pone.0247836.g008

cellular amide metabolism process negative regulation, RNA splicing, cytoplasmic ribonucleo-protein granule, ribonucleoprotein granule, P granule, pole plasm, germplasm, chromatoid body, cytoplasmic stress granule ribosome, CCR4-NOT complex, apical dendrite, helicase activity, mRNA 3'-UTR binding, catalytic activity, acting on RNA, RNA helicase activity, translation repressor activity, mRNA regulatory element-binding, AU-rich element-binding, mRNA 3'-UTR AU-rich region binding, translation repressor activity, translation regulator activity, nuclease activity, and mRNA surveillance pathway were enriched. Our results proved that the expectation of enriched pathways of a group of RBPs genes are RNA-related functions. Some RNA-related functions listed on the top might be critical in thyroid cancer development but need further study.

Later, we did PPI network integration to assess those RBPs who are likely to have major impact on the RBPs network in thyroid cancer. The changes in the expression level of those RBPs with more interactions might influence a larger amount of downstream pathways. Next, we conducted univariate Cox regression analysis on these genes and identified 6 genes
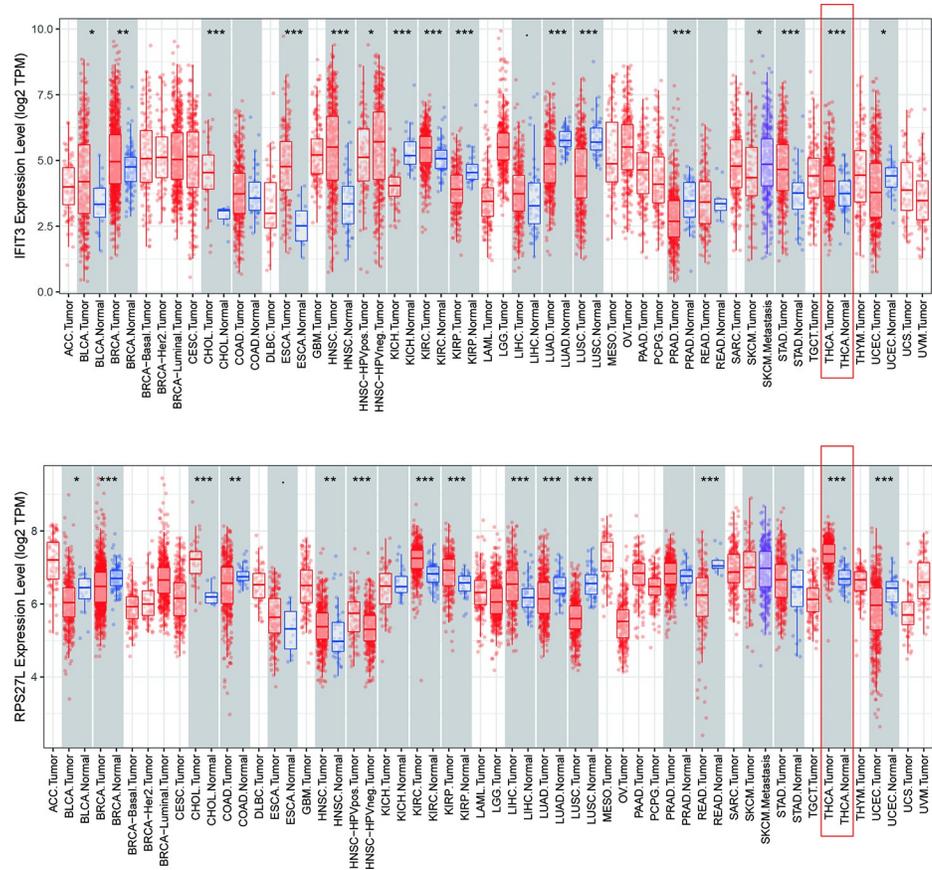
Fig 9. The expression of the two predictive genes in cancers. Data were from the TIMER database. $p^* < 0.05$; $p^{**} < 0.01$; $p^{***} < 0.001$.

https://doi.org/10.1371/journal.pone.0247836.g009

(*NOLC1*, *RPS27L*, *TDRD9*, *TDRD6*, *IFIT2*, and *IFIT3*) involved in prognosis. We further verified these prognostic genes through the online website Kaplan-Meier plotter. Based on their correlation with prognosis, they might work as potential biomarkers or play an important part in the development of thyroid cancer. We have launched a new round of selection for the key RBP related to prognosis. Meanwhile, we build a risk model based on the characteristics of dual RBP genes (*RPS27L* and *IFIT3*) that can predict the prognosis of THCA. Moreover, to begin a new round of prognostic performance of the judging model, we carried out an ROC study on the time dependence of this model with a decent effectivity (AUC = 0.707). To assess the predictive ability of this model, we divided THCA patients into two groups including a high-risk group and a low-risk group and completed survival and clinical correlation examination, respectively. We observed a strong correlation between high-risk subgroups and low survival rates ($p = 0.038$), which suggested that those THCA patients with high risk predicted by the 2-gene model are prone to a worse prognosis. The better performance of this model in the validation subsets supported its comparable prognostic ability. Also, the high-risk subgroups in both the male group ($p = 0.011$) and the T3-4 group ($p = 0.046$) were associated with a worse prognosis. Finally, we explored the genes that make up the model. TCGA data analysis showed that both genes were highly expressed in tumors. HPA database revealed that IFIT3 was highly expressed in THCA. TIMER database results displayed that both *RPS27L* and *IFIT3* were highly expressed in tumors. The mutation investigation outcomes concluded that both

genes were very low mutation frequency, indicating their dysregulation might be mediated by other means. However, the stable mutation status of *RPS27L* and *IFIT3* together with their high expression in tumor tissues makes them potential biomarkers as well as therapeutic targets in thyroid cancer. Furthermore, we found that *IFIT3* has a positive correlation with Macrophage (r = 0.448, $p < 0.001$), Neutrophil (r = 0.648, $p < 0.001$), CD4 + T cell (r = 0.38, $p < 0.01$), B cell (r = 0.523, $p < 0.01$), Dendritic cell (r = 0.693, $p < 0.001$) and CD8 + T cell (r = 0.215, $p < 0.01$) immune infiltration levels of THCA. Although we have not found any report about how *IFIT3* works in cancer, based on the significant association of *IFIT3* and prognosis of thyroid cancer patients, we think its insight should be further studied. These correlations with multiple immune subsets suggest the function of *IFIT3* might be associated with the cancer microenvironment and immune system.

However, we faced some limitations, such as the interesting finding is the individual RBP gene was associated with a better prognosis, while the established model was associated with a worse prognosis. This may be the direction we need to study next such as the lack of verification of the 2-gene predictive model by another independent cohort and the wet experiments of this model for reliability.

Overall, we systematically examined the role of RBPs prognosis of THCA and provided a new perspective on the role of RBPs in the THCA. Mainly, this 2-gene model containing *RPS27L* and *IFIT3* may provide us with great prognostic indicators for the development of THCA. In addition, our study of *IFIT3* implies its great potential in thyroid cancer progressing which has not been reported yet, and these RBPs especially *RPS27L* and *IFIT3* may also be used in clinical adjuvant therapy.

Supporting information

S1 Table. Identification of differentially expressed RBPs.
(DOCX)

S2 Table. GO and KEGG pathway analysis results.
(DOCX)

Author Contributions

Conceptualization: Jing Zhen, Qin Sun.

Data curation: Jing Zhen, JiaCen Li, Qin Sun.

Formal analysis: Jing Zhen, JiaCen Li, Qin Sun.

Methodology: Zhe Song.

Project administration: Zhe Song.

Resources: JiaCen Li.

Writing – original draft: WenJie Su, Qing-Cui Zeng.

Writing – review & editing: WenJie Su, Qing-Cui Zeng.

References

1. Nilsson M, Fagman H. Development of the thyroid gland. *Development*. 2017; 144(12):2123–2140. https://doi.org/10.1242/dev.145615 PMID: 28634271

2. TAYLOR S. Thyroid extract. *Lancet*. 1961; 1(7172):332–333. https://doi.org/10.1016/s0140-6736(61)91499-4 PMID: 13775648

3. LiVolsi VA. The pathology of autoimmune thyroid disease: a review. *Thyroid*. 1994; 4(3):333–339. https://doi.org/10.1089/thy.1994.4.333 PMID: 7833672

4. Pearce EN, Farwell AP, Braverman LE. Thyroiditis [published correction appears in N Engl J Med. 2003 Aug 7;349(6):620]. *N Engl J Med*. 2003; 348(26):2646–2655. https://doi.org/10.1056/NEJMra021194 PMID: 12826640

5. Lamartina L, Grani G, Durante C, Filetti S, Cooper DS. Screening for differentiated thyroid cancer in selected populations. *Lancet Diabetes Endocrinol*. 2020; 8(1):81–88. https://doi.org/10.1016/S2213-8587(19)30324-9 PMID: 31591051

6. Kim J, Gosnell JE, Roman SA. Geographic influences in the global rise of thyroid cancer. *Nat Rev Endocrinol*. 2020; 16(1):17–29. https://doi.org/10.1038/s41574-019-0263-x PMID: 31616074

7. Singh Ospina N, Iñiguez-Ariza NM, Castro MR. Thyroid nodules: diagnostic evaluation based on thyroid cancer risk assessment. *BMJ*. 2020; 368:l6670. Published 2020 Jan 7. https://doi.org/10.1136/bmj.l6670 PMID: 31911452

8. Ancker OV, Krüger M, Wehland M, Infanger M, Grimm D. Multikinase Inhibitor Treatment in Thyroid Cancer. *Int J Mol Sci*. 2019; 21(1):10. Published 2019 Dec 18. https://doi.org/10.3390/ijms21010010 PMID: 31861373

9. Seib CD, Sosa JA. Evolving Understanding of the Epidemiology of Thyroid Cancer. *Endocrinol Metab Clin North Am*. 2019; 48(1):23–35. https://doi.org/10.1016/j.ecl.2018.10.002 PMID: 30717905

10. Cabanillas ME, McFadden DG, Durante C. Thyroid cancer. *Lancet*. 2016; 388(10061):2783–2795. https://doi.org/10.1016/S0140-6736(16)30172-6 PMID: 27240885

11. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*. 2016; 26(1):1–133. https://doi.org/10.1089/thy.2015.0020 PMID: 26462967

12. Schmidbauer B, Menhart K, Hellwig D, Grosse J. Differentiated Thyroid Cancer-Treatment: State of the Art. *Int J Mol Sci*. 2017; 18(6):1292. Published 2017 Jun 17. https://doi.org/10.3390/ijms18061292 PMID: 28629126

13. Fleeman N, Houten R, Bagust A, et al. Lenvatinib and sorafenib for differentiated thyroid cancer after radioactive iodine: a systematic review and economic evaluation. *Health Technol Assess*. 2020; 24 (2):1–180. https://doi.org/10.3310/hta24020 PMID: 31931920

14. Zhao H, Huang T, Li H. Risk factors for skip metastasis and lateral lymph node metastasis of papillary thyroid cancer. *Surgery*. 2019; 166(1):55–60. https://doi.org/10.1016/j.surg.2019.01.025 PMID: 30876667

15. Filetti S, Durante C, Hartl D, et al. Thyroid cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann Oncol*. 2019; 30(12):1856–1883. https://doi.org/10.1093/annonc/mdz400 PMID: 31549998

16. Seethala RR, Baloch ZW, Barletta JA, et al. Noninvasive follicular thyroid neoplasm with papillary-like nuclear features: a review for pathologists. *Mod Pathol*. 2018; 31(1):39–55. https://doi.org/10.1038/modpathol.2017.130 PMID: 29052599

17. Hartl DM, Hadoux J, Guerlain J, et al. Risk-oriented concept of treatment for intrathyroid papillary thyroid cancer. *Best Pract Res Clin Endocrinol Metab*. 2019; 33(4):101281. https://doi.org/10.1016/j.beem.2019.05.005 PMID: 31208873

18. Lewiński A, Adamczewski Z. Papillary thyroid carcinoma: a cancer with an extremely diverse genetic background and prognosis. *Pol Arch Intern Med*. 2017; 127(6):388–389. https://doi.org/10.20452/pamw.4058 PMID: 28680025

19. Na KJ, Choi H. Immune landscape of papillary thyroid cancer and immunotherapeutic implications. *Endocr Relat Cancer*. 2018; 25(5):523–531. https://doi.org/10.1530/ERC-17-0532 PMID: 29507047

20. McLeod DSA, Zhang L, Durante C, Cooper DS. Contemporary Debates in Adult Papillary Thyroid Cancer Management. *Endocr Rev*. 2019; 40(6):1481–1499. https://doi.org/10.1210/er.2019-00085 PMID: 31322698

21. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. *Nat Rev Genet*. 2014; 15 (12):829–845. https://doi.org/10.1038/nrg3813 PMID: 25365966

22. Hentze MW, Castello A, Schwarzl T, Preiss T. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol*. 2018; 19(5):327–341. https://doi.org/10.1038/nrm.2017.130 PMID: 29339797

23. Zhao M, Kim JR, van Bruggen R, Park J. RNA-Binding Proteins in Amyotrophic Lateral Sclerosis. *Mol Cells*. 2018; 41(9):818–829. https://doi.org/10.14348/molcells.2018.0243 PMID: 30157547

24. Chatterji P, Rustgi AK. RNA Binding Proteins in Intestinal Epithelial Biology and Colorectal Cancer. *Trends Mol Med*. 2018; 24(5):490–506. https://doi.org/10.1016/j.molmed.2018.03.008 PMID: 29627433

25. Kapeli K, Martinez FJ, Yeo GW. Genetic mutations in RNA-binding proteins and their roles in ALS. *Hum Genet*. 2017; 136(9):1193–1214. https://doi.org/10.1007/s00439-017-1830-7 PMID: 28762175

26. Pereira B, Billaud M, Almeida R. RNA-Binding Proteins in Cancer: Old Players and New Actors. *Trends Cancer*. 2017; 3(7):506–528. https://doi.org/10.1016/j.trecan.2017.05.003 PMID: 28718405

27. Wang E, Lu SX, Pastore A, et al. Targeting an RNA-Binding Protein Network in Acute Myeloid Leukemia. *Cancer Cell*. 2019; 35(3):369–384.e7. https://doi.org/10.1016/j.ccell.2019.01.010 PMID: 30799057

28. Li T, Li S, Chen D, et al. Transcriptomic analyses of RNA-binding proteins reveal eIF3c promotes cell proliferation in hepatocellular carcinoma. *Cancer Sci*. 2017; 108(5):877–885. https://doi.org/10.1111/cas.13209 PMID: 28231410

29. Correa BR, de Araujo PR, Qiao M, et al. Functional genomics analyses of RNA-binding proteins reveal the splicing regulator SNRPB as an oncogenic candidate in glioblastoma. *Genome Biol*. 2016; 17 (1):125. Published 2016 Jun 10. https://doi.org/10.1186/s13059-016-0990-4 PMID: 27287018

30. Bakkar N, Kovalik T, Lorenzini I, et al. Artificial intelligence in neurodegenerative disease research: use of IBM Watson to identify additional RNA-binding proteins altered in amyotrophic lateral sclerosis. *Acta Neuropathol*. 2018; 135(2):227–247. https://doi.org/10.1007/s00401-017-1785-8 PMID: 29134320

31. Yang C, Kelaini S, Caines R, Margariti A. RBPs Play Important Roles in Vascular Endothelial Dysfunction Under Diabetic Conditions. *Front Physiol*. 2018; 9:1310. Published 2018 Sep 20. https://doi.org/10.3389/fphys.2018.01310 PMID: 30294283

32. Ferrè F, Colantoni A, Helmer-Citterich M. Revealing protein-lncRNA interaction. *Brief Bioinform*. 2016; 17(1):106–116. https://doi.org/10.1093/bib/bbv031 PMID: 26041786

33. Dominguez D, Freese P, Alexis MS, et al. Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol Cell*. 2018; 70(5):854–867.e9. https://doi.org/10.1016/j.molcel.2018.05.001 PMID: 29883606

34. Mohibi S, Chen X, Zhang J. Cancer the'RBP'eutics-RNA-binding proteins as therapeutic targets for cancer. *Pharmacol Ther*. 2019; 203:107390. https://doi.org/10.1016/j.pharmthera.2019.07.001 PMID: 31302171

35. Blech-Hermoni Y, Ladd AN. RNA binding proteins in the regulation of heart development [published correction appears in Int J Biochem Cell Biol. 2014 Oct;55:348]. *Int J Biochem Cell Biol*. 2013; 45 (11):2467–2478. https://doi.org/10.1016/j.biocel.2013.08.008 PMID: 23973289

36. Ignarski M, Rill C, Kaiser RWJ, et al. The RNA-Protein Interactome of Differentiated Kidney Tubular Epithelial Cells. *J Am Soc Nephrol*. 2019; 30(4):564–576. https://doi.org/10.1681/ASN.2018090914 PMID: 30867249

37. Hodson DJ, Screen M, Turner M. RNA-binding proteins in hematopoiesis and hematological malignancy. *Blood*. 2019; 133(22):2365–2373. https://doi.org/10.1182/blood-2018-10-839985 PMID: 30967369

38. Ostareck DH, Ostareck-Lederer A. RNA-Binding Proteins in the Control of LPS-Induced Macrophage Response. *Front Genet*. 2019; 10:31. https://doi.org/10.3389/fgene.2019.00031 PMID: 30778370

39. Wang K, Li L, Fu L, et al. Integrated Bioinformatics Analysis the Function of RNA Binding Proteins (RBPs) and Their Prognostic Value in Breast Cancer. *Front Pharmacol*. 2019; 10:140. https://doi.org/10.3389/fphar.2019.00140 PMID: 30881302

40. Nyati KK, Agarwal RG, Sharma P, Kishimoto T. Arid5a Regulation and the Roles of Arid5a in the Inflammatory Response and Disease. *Front Immunol*. 2019; 10:2790. Published 2019 Dec 5. https://doi.org/10.3389/fimmu.2019.02790 PMID: 31867000

41. Lian S, Li L, Zhou Y, et al. The co-expression networks of differentially expressed RBPs with TFs and LncRNAs related to clinical TNM stages of cancers. PeerJ. 2019; 7:e7696. https://doi.org/10.7717/peerj.7696 PMID: 31576243

42. Schuschel K, Helwig M, Hüttelmaier S, et al. RNA-Binding Proteins in Acute Leukemias. Int J Mol Sci. 2020 May 12; 21(10):3409. https://doi.org/10.3390/ijms21103409 PMID: 32408494

43. Apicco DJ, Zhang C, Maziuk B, et al. Dysregulation of RNA Splicing in Tauopathies. Cell Rep. 2019 Dec 24; 29(13):4377–4388.e4. https://doi.org/10.1016/j.celrep.2019.11.093 PMID: 31875547

44. Cao J, Mu Q, Huang H. The Roles of Insulin-Like Growth Factor 2 mRNA-Binding Protein 2 in Cancer and Cancer Stem Cells. Stem Cells Int. 2018 Mar 15; 2018:4217259. https://doi.org/10.1155/2018/4217259 PMID: 29736175

45. Yasuda K, Mili S. Dysregulated axonal RNA translation in amyotrophic lateral sclerosis. Wiley Interdiscip Rev RNA. 2016 Sep; 7(5):589–603. https://doi.org/10.1002/wrna.1352 Epub 2016 Mar 31. PMID: 27038103

46. Moore MJ, Blachere NE, Fak JJ, et al. ZFP36 RNA-binding proteins restrain T cell activation and antiviral immunity. Elife. 2018 May 31; 7:e33057. https://doi.org/10.7554/eLife.33057 PMID: 29848443

47. Zhu S, Rooney S, Michlewski G. RNA-Targeted Therapies and High-Throughput Screening Methods. Int J Mol Sci. 2020 Apr 23; 21(8):2996. https://doi.org/10.3390/ijms21082996 PMID: 32340368

48. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019 Jan; 47:D607–613. https://doi.org/10.1093/nar/gky1131 PMID: 30476243

49. Szklarczyk D, Morris JH, Cook H, Kuhn M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017 Jan; 45:D362–68. https://doi.org/10.1093/nar/gkw937 PMID: 27924014

50. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015 Jan; 43:D447–52. https://doi.org/10.1093/nar/gku1003 PMID: 25352553

51. Franceschini A, Lin J, von Mering C, Jensen LJ. SVD-phy: improved prediction of protein functional associations through singular value decomposition of phylogenetic profiles. *Bioinformatics*. 2015 Nov; btv696. https://doi.org/10.1093/bioinformatics/btv696 PMID: 26614125

52. Nagy A, Lánczky A, Menyhárt O, Győrffy B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets,*Scientific Reports*, 2018; 8:9227. https://doi.org/10.1038/s41598-018-27521-y PMID: 29907753

53. Uhlen M, Zhang C, Lee S, et al. A pathology atlas of the human cancer transcriptome. *Science*. 2017 357(6352). https://doi.org/10.1126/science.aan2507 PMID: 28818916

54. Uhlén M, Björling E, Agaton C, et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics*. 2005 4(12):1920–32. https://doi.org/10.1074/mcp.M500279-MCP200 PMID: 16127175

55. Uhlén M, Fagerberg L, Hallström BM, et al. Tissue-based map of the human proteome. *Science* 2015 347(6220):1260419. https://doi.org/10.1126/science.1260419 PMID: 25613900

56. Li Taiwen, Fan Jingyu, Wang Binbin, et al. TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Research*.2017; 77(21):e108–e110. https://doi.org/10.1158/0008-5472.CAN-17-0307 PMID: 29092952

57. Li Bo, Severson Eric, Pignon Jean-Christophe, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biology*. 2016; 17(1):174. https://doi.org/10.1186/s13059-016-1028-7 PMID: 27549193

58. Wang Z, Tang W, Yuan J, Qiang B, Han W, Peng X. Integrated Analysis of RNA-Binding Proteins in Glioma. *Cancers (Basel)*. 2020; 12(4):892. Published 2020 Apr 7.

59. Jin Y, Qin X. Comprehensive Analysis of the Roles and Prognostic Value of RNA-Binding Proteins in Head and Neck Squamous Cell Carcinoma [published online ahead of print, 2020 Jul. *DNA Cell Biol*. 2020; https://doi.org/10.1089/dna.2020.5641 PMID: 32716650

60. Li W, Li X, Gao LN, You CG. Integrated Analysis of the Functions and Prognostic Values of RNA Binding Proteins in Lung Squamous Cell Carcinoma. *Front Genet*. 2020; 11:185. https://doi.org/10.3389/fgene.2020.00185 PMID: 32194639