Construction And Verification of A Prognostic Model Related To RNA-Binding Proteins In Uterine Corpus Endometrial Carcinoma

Yukun Jia

The Second School of Clinical Medicine, Southern Medical University; Affiliated Baoan Hospital of Shenzhen, Southern Medical University

Zhan Peng

The Second School of Clinical Medicine, Southern Medical University; Affiliated Baoan Hospital of Shenzhen, Southern Medical University

Guangye Wang (✉ doctorwangtop@163.com)

The Second School of Clinical Medicine, Southern Medical University; Affiliated Baoan Hospital of Shenzhen, Southern Medical University

Research Article

Keywords: RNA binding proteins, uterine corpus endometrial carcinoma, prognostic, survival analysis

Posted Date: September 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-846096/v1>

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

[Read Full License](#)

Abstract

Background: RNA binding proteins (RBP) plays an important role in post-transcriptional regulation. Although the dysregulation of RBP expression is closely related to the occurrence and metastasis of a variety of tumors, there are few reports on RBP in endometrial carcinoma (UCEC). This study aims to establish a RBP-related prognostic model of UCEC.

Methods: We downloaded UCEC gene expression and clinical information data from the Cancer Genome Atlas (TCGA) and GEO database, and determined RBPs that are differentially expressed between tumors and normal tissues. Then, used functional enrichment analysis to analyze the biological functions of the differentially expressed RBP. Used univariate Cox regression analysis to screen prognostic-related RBP and construct a prognostic model. Subsequently, Kaplan-Meier and recipient operating characteristic (ROC) curves were drawn to evaluate the model. Finally, established a nomogram.

Results: This study identified 531 differentially expressed RBPs, including 325 up-regulated and 206 down-regulated RBPs, respectively. Then six independent prognostic-related RBPs (REX02, MARS2, XPO5, YBX1, YBX2, and CELF4) were used to construct a prognostic model. According to this model, the overall survival (OS) of patients in the high-risk score group was significantly lower than that of the low-risk score group. In the training queue and the test queue, the areas under the ROC curve were 0.799 and 0.669, respectively, showing the moderate predictive value of the model.

Conclusion: We have developed and validated the RBP-related prognostic model.

Introduction

Uterine corpus endometrial cancer (UCEC) is the second most common gynecological malignancy after cervical cancer. Every year, about 142,000 women worldwide suffer from UCEC, and an estimated 42,000 women die from this cancer[1]. The 5-year survival rate of patients with early EC is greater than 90%. However, advanced UCEC has the characteristics of strong aggressiveness and high recurrence rate, and the 5-year survival rate of patients is less than 20%[2]. Therefore, the early diagnosis of UCEC is particularly important, which urgently requires new reliable and prospective diagnostic markers and therapeutic targets.

RNA binding protein (RBP) is a protein that is ubiquitously expressed in cells and interacts with various RNAs[3, 4]. Play an important role in post-transcriptional regulation[5]. RBPs participate in almost all post-transcriptional regulation processes, including RNA stability, splicing, translation, localization, transport and degradation. According to research reports, RBP dysregulation is related to the occurrence and metastasis of a variety of tumors. The RBP PCBP1 can inhibit the progression of prostate cancer by inhibiting the expression of mitogen-activated protein kinase 1 [6, 7]. SORBS2 inhibits the metastasis of hepatocellular carcinoma by stabilizing RORA mRNA [8] However, to date, only a small part of RBP has been thoroughly studied and found to have a key role in cancer. And there are few research reports on the

relationship between RBPs and UCEC. Therefore, clarifying the role of RBP in UCEC will help us better understand tumor pathogenesis and develop biomarkers for diagnosis and prognosis.

This study downloads UCEC-related data from TCGA and GEO databases, and screens RBPs that are differentially expressed between tumor tissues and normal tissues. Subsequently, based on these differences RBP, conducted a series of bioinformatics analysis methods, and finally determined 6 independent prognostic-related RBPs, and then constructed a prognostic model. The results of this study may help develop a prognostic assessment model based on RBP of UCEC patients.

Materials And Methods

Data set download

Download the expression data and clinical information of 552 UCSC samples and 35 normal samples from the TCGA database[9]. The GSE17025 data set, 92 UCEC samples expression data and clinical information were downloaded from GEO. Finally, a list of RBPs was obtained from Gerstberger's research[10], and a total of 1489 RBPs were included in our research.

Identify differentially expressed genes

Used limma[package to identify differentially expressed genes, the screening criteria were: $P < 0.05$ and $|\text{Logfold change (FC)}| > 1$. And drew a heat map and a volcano map.

GO and KEGG functional enrichment analysis

GO analysis was used to study the biological functions of differentially expressed genes. KEGG analysis was performed to study the potential biological pathways of differentially expressed gens. All GO[11] and KEGG[12] pathway enrichment analyses used the R packages clusterProfiler[13], enrichplot and ggplot2, and set the thresholds as: $P < 0.05$ h and q (adjusted p-value) < 0.05 .

Construction of PPI network

The differentially expressed RBP genes were submitted in the STRING database[14] to build a PPI network, then Cytoscape 3.7.0 software[15] was used to construct their network. Finally, used the MCODE plug-in to filter out the key modules.

Screening for prognostic-related RBP

Used the "survival" package in R to evaluate the association between survival time and the expression level of differentially expressed RBP through univariate Cox regression analysis. The threshold for screening was $P < 0.05$.

Construction and evaluation of prognostic models

The UCEC patients downloaded by TCGA were randomly stratified into training group and verification group. The multivariate Cox regression was used to construct a prognostic model for UCEC patients. Risk

score = (RBP1 coefficient value × RBP1 gene expression level) + (RBP2 coefficient value × RBP2 gene expression level) + ... + (RBPn coefficient value × RBPn gene expression level). In order to further evaluate the predictive performance of the model, UCEC patients were divided into high and low risk groups according to the median risk score, used the "survival" and "survminer in R to perform survival analysis and drew the survival curve and risk curve, and used the survivalROC" package for ROC curve analysis. Univariate and multivariate Cox regression analysis were used to study whether the risk score is an independent prognostic factor. Finally, the nomogram plot was drawn in R.

RBPs expression verification

Used the Human Protein Atlas (HPA) database [15] to detect the expression of 6 prognostic genes at the translation level.

Result

Differential expression RBP acquisition

The flow chart of the research is shown in Fig. 1. We obtained 531 differentially expressed RBP genes, including 325 up-regulated and 206 down-regulated RBPs. The heat map and volcano map are shown in Fig. 2. The up-regulated differential expression of RBPs were significantly enriched in: "ncRNA metabolic process" for biological processes; "organellar ribosome" for cellular component; and "catalytic activity, acting on RNA" for molecular function. The down-regulated differentially expressed RBPs were significantly enriched in: "RNA splicing" for biological processes; "cytoplasmic ribonucleoprotein granule" for cellular component; and "translation regulator activity" for molecular function. KEGG enrichment showed that the up-regulated RBPs were mainly enriched in the "RNA transport," "Ribosome" and "Ribosome biogenesis in eukaryotes" pathways. The down-regulated RBPs were mainly enriched in the "Ribosome," "Coronavirus disease-COVID-19" and "mRNA surveillance pathway". The results are shown in Fig. 3.

PPI network construction

In order to further explore the molecular mechanism of RBP in the occurrence of UCEC, a PPI network with differentially expressed RBPs genes was constructed. The up-regulated and down-regulated RBPs were shown in red and green circles, respectively. In addition, three key modules were obtained using MCODE software (Fig. 4). We found that the module 1 GO enrichment analysis is mainly enriched in "RNA splicing", "ribosomal", "structural" KEGG: "Ribosome", while in module 2 GO enrichment analysis was mainly enriched in "ribonucleoprotein", "organellar", translation KEGG: "Ribosome" module 3 GO enrichment analysis is mainly enriched in "tRNA 5'-leader removal", "endoribonuclease", "ribonuclease", KEGG: "Ribosome".

Establishment and verification of RBP models related to prognosis

Among 531 differentially expressed RBPs, 29 RBPs related to prognosis were screened by univariate Cox regression analysis. In the training cohort, prognostic model was constructed through multivariate Cox regression analysis, and then the risk score of UCEC patients was calculated according to the following formula: Risk score= $(-0.758882494 \times \text{REX02}) + (0.616466163 \times \text{MARS2}) + (0.870208545 \times \text{XPO5}) + (-0.438901404 \times \text{YBX1}) + (0.382299276 \times \text{YBX2}) + (0.726598644 \times \text{CELF4})$. Patients were divided into high-risk groups and low-risk groups based on the median risk score. Compared with low-risk patients, the prognosis of patients in the high risk score group was poor. In addition, the ROC curve shows the good performance of the model. The expression heat map and survival status of 6-RBP gene biomarker patients in the two groups are shown in the Fig. 5–6. These results indicated that our prognostic model has moderate sensitivity and specificity. In addition, we evaluated whether the 6-RBP gene feature prediction model has similar prognostic ability in the validated patient cohort; the results showed that in the GSE17025 cohort, the OS of patients with high risk scores were worse than that of patients with low risk scores, as shown in Fig. 5. Univariate Cox regression analysis showed that age, tumor stage and risk score were related to the prognosis of UCEC. Multivariate Cox regression analysis showed that risk score is an independent prognostic factor for UCEC patients (Fig. 7). In order to predict the prognosis of the patient more intuitively, the risk score was used to construct a nomogram (Fig. 8). In order to further determine the expression of these prognostic genes in UCEC, we used immunohistochemical results from the HPA database to show that REX02, MARS2, XPO5, YBX1, YBX2 and CELF4 have higher expression levels in UCEC compared with normal endometrial tissue (Fig. 9).

Discussion

Numerous studies have reported that the imbalance of the expression of RBPs leads to the progression of many cancers [16–18]. However, there are few research reports on the role of RBP in LUEC and its influence on tumor progression. In this study, we integrated LUEC's TCGA RNA and GSE17025 RNA sequencing data, and determined the differentially expressed RBP between cancer and normal tissues. We systematically studied related biological pathways and constructed PPIs for these RBPs. Then, the key RBP related to the prognosis was screened, and a risk model with 6 RBPs genes characteristics was constructed to predict the prognosis of LUEC.

In this study, GO biological process analysis found that the differentially expressed up-regulated RBPs were mainly concentrated in ncRNA metabolic process, mitochondrial matrix, catalytic activity, acting on RNA. The down-regulated RBPs were mainly concentrated on ribonucleoprotein granule, RNA splicing, catalytic activity, cytoplasmic ribonucleoprotein granule, acting on RNA. NcRNA has long been the focus of cancer research [19]. Abnormal regulation of ncRNA is related to the metastasis of many cancers [20]. RNA splicing is a very important biological process in the human body. RNA splicing can promote the growth and survival of cancer. Abnormal splicing is involved in the pathogenesis and pathogenesis of many cancers. Progress [21, 22]. KEGG enrichment analysis showed that abnormally expressed RBP was enriched in RNA transport, Ribosome and Spliceosome. Ribosomal biogenesis plays an important role in tumor genesis and progression. Ribosomal proteins affect tumor development by regulating the P53

pathway and mRNA translation[23, 24]. Overall, GO and KEGG analysis showed that these RBPs are closely related to the progress of UCEC through the above-mentioned pathways.

Subsequently, a PPI network was constructed based on these abnormally expressed RBPs. We obtained 145 key RBPs from three key modules. It has been reported that many of these key RBPs have a significant impact on tumorigenesis and tumor development. CARS is a cysteyl-tRNA synthetase, which enhances anti-tumor immunity through specific interaction with TLR2/6 of antigen presenting cells [25].

RPL22L1 is a ribosomal protein that can regulate the splicing of mRNA precursors [26]. Studies have found that RPL22L1 is highly expressed in tumor tissues and participates in the proliferation and metastasis of tumor cells[27, 28].

Univariate Cox regression analysis was performed to analyze and screen the central RBP related to prognosis. We finally identified 631 RBP coding genes. Next, through multiple stepwise Cox regression analysis, 6 RBPs were used to construct a risk model to predict the prognosis of LUSC patients. The ROC curve of the prognostic model showed that the 6-RBP gene feature (AUC = 0.799) has a moderate performance in OS. Multivariate Cox regression analysis showed that this risk-sharing model can independently predict the prognosis of UCEC patients. The immunohistochemistry expression levels of 6 RBPs were verified by HPA database. Constructed a nomogram to help clinicians predict the prognosis of UCEC patients. In summary, our prediction model is relatively reliable and could be used to identify UCEC patients with a poor prognosis, which is helpful for the early diagnosis and treatment of UCEC patients.

This study has certain limitations. First, our results are only based on RNA sequencing; in addition, our prognostic model is built on the UCEC data set of the TCGA and GEO databases, and has not been validated by the clinical patient cohort; finally, for the selected RBP The specific role of UCEC should be carried out in vitro experiments and animal experiments to reveal the mechanism.

Conclusion

This study screened RBPs related to the prognosis of UCEC and established a prognostic model of UCEC. Survival analysis and independent prognostic analysis results showed that this model can be used as an independent predictor of UCEC prognosis. By referring to the literature, it was found that this is the first UCEC prognostic model related to RBP. This study is helpful to further explore the mechanism of UCEC tumor occurrence and development.

Declarations

Ethics approval and consent to participate

NA

Consent for publication

Not applicable

Data Availability Statement

Publicly available datasets were analyzed in this study. This data can be found here: TCGA data (<https://portal.gdc.cancer.gov>) and GEO data(<https://www.ncbi.nlm.nih.gov/gds>).

Conflicts of interest

The authors declare that they have no competing interests

Funding

The authors report no involvement in the research by the sponsor that could have influenced the outcome of this work.

Authors' contributions

Y J: Formal analysis, Data curation, Writing- Original draft preparation, Writing- Reviewing and Editing.

Z P: Resources, Supervision, Writing- Reviewing and Editing.

G W: Conceptualization, Methodology, Project administration.

Acknowledgments

We would like to thank the entire staff of the Cancer Genome Atlas.

References

1. Amant F, Moerman P, Neven P, Timmerman D, Van Limbergen E, Vergote I: **Endometrial cancer**. *Lancet (London, England)* 2005, **366**(9484):491–505.
2. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2018**. *CA: a cancer journal for clinicians* 2018, **68**(1):7–30.
3. New J, Subramaniam D, Ramalingam S, Enders J, Sayed AAA, Ponnurangam S, Standing D, Ramamoorthy P, O'Neil M, Dixon DA *et al*: **Pleotropic role of RNA binding protein CELF2 in autophagy induction**. *Molecular carcinogenesis* 2019, **58**(8):1400–1409.
4. Otsuka H, Fukao A, Funakami Y, Duncan KE, Fujiwara T: **Emerging Evidence of Translational Control by AU-Rich Element-Binding Proteins**. *Frontiers in genetics* 2019, **10**:332.
5. Wurth L, Gebauer F: **RNA-binding proteins, multifaceted translational regulators in cancer**. *Biochimica et biophysica acta* 2015, **1849**(7):881–886.
6. Hua X, Ge S, Chen J, Zhang L, Tai S, Liang C: **Effects of RNA Binding Proteins on the Prognosis and Malignant Progression in Prostate Cancer**. *Frontiers in genetics* 2020, **11**:591667.

7. Zhang Y, Meng L, Xiao L, Liu R, Li Z, Wang YL: **The RNA-Binding Protein PCBP1 Functions as a Tumor Suppressor in Prostate Cancer by Inhibiting Mitogen Activated Protein Kinase 1.** *Cellular physiology and biochemistry: international journal of experimental cellular physiology, biochemistry, and pharmacology* 2018, **48**(4):1747–1754.
8. Han L, Huang C, Zhang S: **The RNA-binding protein SORBS2 suppresses hepatocellular carcinoma tumorigenesis and metastasis by stabilizing RORA mRNA.** *Liver international: official journal of the International Association for the Study of the Liver* 2019, **39**(11):2190–2203.
9. Tomczak K, Czerwińska P, Wiznerowicz M: **The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge.** *Contemporary oncology (Poznan, Poland)* 2015, **19**(1a):A68-77.
10. Gerstberger S, Hafner M, Tuschl T: **A census of human RNA-binding proteins.** *Nature reviews Genetics* 2014, **15**(12):829–845.
11. **The Gene Ontology (GO) project in 2006.** *Nucleic acids research* 2006, **34**(Database issue):D322-326.
12. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K: **KEGG: new perspectives on genomes, pathways, diseases and drugs.** *Nucleic acids research* 2017, **45**(D1):D353-d361.
13. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing biological themes among gene clusters.** *Omics: a journal of integrative biology* 2012, **16**(5):284–287.
14. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P *et al*: **STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.** *Nucleic acids research* 2019, **47**(D1):D607-d613.
15. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, Benfeitas R, Arif M, Liu Z, Edfors F *et al*: **A pathology atlas of the human cancer transcriptome.** *Science (New York, NY)* 2017, **357**(6352).
16. Masuda K, Kuwano Y: **Diverse roles of RNA-binding proteins in cancer traits and their implications in gastrointestinal cancers.** *Wiley interdisciplinary reviews RNA* 2019, **10**(3):e1520.
17. Soni S, Anand P, Padwad YS: **MAPKAPK2: the master regulator of RNA-binding proteins modulates transcript stability and tumor progression.** *Journal of experimental & clinical cancer research: CR* 2019, **38**(1):121.
18. Velasco MX, Kostı A, Penalva LOF, Hernández G: **The Diverse Roles of RNA-Binding Proteins in Glioma Development.** *Advances in experimental medicine and biology* 2019, **1157**:29–39.
19. Ye B, Shi J, Kang H, Oyebamiji O, Hill D, Yu H, Ness S, Ye F, Ping J, He J *et al*: **Advancing Pan-cancer Gene Expression Survival Analysis by Inclusion of Non-coding RNA.** *RNA biology* 2020, **17**(11):1666–1673.
20. Hajjari M, Salavaty A: **HOTAIR: an oncogenic long non-coding RNA in different cancers.** *Cancer biology & medicine* 2015, **12**(1):1–9.
21. Chen J, Weiss WA: **Alternative splicing in cancer: implications for biology and therapy.** *Oncogene* 2015, **34**(1):1–14.

22. Yang X, Huang WT, He RQ, Ma J, Lin P, Xie ZC, Ma FC, Chen G: **Determining the prognostic significance of alternative splicing events in soft tissue sarcoma using data from The Cancer Genome Atlas.** *Journal of translational medicine* 2019, **17**(1):283.
23. Genuth NR, Barna M: **The Discovery of Ribosome Heterogeneity and Its Implications for Gene Regulation and Organismal Life.** *Molecular cell* 2018, **71**(3):364–374.
24. Goudarzi KM, Lindström MS: **Role of ribosomal protein mutations in tumor development (Review).** *International journal of oncology* 2016, **48**(4):1313–1324.
25. Cho S, Kim SB, Lee Y, Song EC, Kim U, Kim HY, Suh JH, Goughnour PC, Kim Y, Yoon I *et al*: **Endogenous TLR2 ligand embedded in the catalytic region of human cysteinyl-tRNA synthetase 1.** *Journal for immunotherapy of cancer* 2020, **8**(1).
26. Zhang Y, O'Leary MN, Peri S, Wang M, Zha J, Melov S, Kappes DJ, Feng Q, Rhodes J, Amieux PS *et al*: **Ribosomal Proteins Rpl22 and Rpl22l1 Control Morphogenesis by Regulating Pre-mRNA Splicing.** *Cell reports* 2017, **18**(2):545–556.
27. Liang Z, Mou Q, Pan Z, Zhang Q, Gao G, Cao Y, Gao Z, Pan Z, Feng W: **Identification of candidate diagnostic and prognostic biomarkers for human prostate cancer: RPL22L1 and RPS21.** *Medical oncology (Northwood, London, England)* 2019, **36**(6):56.
28. Rao S, Peri S, Hoffmann J, Cai KQ, Harris B, Rhodes M, Connolly DC, Testa JR, Wiest DL: **RPL22L1 induction in colorectal cancer is associated with poor prognosis and 5-FU resistance.** *PloS one* 2019, **14**(10):e0222392.

Figures

Figure 1

The workflow for analyzing the RBPs in UCEC.

Figure 2

RBP differentially expressed in UCEC: (A) Heat map, red represents up-regulation, green represents down-regulation. (B) Volcano map. UCEC, endometrial cancer of the uterus; RBP, RNA binding protein.

Figure 3

Enrichment analysis of differentially expressed RBP. (A), (C) GO analysis of up-regulated RBP and enrichment analysis of KEGG pathway. (B), (D) GO analysis of down-regulated RBP and enrichment analysis of KEGG pathway.

Figure 5

Risk score analysis of the prognostic model. (A), (C) KM curve and ROC analysis of the prognostic model in the TCGA training cohort; (B), (D) KM curve and ROC analysis of the prognostic model in the TCGA validation cohort

Figure 6

(A), (C) and (E) The risk factor correlation diagram of UCEC patients in the TCGA training cohort; (B), (D) and (F) The risk factor correlation diagram of UCEC patients in the TCGA validation cohort.

Figure 7

Cox regression analysis of independent prognostic value. (A) Univariate and (B) Multivariate Cox regression analysis.

Figure 8

The nomogram of OS for UCEC patients at 1, 2, and 3 years.

Figure 9

HPA database UCEC and the expression of 6 prognostic-related RBPs in normal endometrial tissue. (A) REXO2, (B) MARS2, (C) XPO55, (D) YBX1. (E) YBX2, (F) CELF4.