

Development and validation of novel prognostic models based on RNA-binding proteins in breast cancer

Wei Fan^{1,*}, Jun Ding^{2,*}, Shushu Liu¹ and Wei Zhong¹ 

Abstract

Objectives: We aimed to construct novel prognostic models based on RNA-binding proteins (RBPs) in breast cancer (BRCA) and explore their roles in this disease and their effects on tumor-infiltrating immune cells (TIICs).

Methods: Datasets were downloaded from the Gene Expression Omnibus (GEO) database. Functions and prognostic values of RBPs were systematically investigated using a series of bioinformatics analysis methods. TIICs were assessed using CIBERSORT.

Results: Overall, 138 differentially expressed RBPs were identified, of which 86 were upregulated and 52 were downregulated. Of these, 13 RBPs were identified as prognosis-related and adopted to construct an overall survival (OS) model, while 12 RBPs were used for the relapse-free survival (RFS) model. High-risk patients had poorer OS and RFS rates than low-risk patients. The results indicate that the OS and RFS models are good prognostic models with reliable predictive abilities. In addition, the proportions of CD8, CD4 naïve, and CD4 memory resting T cells, as well as resting dendritic cells, were significantly different between the low-risk and high-risk groups in the OS model.

Conclusions: OS and RFS signatures can be used as reliable BRCA prognostic biomarkers. This work will help understand the prognostic roles and functions of RBPs in BRCA.

¹Department of Breast Cancer, Hubei Cancer Hospital; Tongji Medical College, Huazhong University of Science and Technology and Hubei Provincial Clinical Research Center for Breast Cancer, Wuhan, Hubei, China

²Department of Neurology, Wuhan First Hospital, Wuhan, Hubei, China

*These authors contributed equally to this work.

Corresponding author:

Wei Zhong, Department of Breast Cancer, Hubei Cancer Hospital; Tongji Medical College, 116 Quannan Road, Wuhan, Hubei 430079, China.

Email: hustzw168@126.com



Keywords

RNA-binding protein, breast cancer, prognostic model, tumor-infiltrating immune cell, overall survival, Gene Expression Omnibus, bioinformatics

Date received: 28 March 2022; accepted: 19 May 2022

Introduction

Currently, there is no satisfactory treatment for patients with aggressive or relapsed breast cancer (BRCA).^{1,2}

Studies have demonstrated that abnormal expression of RNA-binding proteins (RBPs) is closely associated with human cancer progression.³⁻⁶ Wang et al.⁷ explored the prognostic value and function of RBPs in BRCA, but prognostic models were not developed and validated, nor were the effects on immune-infiltrating cells explored. Thus, in this study, we developed and validated novel prognostic models based on RBPs and explored the relationship between risk models and tumor-infiltrating immune cells (TIICs).

Materials and methods

Data processing

RNA sequencing and clinical datasets for BRCA were downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). The GSE42568 dataset included gene expression data from 17 normal breast tissues and 107 tumor tissues obtained using the Affymetrix Human Genome U133 Plus 2.0 Array platform. In addition, the names of 1542 RBPs were extracted from a previous study.⁸ The differentially expressed genes (DEGs) between tumor tissues and normal breast tissues were identified using the limma package in R software.⁹ DEGs were defined when $P < 0.05$ and $\log_2|\text{fold change}|$ values were >1 . The validation BRCA dataset,

including the transcription profile based on the GPL15048 platform (Rosetta/Merck Human RSTA Custom Affymetrix 2.0) and clinical information, was obtained from the GSE86166 GEO dataset.

Construction and validation of prognostic models

Univariate Cox regression analysis was performed on all differentially expressed RBPs in the GSE42568 dataset using the survival R package. The significance of the candidate genes was verified. Subsequently, a least absolute shrinkage and selection operator (LASSO) regression model was constructed based on the above significant candidate genes, and the risk score (RS) was calculated to assess patient prognosis outcomes. The RS formula for each sample was as follows: $RS = \beta_1 \times \text{Exp gene}_1 + \beta_2 \times \text{Exp gene}_2 + \beta_i \times \text{Exp gene}_i$, where β represents the coefficient value and Exp represents the gene expression level. According to median RS survival analysis, BRCA patients were divided into low- and high-risk groups. The overall survival (OS) and relapse-free survival (RFS) of the two subgroups were compared using log-rank tests. The predictive capability of the prognostic model was evaluated using the survival R package.¹⁰ The GSE86166 dataset was used as a validation cohort to confirm the predictive capability of the prognostic models. Finally, a nomogram was constructed using the R package (www.r-project.org) to forecast the likelihood of OS and RFS.

Gene ontology (GO) enrichment analysis and biology network

DEGs were classified using the clusterProfiler package¹¹ in R software with hypergeometric distributions that were significant when $P < 0.05$. Functional enrichment analyses were performed for GO enrichment. A protein–protein interaction (PPI) network was established via Cytoscape3.7.2 and the STRING 11.5 database (<https://string-db.org/>). Candidate hub genes were identified using the degree topology method.¹² Transcription factor (TF)–gene interaction and microRNA (miRNA)–gene interaction networks of prognosis-related RBPs were established via Cytoscape3.7.2 based on the data from NetworkAnalyst 3.0 (<https://www.networkanalyst.ca/>).

TIIC analysis based on CIBERSORT

The TIICs used in the GSE42568 dataset were assessed using the CIBERSORT analytical tool (<https://cibersort.stanford.edu/>).¹³ The abundance ratio matrix for the 22 immune cell types was determined at $P < 0.05$.

Statistical analysis

R software was used for most bioinformatics and statistical analyses in this study, including RNA-seq data normalization and transformation, CIBERSORT, DEG analysis, survival analyses, receiver operating characteristic (ROC) curve analysis, and Spearman rank correlation analysis. Univariate and multivariate Cox regression analyses were performed using the “coxph” command of the survival package. Statistical significance was set at $P < 0.05$.

Results

Differentially expressed RBPs in BRCA

Differentially expressed RBPs in BRCA were identified using the limma package

for R software. Among the 1542 RBPs (Table S1), 138 genes were differentially expressed, with 86 upregulated and 52 downregulated genes in the GSE42568 dataset (Figure 1, Table S2).

Development and validation of prognostic models

To analyze the effects of RBPs on the prognosis of BRCA patients, we assessed the relationship between the differentially expressed RBPs and OS using univariate Cox regression analysis and the Kaplan–Meier method in the training cohort (GSE42568 dataset). The results showed a significant association between 34 candidate RBPs and OS (Table S3). Similarly, 19 candidate RBPs were significantly associated with RFS (Table S4).

LASSO regression analysis was performed to select reliable RBPs for predicting OS in the GSE42568 dataset (Figure 2a and 2b). Subsequently, we selected the composition of the final gene signature and used it to generate the RS.

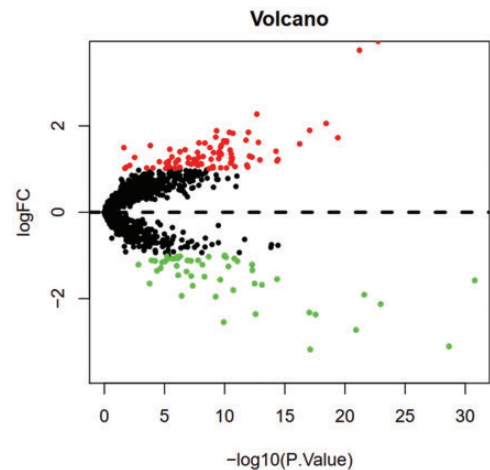


Figure 1. Volcano plot of differentially expressed RNA-binding proteins (RBPs) in the GSE42568 dataset.

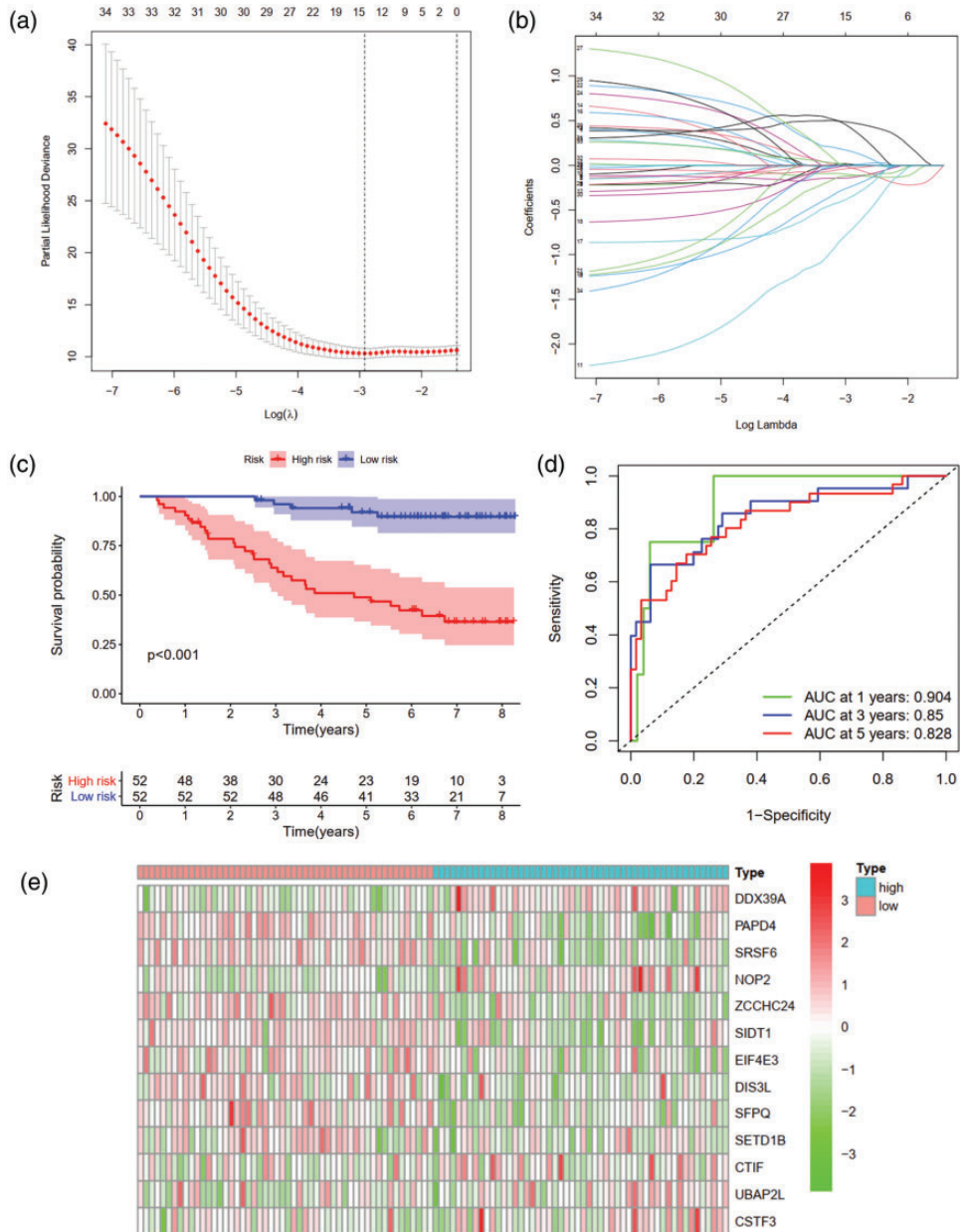


Figure 2. Least absolute shrinkage and selection operator (LASSO) regression analysis based on RNA-binding proteins (RBPs) and overall survival (OS) prediction in the GSE42568 dataset. (a) Confidence interval of each lambda. (b) Changing trajectory of each independent variable. (c) Survival curves for the low- and high-risk subgroups. (d) Receiver operating characteristic (ROC) curves for forecasting OS based on risk score and (e-g) Expression heat map, risk score, and survival status.

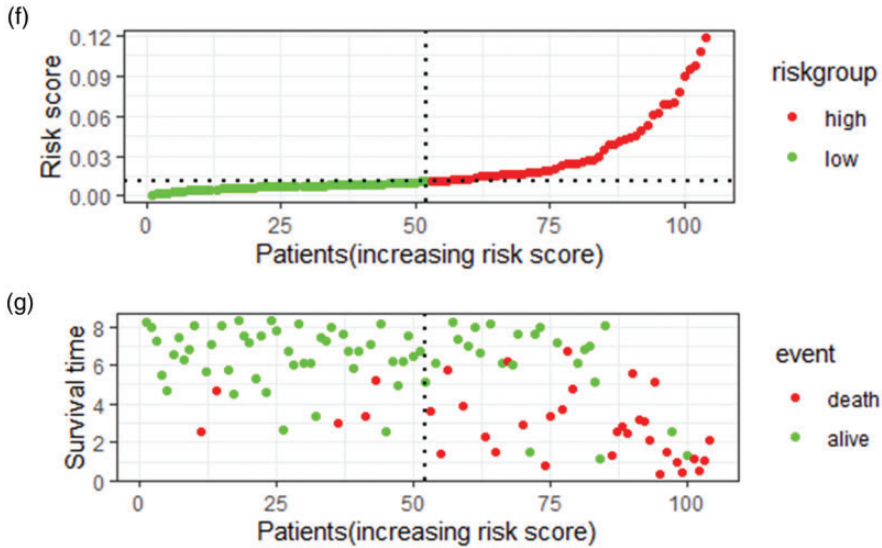


Figure 2. Continued.

$$\begin{aligned}
 \text{RS} = & (0.479 * \text{ExpDDX39A}) \\
 & + (-0.025 * \text{ExpPAPD4}) \\
 & + (-0.104 * \text{ExpSRSF6}) \\
 & + (0.005 * \text{ExpNOP2}) \\
 & + (-0.138 * \text{ExpZCCHC24}) \\
 & + (-0.032 * \text{ExpSIDT1}) \\
 & + (-0.118 * \text{ExpEIF4E3}) \\
 & + (-0.239 * \text{ExpDIS3L}) \\
 & + (-0.732 * \text{ExpSFPQ}) \\
 & + (-0.356 * \text{ExpSETD1B}) \\
 & + (0.012 * \text{ExpCTIF}) \\
 & + (0.158 * \text{ExpUBAP2L}) \\
 & + (0.419 * \text{ExpCSTF3})
 \end{aligned}$$

According to the median RS, 104 patients from the GSE42568 dataset were assigned to the low- and high-risk groups. Survival analysis demonstrated that patients in the high-risk group had significantly poorer OS than those in the low-risk group ($P < 0.001$, Figure 2c). Time-dependent ROC analysis was performed to further evaluate the prognostic capability of the 13 identified RBPs. The area under the ROC curve (AUC) was 0.904, 0.85, and 0.828 at 1-, 3-, and 5-years,

respectively (Figure 2d), indicating a moderate diagnostic performance for this model. The survival status of patients, RS, and expression heat map of the signature consisting of the 13 RBPs in the low- and high-risk subgroups are shown in Figure 2e-g.

To verify the validity of the OS predictive model, the GSE86166 dataset was analyzed. The results showed that patients with a high-risk score had significantly worse OS than those with a low-risk score ($P < 0.001$, Figure 3a). The AUCs of the GSE86166 dataset were 0.552, 0.683, and 0.673 at 1-, 3-, and 5-years, respectively (Figure 3b), which suggests good sensitivity and specificity of the predictive model. The survival status of patients, RS, and the expression heat map of the signature consisting of the 13 RBPs in the GSE86166 dataset are displayed in Figure 3c-e.

Using the same method, 12 gene signatures were selected, and a predictive RFS model was generated for the GSE86166 dataset (Figure 4a and 4b).

$$\text{RS} = (0.498 * \text{ExpDDX39A})$$

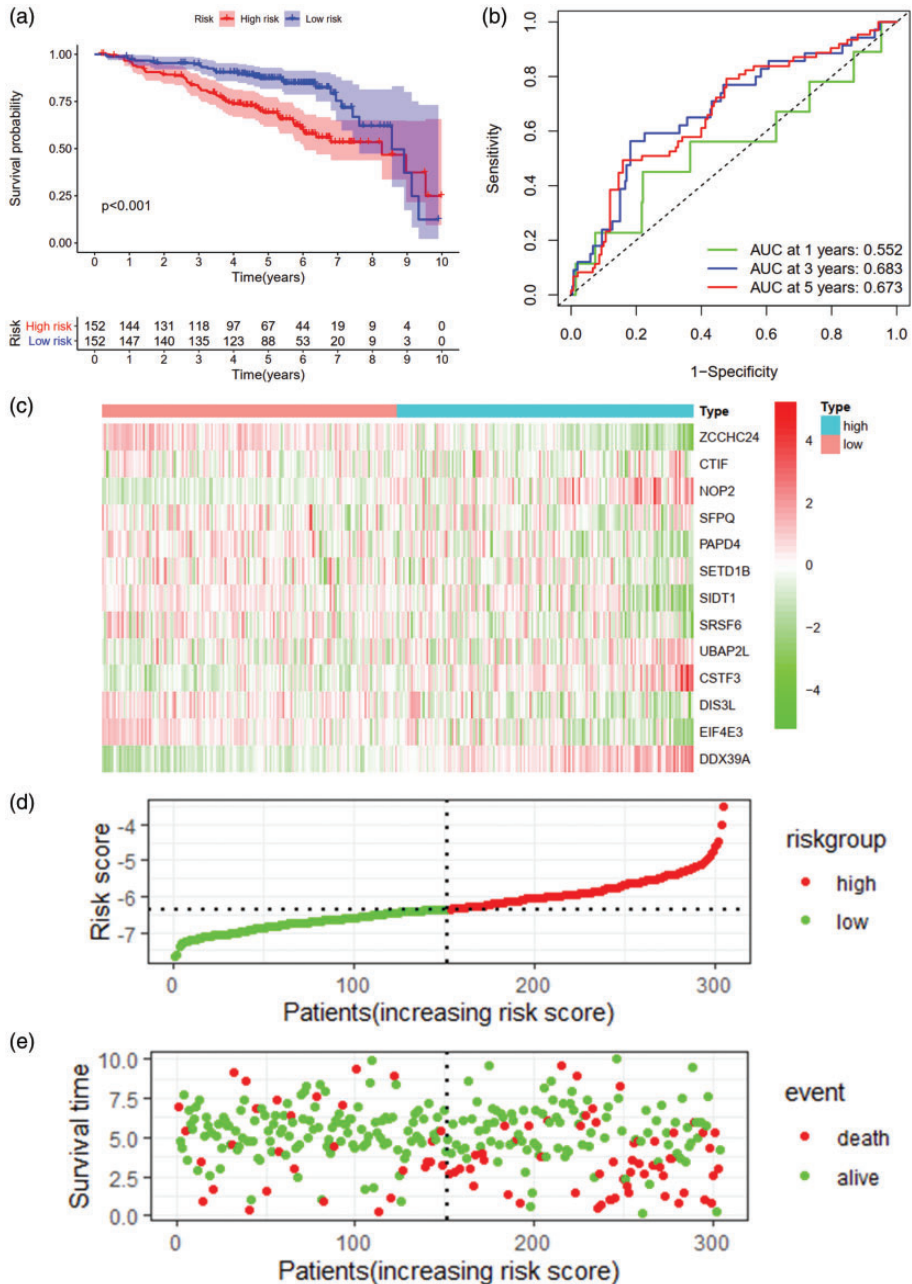


Figure 3. Validation of the overall survival (OS) model using the GSE86166 dataset. (a) Survival curves for the low- and high-risk subgroups. (b) Receiver operating characteristic (ROC) curves for forecasting OS based on risk score and (c–e) Expression heat map, risk score, and survival status.

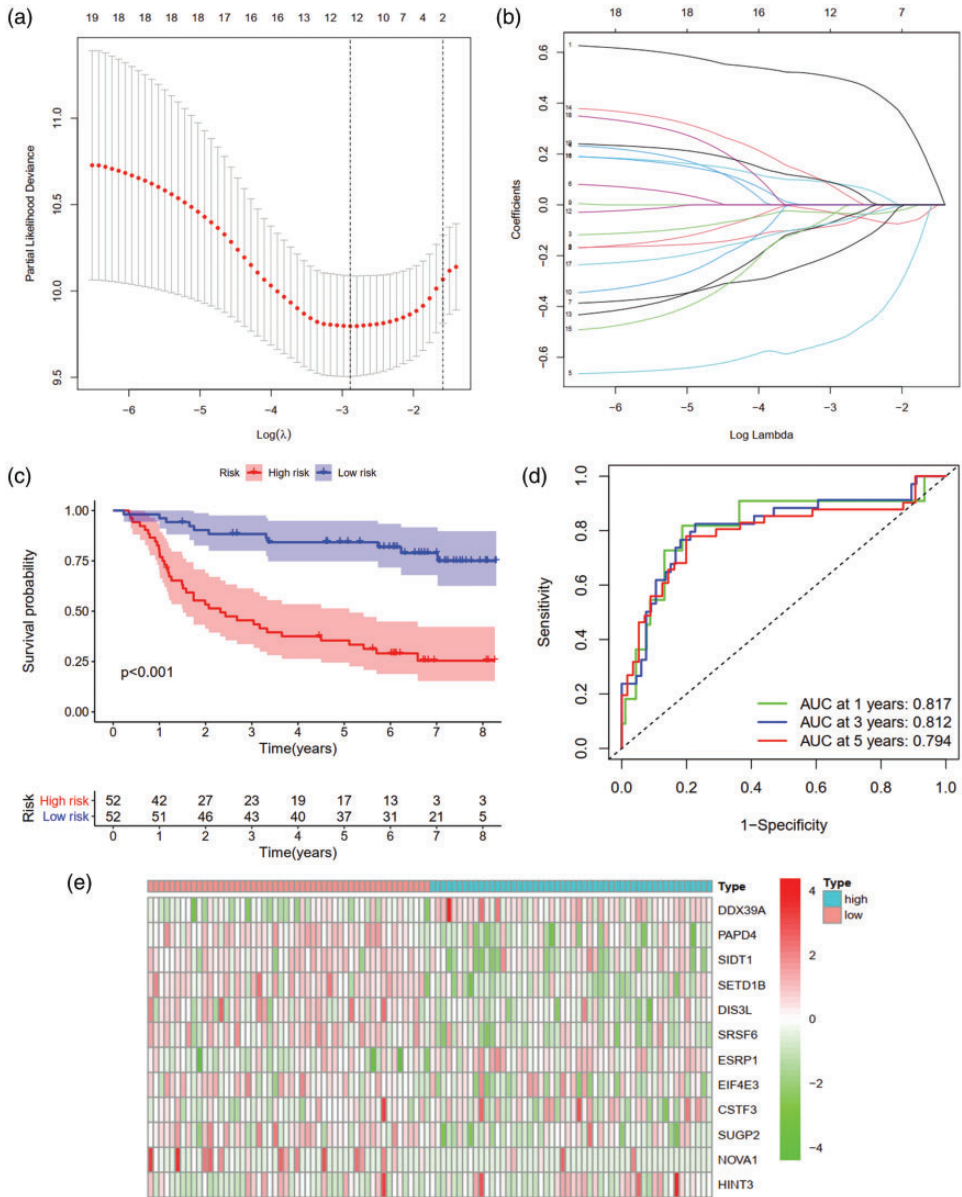


Figure 4. Least absolute shrinkage and selection operator (LASSO) regression analysis based on RNA-binding proteins (RBPs) and relapse-free survival (RFS) prediction from the GSE42568 dataset. (a) Confidence interval of each lambda. (b) Changing trajectory of each independent variable. (c) Survival curves for the low- and high-risk subgroups. (d) Receiver operating characteristic (ROC) curves for forecasting overall survival (OS) based on risk score. (e-g) Expression heatmap, risk score, and survival status.

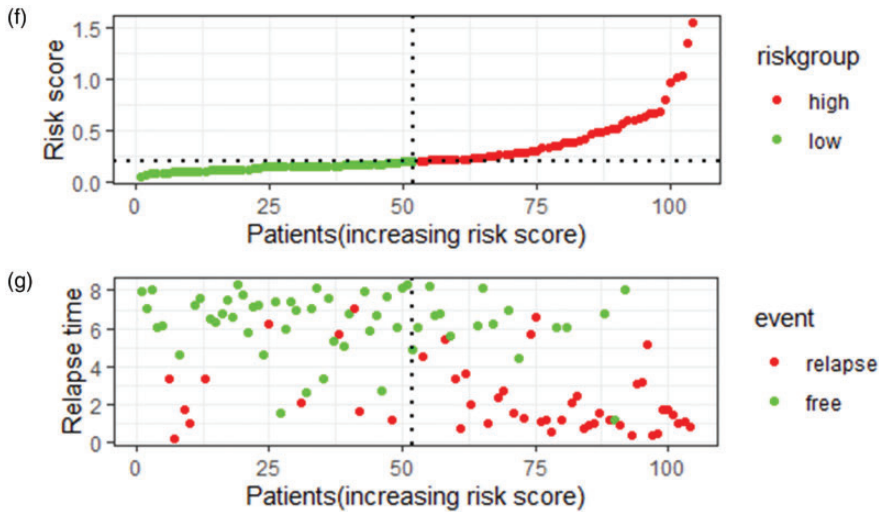


Figure 4. Continued.

$$\begin{aligned}
 &+(-0.034 * \text{ExpPAPD4}) \\
 &+(-0.079 * \text{ExpSRSF6}) \\
 &+(0.083 * \text{ExpESRP1}) \\
 &+(-0.024 * \text{ExpSUGP2}) \\
 &+(-0.033 * \text{ExpSIDT1}) \\
 &+(-0.059 * \text{ExpEIF4E3}) \\
 &+(-0.185 * \text{ExpDIS3L}) \\
 &+(-0.065 * \text{ExpNOVA1}) \\
 &+(-0.529 * \text{ExpSETD1B}) \\
 &+(0.081 * \text{ExpHINT3}) \\
 &+(0.060 * \text{ExpCSTF3})
 \end{aligned}$$

Survival analysis for the GSE42568 dataset demonstrated that patients in the high-risk group had significantly poorer RFS than those in the low-risk group (Figure 4c). The AUCs for the dataset were 0.817, 0.812, and 0.794 at 1-, 3-, and 5-years, respectively (Figure 4d). The relapse status of patients, RS, and expression heat map of the 12 RBPs in the GSE42568 dataset are displayed in Figure 4e–g.

We also verified the validity of the RFS predictive model by analyzing the GSE86166 dataset. Patients with high-risk scores had significantly worse RFS than those with low-risk scores ($P < 0.001$,

Figure 5a). The AUCs of the GSE86166 dataset were 0.735, 0.714, and 0.689 at 1-, 3-, and 5-years, respectively (Figure 5b). The relapse status of patients, RS, and expression heat map of the 12 RBPs in the GSE86166 dataset are displayed in Figure 5c–e.

Additionally, the prognostic significance of different clinicopathological variables was assessed among patients in the GSE42568 dataset using Cox regression analysis. The results demonstrated that RS could independently predict both OS (Table 1) and RFS (Table 2). Additionally, there were similar results in the GSE86166 dataset (Table S5 and S6). Taken together, these data show that the RBP-based OS and RFS models are reliable in predicting the outcomes of BRCA patients.

Building predictive nomograms

Nomograms for OS (Figure 6a) and RFS (Figure 6c) were constructed to generate clinically practical models that would enable physicians to evaluate the prognosis

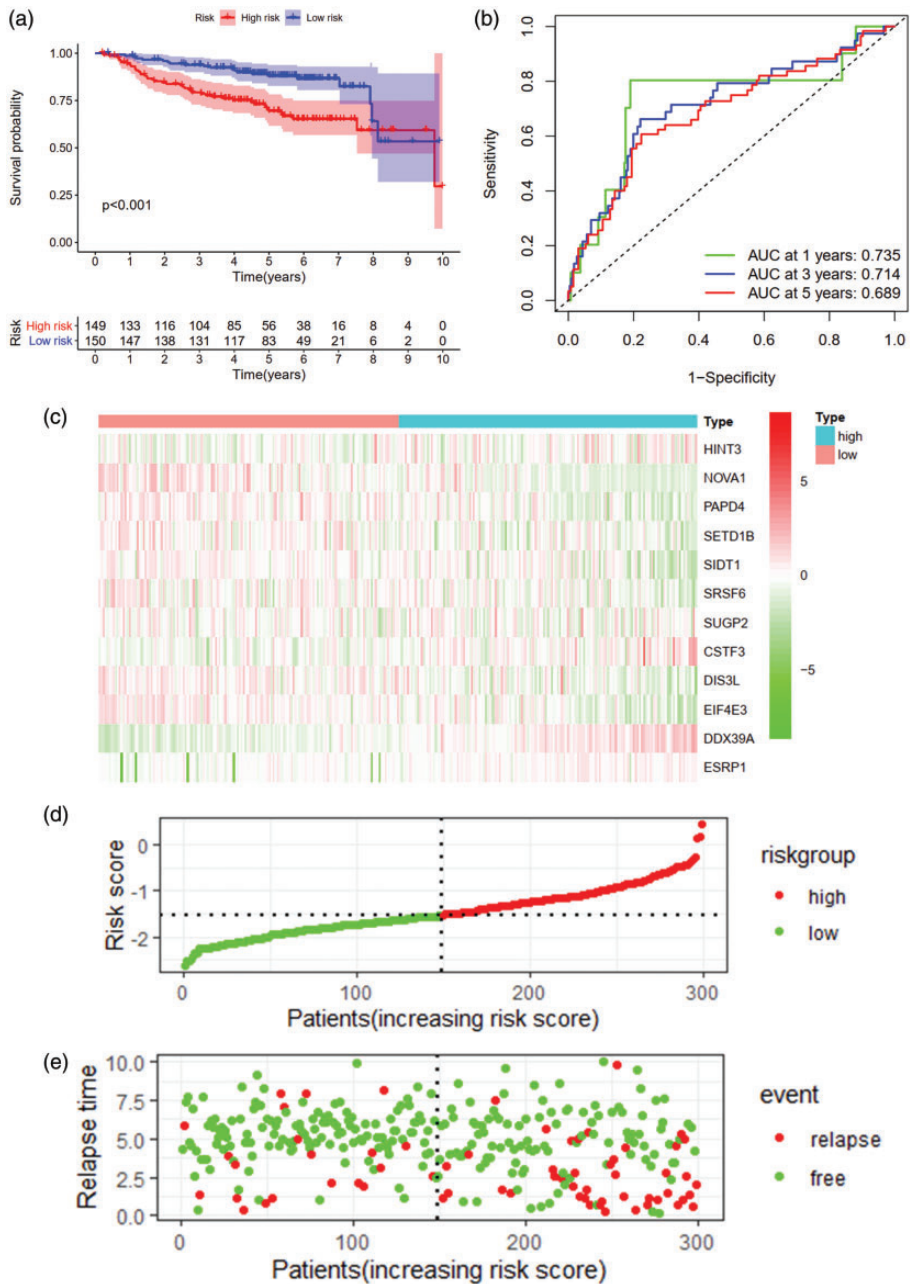


Figure 5. Validation of the relapse-free survival (RFS) model using the GSE86166 dataset. (a) Survival curves for the low- and high-risk subgroups. (b) Receiver operating characteristic (ROC) curves for forecasting overall survival (OS) based on risk score and (c–e) Expression heat map, risk score, and survival status.

Table 1. Cox proportional-hazard regression analysis for overall survival in the GSE42568 dataset.

Characteristic	Univariate analysis			Multivariate analysis		
	HR	HR (95% CI)	P-value	HR	HR (95% CI)	P-value
Age	0.998	0.969–1.028	0.889	1.009	0.975–1.046	0.581
T	1.783	0.942–3.374	0.076	1.083	0.529–2.217	0.828
N	4.555	1.877–11.054	0.001	3.112	1.152–8.411	0.025
Grade	3.038	1.534–6.018	0.001	1.503	0.711–3.179	0.286
ER	0.532	0.268–1.056	0.071	0.647	0.274–1.528	0.321
Risk score	1.77E+18	1.75E+13–1.80E+23	<0.001	4.83E+14	1.74E+09–1.34E+20	<0.001

HR, hazard ratio; CI, confidence interval; ER, estrogen receptor.

Table 2. Cox proportional-hazards regression analysis for relapse free survival in GSE42568.

Characteristic	Univariate analysis			Multivariate analysis		
	HR	HR (95% CI)	P-value	HR	HR (95% CI)	P-value
Age	0.996	0.970–1.022	0.747	1.016	0.986–1.047	0.303
T	2.467	1.306–4.659	0.005	1.543	0.854–2.789	0.151
N	4.549	2.184–9.479	<0.001	3.689	1.656–8.218	0.001
Grade	2.365	1.375–4.069	0.002	1.109	0.607–2.029	0.736
ER	0.438	0.243–0.792	0.006	0.504	0.246–1.034	0.062
Risk score	17.425	7.289–41.655	<0.001	9.241	3.205–26.643	<0.001

HR, hazard ratio; CI, confidence interval; ER, estrogen receptor.

of BRCA patients using the selected RBPs. According to the results of the LASSO regression analysis, each individual variable was assigned a corresponding point based on the scale obtained in the nomograms. The point for each variable was determined by drawing a horizontal line, and then the points of all the variables were added to obtain the patient's total score. This was used to estimate the survival rate of each patient at 1, 3, and 5 years. Furthermore, calibration plots (Figure 6b and 6d) were generated to evaluate the validity and accuracy of the nomograms.

GO enrichment analysis and biology network

To explore the functions of the differentially expressed RBPs and mechanisms through which they can promote BRCA

progression, we performed functional analyses of the downregulated and upregulated RBPs via the clusterProfiler package in R software. As shown in Figure 7a and 7b, significant differences were observed in the functional enrichment of downregulated and upregulated RBPs. Upregulated RBPs were enriched in the spliceosomal complex, nuclear speckles, catalytic step 2 spliceosome, and methyltransferase complex, whereas downregulated RBPs were enriched in ribosomal subunits, organellar ribosomes, mitochondrial ribosomes, mRNA cap binding complexes, and RNA polymerase I complexes. Biological process analysis showed that upregulated RBPs were related to RNA splicing and non-coding RNA processing, while downregulated RBPs were related to RNA splicing, translation, cellular amide metabolic processes, and RNA phosphodiester bond

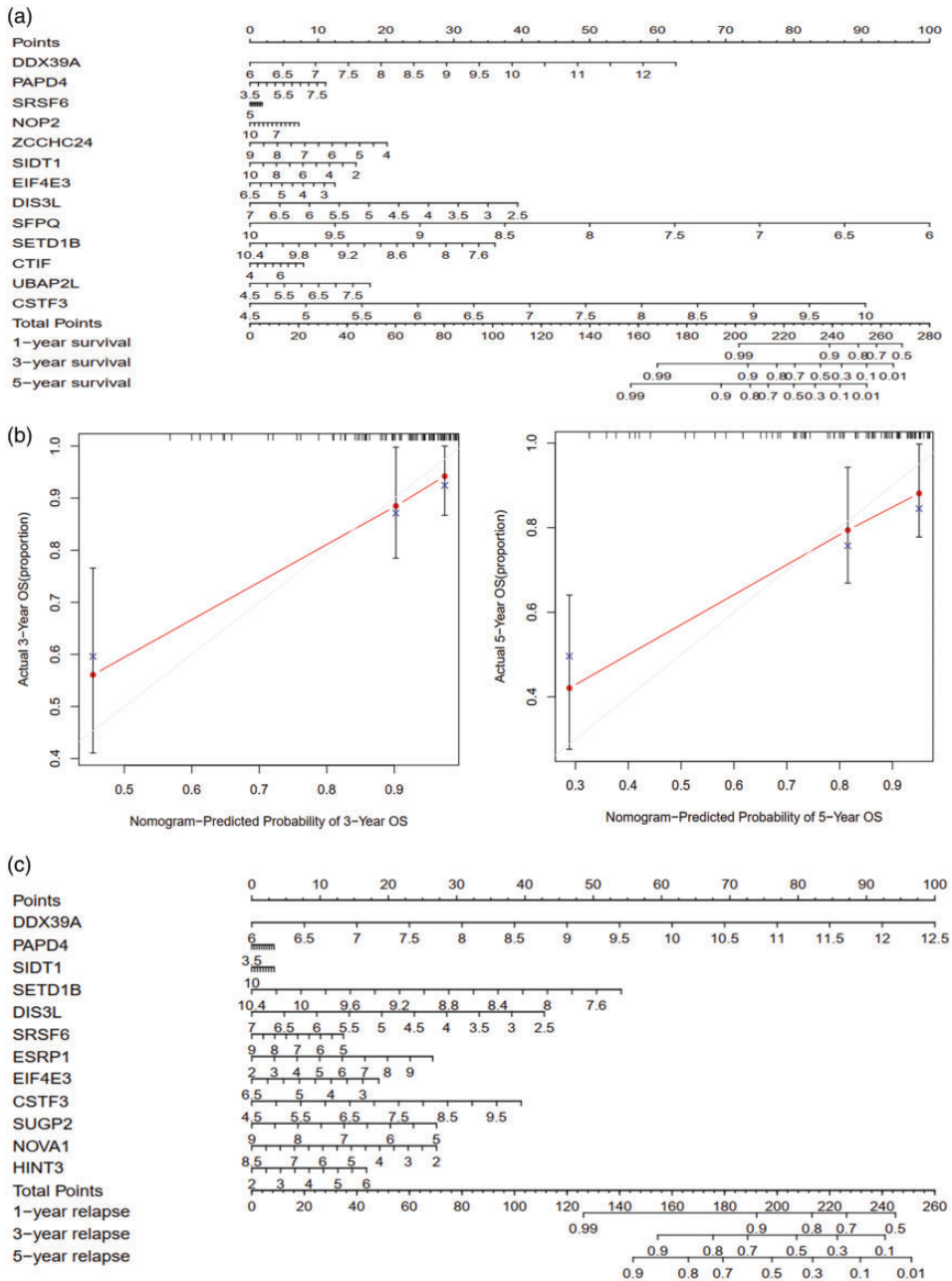


Figure 6. Nomograms for predicting 1-, 3-, and 5-year overall survival (OS) and relapse-free survival (RFS) of breast cancer patients in the GSE42568 dataset. (a) Nomogram for predicting OS. (b) Calibration plots of the nomogram for predicting OS at 3- and 5-years in the dataset. (c) Nomogram for predicting RFS and (d) Calibration plots of the nomogram for predicting RFS at 3- and 5-years in the training cohort.

hydrolysis. Molecular functional analysis demonstrated that upregulated RBPs participated in catalytic, methyltransferase, and transferase activities, as well as single-stranded RNA binding, whereas downregulated RBPs participated in catalytic, endonuclease, nuclease, and ribonuclease activities.

To investigate the functions of the differentially expressed RBPs in BRCA, we created PPI and TF-miRNA coregulatory networks and identified the top 20 hub RBPs (Figure 7c). As shown in Figure 7d, the TF-miRNA co-regulatory network suggested that the prognosis-related RBPs are regulated by many TFs and miRNAs.

TIICs based on CIBERSORT

To explore the impact of the risk score on TIICs, the infiltration proportions of 22 immune cell types in the GSE42568 dataset were calculated using the CIBERSORT algorithm. For the BRCA samples, follicular helper M0, M1, and M2 macrophages were the major constituents of TIICs (Figure S1). Compared with normal breast tissues, the infiltration proportions of follicular helper T cells ($P < 0.001$), M0 macrophages ($P < 0.001$), and M1 macrophages ($P < 0.001$) were significantly increased in BRCA tissues, while the infiltration proportions of activated natural killer (NK) cells ($P = 0.001$), monocytes ($P < 0.001$), M2 macrophages ($P < 0.001$), and resting mast cells ($P = 0.004$) were reduced (Figure 8a). We also analyzed the relationship between TIICs and OS or RFS. Patients with high proportions of follicular helper T cells ($P = 0.001$, Figure 8b) or activated mast cells ($P = 0.015$, Figure 8b) had poorer OS than those with low infiltration proportions, whereas patients with high proportions of resting mast cells had longer OS ($P = 0.017$, Figure 8b). In addition, patients with high infiltration proportions of follicular helper T cells had poorer RFS than

those with low infiltration proportions ($P = 0.034$, Figure 8b).

Compared with the low-risk group designated by the OS model (Figure 8c), the proportions of CD4 naïve T cells ($P = 0.016$), resting CD4 memory T cells ($P = 0.029$), resting dendritic cells ($P = 0.003$), and resting mast cells ($P = 0.048$) were significantly reduced in the high-risk group, while the infiltration proportions of CD8 T cells ($P = 0.028$), monocytes ($P = 0.031$), and activated mast cells ($P = 0.028$) were increased. In addition, the prognosis-related RBPs had different associations with various TIICs (Figure S2). These results suggest that the prognosis-related RBPs interact with some TIICs and influence the progression and prognosis of BRCA.

Discussion

In the present study, we identified 138 differentially expressed RBPs. OS- and RFS-risk models were established based on prognosis-related RBPs that were identified and validated using the GSE42568 and GSE86166 datasets, respectively. Both models proved reliable in predicting OS and RFS. To improve the clinical practicality, we constructed nomograms to evaluate the survival or relapse of patients at 1-, 3-, and 5-years. Additionally, we analyzed the relevant biological functions and networks of the RBPs, and assessed the effects of risk score on TIICs.

To explore the prognostic significance of the differentially expressed RBPs, we performed Cox survival analysis on the GSE42568 dataset, which showed that 13 RBPs (DDX39A, PAPD4, SRSF6, NOP2, ZCCHC24, SIDT1, EIF4E3, DIS3L, SFPQ, SETD1B, CTIF, UBAP2L, and CSTF3) were OS-related, and 12 RBPs (DDX39A, PAPD4, SIDT1, SETD1B, DIS3L, SRSF6, ESRP1, EIF4E3, CSTF3, SUGP2, NOVA1, and HINT3) were RFS-related. Subsequently, we established

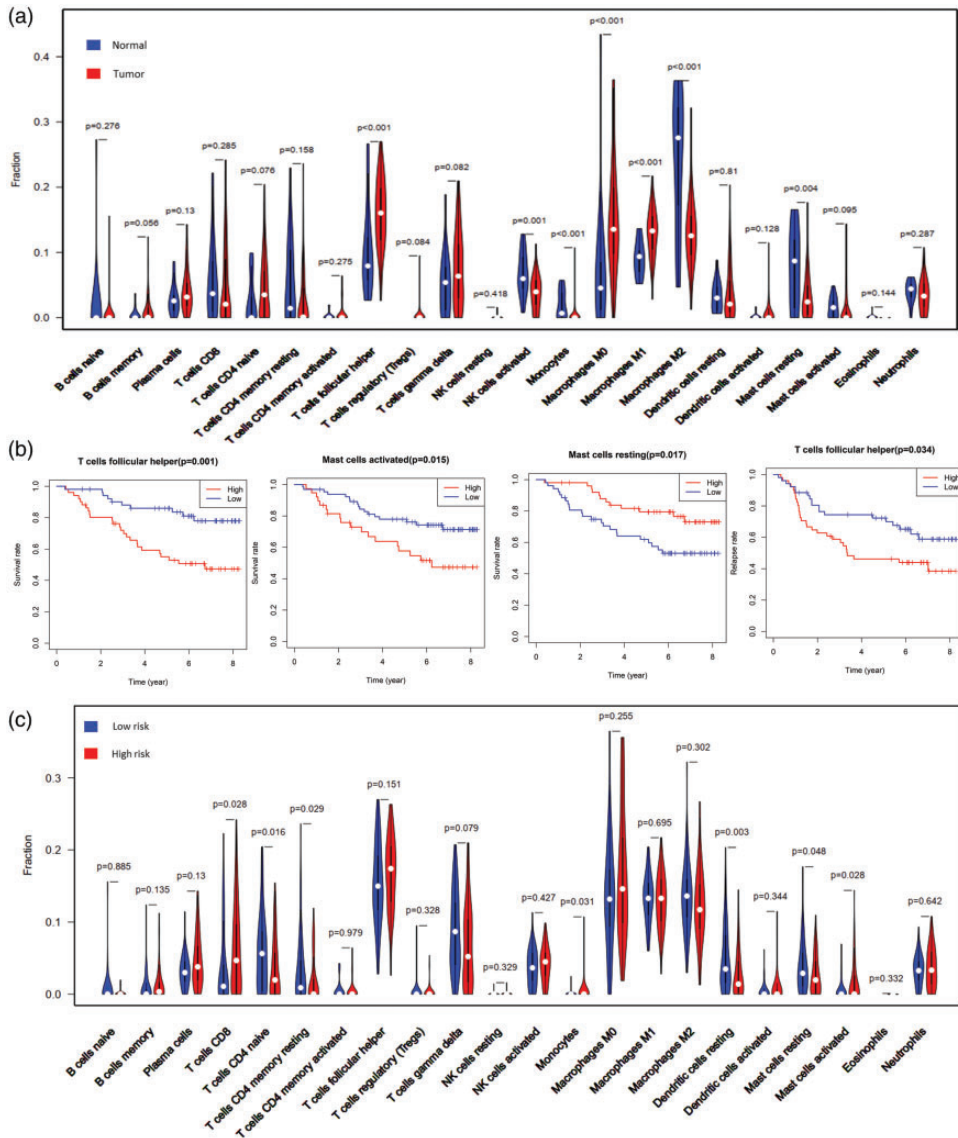


Figure 8. Differences in immune infiltration characteristics between risk groups in the GSE42568 dataset. (a) Quantity of tumor-infiltrating immune cells (TIICs) in normal and breast cancer tissues. (b) Prognostic value of different TIICs and (c) Difference in TIICs between the low- and high-risk groups determined using the CIBERSORT R package.

OS- and RFS-risk models based on the identified RBPs to predict the prognosis of BRCA patients in the GSE42568 dataset. ROC analysis demonstrated that the RBP-based risk models were reliable in selecting

BRCA patients with poor OS and RFS, which was validated by the GSE86166 dataset. Additionally, RS was shown to be an independent prognostic factor in BRCA patients. Furthermore, we constructed

a nomogram to enable physicians to more accurately predict survival or relapse at 1-, 3-, and 5-years.

Functional enrichment analysis demonstrated the significant enrichment of genes involved in the post-transcriptional regulation of RNA, including RNA splicing, regulation of translation, and non-coding RNA processing. Post-transcriptional regulation of RNA plays an important role in many types of cancers.^{14,15} For example, serine and arginine rich splicing factor 1 (SRSF1) has been shown to play an important role in BRCA by regulating alternative splicing,¹⁶ and heterogeneous nuclear ribonucleoprotein A1 (HNRNPA1) affected the progression of triple-negative breast cancer by regulating hypoxia up-regulated 1 (HYOU1) mRNA expression.¹⁷ In addition, N-acetyltransferase 10 (NAT10) regulates cell cycle checkpoint control and resistance to DNA-damaging chemotherapy and radiotherapy by affecting MORC family CW-type zinc finger 2 (MORC2) acetylation in BRCA.¹⁸ Furthermore, interleukin enhancer binding factor 3 (ILF3) promoted breast tumorigenicity by regulating sustained urokinase-type plasminogen activator (uPA) expression,¹⁹ and DEAD-box helicase 17 (DDX17) enhanced the tumorigenic and stem-like features of SRY-box 2 (SOX2) by promoting its binding to its target genes in ER-positive BRCA.²⁰ Consistent with the results of the functional enrichment analysis, the hub RBPs mainly affected the post-transcriptional regulation of RNA. However, the roles and mechanisms of other hub RBPs, such as SRSF6, PRPF19, POLR2E, SNRPB, and SF1, in BRCA remain unclear and require further study. The TF-miRNA co-regulatory network revealed that these prognosis-related RBPs might play important roles in BRCA. Indeed, studies have shown that CSTF3,²¹ SIDT1,⁷ ESRP1,²² DDX39A,²³ NOP2,²⁴ NOVA1,²⁵ SFPQ,²⁶ and UBAP2L²⁷

participate in BRCA development and progression. Although the mechanisms still need to be defined, these studies support our prognostic models. In addition, the roles of other prognosis-related RBPs (CTIF, EIF4E3, SETD1B, ZCCHC24, and SUGP2) in BRCA warrant further exploration.

Finally, we analyzed TIICs in BRCA. The infiltration proportions of follicular helper T cells ranked among the top four TIICs and correlated with OS and RFS. This finding suggests that follicular helper T cells play an important role in BRCA. Furthermore, the proportion of follicular helper T cells was significantly higher in the BRCA tissues than in the normal tissues. A high proportion of follicular helper T cells in BRCA patients was associated with worse OS and RFS when compared with those with a low proportion of follicular helper T cells. This suggests that follicular helper T cells are associated with a poor prognosis and tumor progression in BRCA. However, follicular helper T cells are direct mediators of the immune checkpoint inhibitor response and are associated with B cell activation and antitumor responses in high mutation burden mouse models of BRCA,²⁸ which have not yet been studied in humans. Nevertheless, these studies suggest that follicular helper T cells play an important role in BRCA.

This study has some limitations. First, the sample size of the GSE42568 dataset is small. Second, detailed clinicopathological parameters were lacking in the GSE42568 and GSE86166 datasets. Third, this study was based on public data and needs more basic experiments to further verify these concepts.

Collectively, we investigated the prognostic value and potential function of differentially expressed RBPs in BRCA. Risk models that reliably predict the prognosis of BRCA patients were constructed based on identified RBPs and validated using an

alternative dataset. In addition, the results showed that follicular helper T cells play an important role in BRCA and are positively correlated with some RBPs. Our study provides important evidence for future studies on the role of RBPs in BRCA.

Author contributions

WZ conceptualized and supervised the study. WF and JD performed the bioinformatic and statistical analyses. SL performed data curation and interpretation. WF wrote the original manuscript.

Data availability statement

The data used to support the findings of this study were obtained from the GEO datasets GSE42568 and GSE86166.


Declaration of conflicting interest

The authors declare that there is no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ORCID iD

Wei Zhong  <https://orcid.org/0000-0002-2366-6118>

Supplemental material

Supplemental material for this article is available online.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021; 71: 209–249.
2. Barzaman K, Karami J, Zarei Z, et al. Breast cancer: Biology, biomarkers, and treatments. *Int Immunopharmacol* 2020; 84: 106535.
3. Wang J, Guo Y, Chu H, et al. Multiple functions of the RNA-binding protein HuR in cancer progression, treatment responses and prognosis. *Int J Mol Sci* 2013; 14: 10015–10041.
4. Zhang H, Wang Y, Dou J, et al. Acetylation of AGO2 promotes cancer progression by increasing oncogenic miR-19b biogenesis. *Oncogene* 2019; 38: 1410–1431.
5. Jeong HM, Han J, Lee SH, et al. ESRP1 is overexpressed in ovarian cancer and promotes switching from mesenchymal to epithelial phenotype in ovarian cancer cells. *Oncogenesis* 2017; 6: e391.
6. Li J, Zhou W, Wei J, et al. Prognostic Value and Biological Functions of RNA Binding Proteins in Stomach Adenocarcinoma. *Onco Targets Ther* 2021; 14: 1689–1705.
7. Wang Y, Li H, Ma J, et al. Integrated Bioinformatics Data Analysis Reveals Prognostic Significance Of SIDT1 In Triple-Negative Breast Cancer. *Onco Targets Ther* 2019; 12: 8401–8410.
8. Gerstberger S, Hafner M and Tuschl T. A census of human RNA-binding proteins. *Nat Rev Genet* 2014; 15: 829–845.
9. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; 43: e47.
10. Heagerty PJ, Lumley T and Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000; 56: 337–344.
11. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 2012; 16: 284–287.
12. Chin CH, Chen SH, Wu HH, et al. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* 2014; Suppl 4: S11.
13. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015; 12: 453–457.
14. Masuda K and Kuwano Y. Diverse roles of RNA-binding proteins in cancer traits and their implications in gastrointestinal cancers. *Wiley Interdiscip Rev RNA* 2019; 10: e1520.

15. Shao Y, Chong W, Liu X, et al. Alternative splicing-derived intersectin1-L and intersectin1-S exert opposite function in glioma progression. *Cell Death Dis* 2019; 10: 431.
16. Anczuków O, Akerman M, Cléry A, et al. SRSF1-Regulated Alternative Splicing in Breast Cancer. *Mol Cell* 2015; 60: 105–117.
17. Hao A, Wang Y, Zhang X, et al. Long non-coding antisense RNA HYOU1-AS is essential to human breast cancer development through competitive binding hnRNPA1 to promote HYOU1 expression. *Biochim Biophys Acta Mol Cell Res* 2021; 1868: 118951.
18. Liu HY, Liu YY, Yang F, et al. Acetylation of MORC2 by NAT10 regulates cell-cycle checkpoint control and resistance to DNA-damaging chemotherapy and radiotherapy in breast cancer. *Nucleic Acids Res* 2020; 48: 3638–3656.
19. Hu Q, Lu YY, Noh H, et al. Interleukin enhancer-binding factor 3 promotes breast tumor progression by regulating sustained urokinase-type plasminogen activator expression. *Oncogene* 2013; 32: 3933–3943.
20. Alqahtani H, Gopal K, Gupta N, et al. DDX17 (P72), a Sox2 binding partner, promotes stem-like features conferred by Sox2 in a small cell population in estrogen receptor-positive breast cancer. *Cell Signal* 2016; 28: 42–50.
21. Miles WO, Lembo A, Volorio A, et al. Alternative Polyadenylation in Triple-Negative Breast Tumors Allows NRAS and c-JUN to Bypass PUMILIO Posttranscriptional Regulation. *Cancer Res* 2016; 76: 7231–7241.
22. Gökmen-Polar Y, Neelamraju Y, Goswami CP, et al. Splicing factor ESRP1 controls ER-positive breast cancer by altering metabolic pathways. *EMBO Rep* 2019; 20: e46078.
23. Elango R, Vishnubalaji R, Shaath H, et al. Molecular subtyping and functional validation of TTK, TPX2, UBE2C, and LRP8 in sensitivity of TNBC to paclitaxel. *Mol Ther Methods Clin Dev* 2021; 20: 601–614.
24. Lewinska A, Bednarz D, Adamczyk-Grochala J, et al. Phytochemical-induced nucleolar stress results in the inhibition of breast cancer cell proliferation. *Redox Biol* 2017; 12: 469–482.
25. Tang S, Zhao Y, He X, et al. Identification of NOVA family proteins as novel β -catenin RNA-binding proteins that promote epithelial-mesenchymal transition. *RNA Biol* 2020; 17: 881–891.
26. De Silva HC, Lin MZ, Phillips L, et al. IGFBP-3 interacts with NONO and SFPQ in PARP-dependent DNA damage repair in triple-negative breast cancer. *Cell Mol Life Sci* 2019; 76: 2015–2030.
27. He J, Chen Y, Cai L, et al. UBAP2L silencing inhibits cell proliferation and G2/M phase transition in breast cancer. *Breast Cancer* 2018; 25: 224–232.
28. Hollern DP, Xu N, Thennavan A, et al. B Cells and T Follicular Helper Cells Mediate Response to Checkpoint Inhibitors in High Mutation Burden Mouse Models of Breast Cancer. *Cell* 2019; 179: 1191–1206.