

# Expression Profile of RNA Binding Protein in Cervical Cancer

Zhiyuan Huang

Tongji University Affiliated East Hospital: Shanghai East Hospital

Fang Li (✉ [fang\\_li@tongji.edu.cn](mailto:fang_li@tongji.edu.cn))

Shanghai East Hospital, School of Medicine, Tongji University. <https://orcid.org/0000-0003-0342-3941>

Qinchuan Li

Tongji University Affiliated East Hospital: Shanghai East Hospital

---

## Primary research

**Keywords:** RNA binding proteins, cervical cancer, prognosis, TCGA, GTEX

**Posted Date:** August 2nd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-655697/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** It has been demonstrated by studies globally that RNA binding proteins (RBPs) took part in the development of cervical cancer (CC). Few studies concentrated on the correlation between RBPs and overall survival of CC patients. We retrieved significant DEGs (differently expressed genes, RNA binding proteins) correlated to the process of cervical cancer development.

**Methods:** Expressions level of genes in cervical cancer and normal tissue samples were obtained from GTEx and TCGA database. Differently expressed RNA binding proteins (DEGs) were retrieved by Wilcoxon sum-rank test. ClusterProfiler package worked in R software was used to perform GO and KEGG enrichment analyses. Univariate proportional hazard cox regression and multivariate proportional hazard cox regressions were applied to identify DEGs equipped with prognostic value and other clinical independent risk factors. ROC curve was drawn for comparing the survival predict feasibility of risk score with other risk factors in CC patients. Nomogram was drawn to exhibit the prediction model and validated by C-index and calibration curve. Correlations between Differentially expressed RNA binding proteins (DEGs) and other clinical features were investigated by t test or Cruskal wallis analysis. Correlation between Immune and DEGs in cervical cancer was investigated by ssGSEA.

**Results:** 347 differentially expressed RBPs (DEGs) were retrieved from cervical cancer tissue and normal tissue samples. GO enrichment analysis showed that these DEGs involved in RNA splicing, catabolic process and metabolism. Cox regression model showed that there were ten DEGs significantly associated with overall survival of cervical cancer patients. WDR43 (HR = 0.423, P=0.008), RBM38 (HR = 0.533, P<0.001), RNASEH2A (HR=0.474, P=0.002) and HENMT1 (HR=0.720, P=0.071) played protective roles in survival among these ten genes. Stage (Stage IV vs Stage I HR = 3.434, P<0.001) and risk score (HR = 1.214, P< 0.001) were sorted as independent prognostic risk factors based on multivariate cox regression. ROC curve validated that risk score was preferable to predict survival of CC patients than other risk factors. Additionally, we found some of these ten predictor DEGs were correlated significantly in statistic with tumor grade or stage, clinical T stage, clinical N stage, pathology or risk score (all P< 0.05). Part of immune cells and immune functions showed a lower activity in high risk group than low risk group which is distinguished by median risk score.

**Conclusion:** Our discovery showed that many RNA binding proteins involved in the progress of cervical cancer, which could probably serve as prognostic biomarkers and accelerate the discovery of treatment targets for CC patients.

## Introduction

One of the most challenging malignancies is cervical cancer (CC) observing among females worldwide[1]. It is showed that CC led to more than about 311 thousand people death all over the world in statistically in 2018[2]. One of major reasons for cervical cancer is infection of high-risk human papillomavirus (HPV), although the occurrence of cervical cancer cannot be fully elucidated by HPV

infection[3]. Radical hysterectomy, radical radiotherapy and chemotherapy based on cisplatin are major treatment methods for CC patients until now[4]. It is reported that patients with locally advanced cervical cancer had their five-year overall survival (OS) increased into about 70% after chemotherapy[5]. Nonetheless, the recurrence of cervical cancer after surgery or radiotherapy remains a problem difficult to solve.[6] The circumstance of limited treatment and a poor prognosis is the reality that CC patients with relapse have to face.[7]

RNA binding proteins (RBPs) belong to one of the crucial series cellular proteins. Their interaction with RNA by means of recognizing special RNA binding domains plays a significant role in various kinds of post-transcriptional regulation. For example, RNA transport, translation control, intracellular localization, shearing, sequence editing are both under the influence of RNA binding proteins [8]. Former studies have discovered more than 1500 proteins who involved in RNA binding in homo sapiens genome [9]. There is a significant district in RBPs, which contains 60–100 residues. This district often adopts an  $\alpha\beta$  topology which assists them to bind the RNA according to concrete nucleic acid sequence [10]. The origination and development of many diseases have been discovered to be correlated with RBPs. For example, spinal muscular atrophy and myotonicdystrophy are two kinds of typical disease [11]. Undoubtedly, the origination and development of cancer has been reported to have something to do with RBPs. For example, HuR, which is a RBP is able to accelerate the proliferation and promote metastasis of gastric cancer[12]. Zhang H et al reported that AGO2 increases oncogenic miR-19b biogenesis by Acetylation which leads to the facilitation of cancer progression[13]. The proliferation of lung cancer cells can be regulated by cancer-associated alternative splicing. This process is inhibited by QKI-5[14]. ESRP1 accelerates the EMT of ovarian carcinoma cells[15]. All these researches revealed RBPs as important adjustment moleculars in the process of cancer development.

Nowadays, FIGO stage serves as a majority tool for doctor to predict the survival of cervical cancer patients in clinical.[16] There is deficiency in FIGO stage system that patients may have different individual survival time even if they are attributed to same FIGO stage[17]. In order to provide doctors with a better prognostic prediction tool for CC patients, more clinical factors should be taken into consideration. Recently, the prognosis model involved with the expression level of RBPs has become popular and been constructed in colorectal cancer[18], hepatocellular cancer[19] and breast cancer[20], etc. So, the prognostic prediction role of RNA binding proteins in CC triggered our interest. To begin with, the differently expressed RNA binding proteins (DEGs, differently expressed genes) were retrieved from gene expression profile of tumor tissues and normal tissues. They were uploaded to STRING database for constructing protein-protein interaction network. A cox regression model for predicting the survival of cervical cancer patients was constructed with DEGs involved in the PPI network prognosis signature. The predict factors involved in this cox regression model had been validated by the Kaplan-Meier analysis. The survival status discrimination efficacy of risk score was compared with other clinical factors by means of ROC curve and quantified by area under the curve (AUC). Moreover, GO and KEGG enrichment analysis was applied to explore the functional pathways that screened DEGs in PPI network and their subnetworks involved in. Finally, we also explored the relationships between the risk scores which was counted by DEGs signature and immune cells or functions.

## Materials And Methods

### Acquisition data from GTEx and TCGA dataset

The expression level of genes in normal cervix was downloaded from GTEx database (<https://www.genome.gov/Funded-Programs-Projects/Genotype-Tissue-Expression-Project>), which is a website containing great quantity of gene expression data resourced from healthy people. The expression level of genes in cervical cancer and the corresponding clinical data were downloaded from TCGA database (<https://portal.gdc.cancer.gov/>), which is a landmark cancer genomics program.

Clinical pathological data of patients from TCGA is available in additional file table S1. Gene array data from GTEx and TCGA was normalized by means of limma package from R bioconductor software.

Totally 1542 RBPs (RNA binding proteins) [21] were obtained to screen the gene expression profile.

### Identification of DEGs (Different expression genes, Different expression RBPs)

Wilcoxon signed-rank test was applied to identify differentially expressed RNA binding protein genes (DEGs). The cut-off values were set based on left parameters, false discovery rate (FDR) < 0.05 and  $|\log_2 \text{fold change} (\log_2 \text{FC})| > 0.5$ .

### Functional enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed by clusterProfiler package [22] in R software. Results were filtered by FDR (false discovery rate) < 0.05, top results were presented and recognized as significant items.

### Construct protein-protein interaction network and the subnetwork.

The sorted DEGs were applied to construct protein-protein interaction network by STRING (The Search Tool for the Retrieval of Interacting Genes). The PPI network was visualized by Cytoscape software. The MCODE which is a plug-in for Cytoscape was applied to obtain the first three relevant sub-network modules. The genes in them were applied to perform GO and KEGG enrichment analysis respectively.

### Construct cox regression model

The correlation between the overall survival (OS) and differently expressed RNA binding proteins (DEGs) was firstly investigated by univariate cox regression. Variables were screened by p-value (< 0.05) presented from Wald  $\chi^2$  test. All the significant variables were applied to construct multivariate cox regression model and sorted by AIC value. Then, the expression level of significant DEGs were included into multivariate cox regression with other clinical factors to construct prognostic prediction model.

### ROC curve analysis

The prediction value of independent risk factors was investigated by Receiver Operating Characteristic (ROC) curve with area under the curve (AUC). ROC curve was drawn by survivalROC package in R

software, which is designed for survival data. AUC was applied to measure the sensitivity and specificity of prediction variable, which ranges from 0.5 to 1. The larger the AUC is, the better the variable predict the prognosis.

## Development of the nomogram

Nomogram was drawn to exhibit the prediction model constructed based on independent prognostic factors sorted by univariate and multivariate cox regression. In order to evaluate the calibration and discrimination of nomograph, calibration curves were plotted and Harrell C-index was calculated. Bootstrapping validation with 100 bootstraps resample was applied to calculate C-index for this nomogram.

## Immune and RNA binding protein in cervical cancer

The infiltrating scores of 16 immune cells was measured by the single-sample gene set enrichment analysis (ssGSEA) in the "gsva" R package. The activities of 13 immune-related pathways was evaluated in the same way[23]. The annotated gene set file applied in ssGSEA analysis is provided in additional file Table S2[24].

## Experimental validation by RT-qPCR

RT-qPCR experiment was conducted to validate the expression level of five differentially expressed RNA binding proteins retrieved by bioinformatic analysis. CC specimens and correspondent normal cervical tissues come from CC patients who received operation from 2020 January to 2021 January in Shanghai East Hospital, School of Medicine, Tongji University. The clinical pathological data of patients is available in additional file Table S3. Internal review board of Shanghai East Hospital, School of Medicine, Tongji University have approved this study.

Total RNA was retrieved by TRIzol (Invitrogen, USA) from CC tissues and correspondent normal tissues. The purified RNA was transcribed into cDNA (Complementary DNA) by PrimeScript® RT reagent Kit with gDNA (genomic DNA) Eraser (Takara). SYBR master mix kit (Vazyme, China) was used to detect the expression level of these DEGs (Differently expressed genes, RNA binding proteins) on the QuantStudio RT-qPCR System (Q6, Applied Biosystems, USA). Endogenous GAPDH (glyceral-dehyde-3-phosphate dehydrogenase) was used to normalize the expression level of each gene by  $2^{-\Delta\Delta C_t}$ . The primers were synthesized by Sangon Company, China.

## Results

### Sort DEGs from RNA binding protein gene expression profile

An expression profile dataset was combined with data from GTEx and TCGA database, which included 13 normal cervix samples and 306 cervical cancer samples. The clinical data of the patients was

downloaded from TCGA database and intergrated into expression matrix by perl software. We obtained 347 differently expressed RNA bind proteins (DEGs) by comparison of the expression level of RNA bind proteins (DEGs) between cervix tissues and cervical cancer tissues. There were 177 genes down-regulated in tumor samples and 170 genes up-regulated in tumor samples compared with normal samples (additional file table S4). The detail of DEGs' expression matrix was presented by heatmap and volcano plot (Figure.1a-b).

## **Bio-functional enrichment analysis of DEGs**

We conducted GO and KEGG enrichment analysis by clusterprofiler package in R software to evaluate the biological function of our retrieved DEGs. These differently expressed RBPs were devided into up-regulated group and down-regulated group for enrichment analysis individually.

In GO analysis, for BP (biological process) category, downregulated DEGs mainly enriched in RNA splicing, RNA catabolic process and mRNA catabolic process. For CC (celluar components) category, downregulated DEGs mainly enriched in cytoplasmic ribonucleoprotein granule, ribonucleoprotein granule and cytoplasmic stress granule. For MF (molecular function) category, downregulated DEGs mainly enriched in catalytic activity, acting on RNA, single – stranded RNA binding and mRNA 3'-UTR binding. In KEGG pathway analysis, down-regulated DEGs mainly enriched in mRNA surveillance pathway, RNA transport and Ribosome (Figure.2a-d) (additional file table S5-S6).

In GO analysis for up regulated group, biological process (BP) category mainly include ncRNA processing, RNA phosphodiester bond hydrolysis, and tRNA metabolic process. They have something to do with the process of cervical cancer. Items in cellular component (CC) mainly include ribonucleoprotein granule and preribosome. They involved in RNA binding and protein expression level adjustment, which may help cervical cancer cells survive. Molecular function (MF) category showed that DEGs were mainly enriched in catalytic activity, acting on RNA, double – stranded RNA binding and ribonuclease activity, respectively. They revealed the function of RBP. In result of KEGG analysis for upregulated DEGs, DEGs were enriched in pathways such as Ribosome biogenesis in eukaryotes, mRNA surveillance pathway and RNA transport. These pathways help cancer cells live a better life (Figure.3a-d) (additional file table S7-S8).

## **Construct Protein-protein interaction (PPI) network**

A PPI network was constructed by Cytoscape software. The information of nodes and network was obtained from STRING database accroding to uploaded DEGs. The PPI network incorporated 2545 edges and 320 nodes (Figure.4a). The co-expression network was treated by MCODE plug-in for Cytoscape to identify the most correlated three subnetworks (Figure.4b). Acquired first important crucial module consist of 27 nodes and 335 edges (Figure.4c). The GO enrichment analysis result shows that the RBPs in the key module 1 were mainly enriched in ribosome biogenesis, preribosome and RNA helicase activity. Moreover, in KEGG analysis they were enriched in Ribosome biogenesis in eukaryotes pathway. The GO and KEGG analysis results of both three subnetworks were displayed in Table1 and Table2.

Table 1  
The GO enrichment analysis of three most significant MCODE components

ONTOLOGY	ID	Description	pvalue	qvalue	Count
subnetwork1					
BP	GO:0042254	ribosome biogenesis	2.8E-32	3.83E-30	20
BP	GO:0006364	rRNA processing	9.39E-26	5.61E-24	16
BP	GO:0034470	ncRNA processing	1.23E-25	5.61E-24	18
CC	GO:0030684	preribosome	7.18E-23	8.31E-22	12
CC	GO:0032040	small-subunit processome	9.73E-12	5.63E-11	6
CC	GO:0030686	90S preribosome	5.39E-10	2.08E-09	5
MF	GO:0003724	RNA helicase activity	5.67E-06	0.000137	4
MF	GO:0140098	catalytic activity, acting on RNA	2.08E-05	0.000252	6
MF	GO:0004386	helicase activity	0.000103	0.000833	4
subnetwork2					
BP	GO:0006414	translational elongation	1.82E-22	4.15E-20	14
BP	GO:0070126	mitochondrial translational termination	1.33E-20	1.52E-18	12
BP	GO:0006415	translational termination	9.55E-20	7.27E-18	12
CC	GO:0044391	ribosomal subunit	3.49E-33	5.15E-32	20
CC	GO:0005840	ribosome	5.99E-30	4.42E-29	20
CC	GO:0015934	large ribosomal subunit	2.13E-23	1.04E-22	14
MF	GO:0003735	structural constituent of ribosome	5.74E-22	1.27E-20	15
MF	GO:0003746	translation elongation factor activity	0.000518	0.005724	2
MF	GO:0004540	ribonuclease activity	0.000931	0.006861	3
subnetwork3					
BP	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	1.03E-20	5.72E-19	17
BP	GO:0000398	mRNA splicing, via spliceosome	1.03E-20	5.72E-19	17
BP	GO:0000375	RNA splicing, via transesterification reactions	1.17E-20	5.72E-19	17

ONTOLOGY	ID	Description	pvalue	qvalue	Count
CC	GO:0005849	mRNA cleavage factor complex	1.46E-10	4.75E-09	5
CC	GO:0005681	spliceosomal complex	5.86E-10	9.57E-09	8
CC	GO:0046540	U4/U6 x U5 tri-snRNP complex	7.79E-09	6.35E-08	5
MF	GO:0008135	translation factor activity, RNA binding	1.02E-08	4.31E-07	6
MF	GO:0140098	catalytic activity, acting on RNA	4.10E-07	8.64E-06	8
MF	GO:0003743	translation initiation factor activity	2.35E-06	3.30E-05	4

Table 2  
The KEGG pathway analysis of three most significant MCODE components

ID	Description	pvalue	qvalue	Count
subnetwork1				
hsa03008	Ribosome biogenesis in eukaryotes	7.88E-15	8.29E-15	8
subnetwork2				
hsa03010	Ribosome	2.16E-16	9.10E-16	12
hsa05171	Coronavirus disease - COVID-19	3.72E-08	7.83E-08	8
hsa03013	RNA transport	0.000973	0.001365	4
hsa03018	RNA degradation	0.015887	0.016723	2
subnetwork3				
hsa03015	mRNA surveillance pathway	4.25E-12	5.37E-11	9
hsa03013	RNA transport	7.41E-07	4.68E-06	7
hsa03040	Spliceosome	3.70E-06	1.56E-05	6
hsa03020	RNA polymerase	0.003703	0.011693	2

Table.3 prognostic related gene sorted by multivariate cox regression

id	coef	HR	HR.95L	HR.95H	pvalue
EIF3C	0.68945	1.992619	0.862064	4.605842	0.106796
WDR43	-0.85954	0.423357	0.224097	0.799793	0.00809
PRPF40B	1.29466	3.649756	1.710637	7.786992	0.000812
RBM38	-0.62881	0.533228	0.370969	0.766457	0.000682
EEF1D	0.484649	1.623605	0.982808	2.682204	0.058457
CTU1	0.569026	1.766545	1.155554	2.700595	0.008599
RNASEH2A	-0.74683	0.473867	0.298073	0.753339	0.001592
HENMT1	-0.32793	0.720411	0.504799	1.028115	0.070737
ZC3HAV1L	0.538778	1.713911	1.058803	2.77435	0.028345
NUFIP1	0.991838	2.696187	1.344851	5.405375	0.005193

### Retrieve DEGs related to prognosis.

To begin with, we retrieved prognostic related DEGs by kaplan-meier analysis and univariate cox regression with wald  $X^2$  test. In kaplan-meier analysis, patients were divided into two groups according to the median of expression level of DEGs. Their survivals were compared by log-rank test. The DEGs were considered significant when p-value of log-rank test is less than 0.05. Those DEGs were verified by univariate cox regression. Variables with wald  $X^2$  test p-value less than 0.05 were selected. It is validated that 18 DEGs (EIF3C, WDR43, BICC1, HEATR1, PRPF40B, RBM4, RBM38, CLK3, EEF1D, SAMD4A, CTU1, RNASEH2A, HENMT1, ENOX1, FBXO17, SMG8, ZC3HAV1L, NUFIP1) were significantly in statistic with the overall survival of CC patients (shown in Figure.5a). Moreover, these genes were applied to construct multivariate cox regression model, AIC value was used to sort the variable. Ten DEGs (EIF3C, WDR43, PRPF40B, RBM38, EEF1D, CTU1, RNASEH2A, HENMT1, ZC3HAV1L, NUFIP1) were preserved at last and regarded as prognostic related DEGs (Fig. 5b and Table 3).

Four genes (WDR43, RBM38, RNASEH2A, HENMT1) played protective roles ( $HR < 1$ ) among these ten DEGs. The other six genes (EIF3C, PRPF40B, EEF1D, CTU1, ZC3HAV1L, NUFIP1) were presented as risk factors for CC patients' survival ( $HR > 1$ ). Finally, the risk score was calculated according to the expression level of these ten genes ( $\text{riskscore} = h_0(t) \exp(\sum_{j=1}^n \text{Coef}_j \times X_j)$ ,  $n = 10$ ,  $\text{Coef}_j$  is the coefficient of each DEG,  $X_j$  is the relative expression levels of each DEG,  $h_0(t)$  is baseline risk function). The median risk score value was regarded as cutoff point to divide the CC patients into high risk group ( $n=152$ ) and low risk group ( $n=152$ ). The patients' overall survival (OS) of high risk group is shorter than that of low risk group significantly (median time = 3.4 years vs. more than 8 years, log rank  $p < 0.001$ , Figure.5c).

## Draw prognostic hazard curves

Prognostic hazard curves was drawn to evaluate the survival time for the patients. It is observed that the survival time diminished with the increasing of risk score for the dead patients (Figure.6a-b). Furthermore, the quantity of patients alive decreased with the ascend of risk score for patients too. RNASEH2A was shown to be down-regulated in group with high-risk according to the risk heatmap. Whereas, CTU1 was regarded as a tumor accelerating gene because it was up-regulated in high-risk group (Figure.6c).

## Prognostic factors and prediction model for OS

The risk score and other clinical factors were combined to construct cox regression model. It is showed in univariate cox regression model that clinical stage and risk score were correlated with overall survival (OS) of CC patients ( $P < 0.001$ ,  $P < 0.001$ , Figure.7a). Multivariate cox regression validated that clinical stage (Stage IV vs Stage I HR = 3.434,  $P < 0.001$ ) and risk score (HR = 1.214,  $P < 0.001$ ) were independent risk factors for survival (Figure.7b).

For the sake of evaluating the discrimination of each predicting factors, ROC curves were constructed in 0.5-year, 1-year, 3-year and 5-year with the prediction factors (age, stage and risk score). Moreover, we assess the feasibility of discrimination of survival or dead patients using the area under curve (AUC) values. ROC curve reveals that the risk score showed a better ability to predict the survival of CC patients (AUC = 0.932, 0.843, 0.805, 0.832 for 0.5-year, 1-year, 3-year and 5-year) than other prediction factors (Figure.8a-d).

## Analyse relationship between clinical features and DEGs predictor

The correlations between the ten prognostic DEGs and clinical features was evaluated by t-test or Kruskal-Wallis test depend on the quantity of categories of clinical features. It showed that the expression level of CTU1 and ZC3HAV1L were significantly different expressed in statistic with each clinical stage ( $P$ -values = 0.013 and 0.040 respectively) (Figure.9a-b). Furthermore, the expression level of CTU1 and ZC3HAV1L were higher in the advanced T stage patients ( $P$ -values = 0.009 and  $< 0.001$  respectively), implying their dangerous roles with the development of cervical cancer (Figure.9c-d). The expression level of CTU1 was significantly associated with N stages which implying that its expression levels increased with progression of lymph node metastasis (Figure.9e). The expression level of EEF1D increased with advanced M stage, which implied that its expression level may be correlated with the organ metastasis ability of cervical cancer (Fig. 9f). The expression level of CTU1, RBM38, WDR43 varies with different pathology of cervical cancer (Fig. 9g-i). In addition, the expression level of EEF1D, RBM38 and WDR43 ascended with higher tumor pathology grade ( $p$ -value = 0,019, 0.020, 0.034) (Figure.9j-l).

## Establish and validate the nomogram

Three prognostic indicators including age, clinical stage and ten prognostic prediction RBPs were selected to establish the nomogram (Fig. 10a). The discrimination and calibration of nomogram was validated based on C-index and calibration curve. Analysis result revealed that the C-index of the

constructed nomogram is 0.808 and the 1-year, 3-year and 5-year calibration curve in Fig. 10b, 10c and 10d demonstrated that the nomogram can partially predict the prognosis of CC patients.

## Enrichment analysis of immune cell and function

SsGSEA R package was used to investigate the enrichment scores of 16 immune cell subpopulations and their 13 correlated immune functions. It is revealed that 5 kinds of immune cells (such as B cells, iDCs, mast cells, NK cells, pDCs) caught a lower score in high risk group than low risk group (Figure. 11a). What is more. The scores of the 2 types immune functions, such as HLA, Inflammation – promoting were significantly higher in low-risk group. Their enrichment scores suggested the immunological functions of high risk group should be injured more than low risk group classified by expression level of prognostic DEGs (Figure. 11b).

## RT-qPCR experiment validation

Five differently expressed genes (POLR2J2, RBMS1, DARS2, GLE1, RPP25) was validated by RT-qPCR. The result shows that POLR2J2, RBMS1 were down-regulated in cervical cancer than normal cervix tissues. RPP25 was up-regulated in cervical cancer than normal tissues. The expression level of DARS2 and GLE1 showed no significant difference between normal and cancer tissue. These results required more validation in future by a larger scale clinical samples (additional file table S9-S11, Fig. 12).

## Discussion

Nowadays, tumor has become one of the greatest intimidation to human health, which has exceed the cardiovascular disease [25]. Cervical cancer has become the second most common malignancies among females all over the world[26]. Especially, in developing countries, where it is not popular for females to take part in cervical screening, cervical cancer posed a greater threat to woman than developed country [27]. Cervical intraepithelial neoplasias (CIN) was recognized as the precursor lesions for cervical cancer. Persistent infection of human papillomaviruses (HPVs) is one of majority reasons led to CIN [28]. The potential mechanism of CIN is assumed that the infection of virus altered gene transcription or affected the posttranscription regulation of message RNA. The possible process of posttranscription included two categories. Firstly, microRNA is able to trigger degradation of the target message RNA by binding ones 3' untranslated regions (UTR) [29]. Secondly, RNA binding proteins involve in the process, editing, stable maintenance, transportation and translation of message RNA [9, 30]. Recently, microarray and RNA sequencing technologies have emerged as favourable tools for scientists to investigate the potential variation of gene or gene modification in the development of cancer [31]. In this study, the RNA sequencing data of 306 cervical cancer tissues and 13 normal cervix tissues from GTEx and TCGA databases was integrated to analyse the expression profile of differently expressed RNA binding proteins (also called DEGs in this article, Differently expressed genes) in cervical cancer. 348 DEGs was retrieved by wilcoxon sum-rank test, of which 177 DEGs were down regulated in tumor samples and 170 DEGs were up regulated in tumor samples. The functional enrichment analysis of GO and KEGG were

performed for the downregulated and upregulated DEGs respectively. The PPI network was constructed for sorting the candidate genes of prognostic prediction model by STRING database. Moreover, this DEGs was screened by cox regression with wald  $X^2$  test and kaplan-meier analyse with log-rank test. Among these DEGs, WDR43, RBM38, RNASEH2A and HENMT1 with HR < 1 played a protective role in survival. Other six genes (EIF3C, PRPF40B, EEF1D, CTU1, ZC3HAV1L, NUFIP1) were considered as risk factors with HR > 1. The nomograph was drawn to present the prognostic prediction model with FIGO stage and RBPs predictor. It was validated by C-index and calibration curve subsequently. In addition, the enrichment analysis of immune cell and function was performed by ssGSEA package in R software.

We investigated the biological functions of These DEGs by GO analysis. To begin with, the enrichment of cell components was located in the ribosome, cytoplasmic ribonucleoprotein granule and the ribonuclease. They play crucial roles in transmission of genetic information from DNA to protein. Protein was synthesised in ribosome by translating the coding information from RNA. The mutation of ribosomal protein may exert an influence on degradation of p53 protein which involved in the process of many kinds of cancer, such as endometrial cancer, T-cell acute lymphoblastic leukemia, chronic lymphocytic leukemia and colorectal cancer[32]. Many kinds of disease have been reported having something to do with RNA processing or RNA metabolism, which exerted influence on RNA translation[33–35]. The forming of ribonucleoprotein complexes has been recognized as the result of interaction of RNA and RBPs. They sustain the stability of target message RNAs, after which the the efficiency of mRNA translation is promoted. For example, oncogenic RNA binding protein SRSF1 is reported to accelerate the proliferation of lung cancer cells by strengthening the message RNA stability of DNA ligase 1 [36]. What is more, ribonucleoprotein granule was discovered as a crucial region for protein synthesis. The development of cancer is affected by the modification of ribonucleoprotein, because of its significant role in RNA translation [32]. Moreover, the category of molecular function in GO analysis revealed the interactions of RNA and proteins such as RNA methyltransferase activity. RBPs have been discovered to bind with many kinds of RNA such as pre-mRNA, snRNA, tRNA and mRNA. The regulation of various enzyme was also displayed in GO analysis such as endoribonuclease, ribonuclease and nuclease. They are correlated to synthesis or repair of DNA and metabolism of RNA. For example, in the field of correlation between cervical cancer and RNA methylation. Jingxin Pan et al. developed a prognostic prediction model for cervical cancer patients based on m6A RNA methylation regulator [37]. While, most of research concentrate on the methylation of protein or DNA such as the promoters of genes instead of RNA. The underlying mechanism of RNA methylation and CC remains to be revealed. Finally, in term of biological process category of GO analysis, differently expressed RBPs have something to do with the processing of both coding RNA and non-coding RNA such as rRNA and tRNA. Both of RNA splicing and metabolism were adjusted by these differently expressed RBPs. Our result was consistent with the consensus reached before. It is reported that RNA binding protein (RBP) quaking (QKI) was able to interact with the QKI response elements (QREs) in SLC26A4 gene introns, which lies in the 3'UTR (3' untranslation region) of mRNA after transcription, thereby promoting circSLC26A4 biogenesis. CircSLC26A4 promotes the proliferation of cervical cancer cells in both vivo and vitro [38]. The RBP HuR was discovered to promote the growth of cervical cancer cells by interaction with the 3'UTR of RBP nucleolin (NCL) mRNA, which

specifically promoted the translation of NCL without the alteration of NCL mRNA levels[39]. In other kind of cancer, RBP Musashi1 (Msi1) promoted the proliferation of colon cancer cells by target the 3'UTR of p21(cip1) [40]. Then, the items in KEGG pathway analysis suggested that the origination and development of cervical cancer is regulated by RBPs through mRNA surveillance pathway, RNA transport and RNA degradation. The underlying correlation between RBPs and signal pathways should be under reaserch further.

The relationships between many RBPs and cancer has been reported by former studies which were consist with our study. We discovered that CTU1 is a risk factor for CC patients. It has been reported that CTU1/2, which is partner enzymes in U34 mcm<sup>5</sup>s<sup>2</sup>-tRNA modification, sustains metastasis and invasion of breast cancer[41]. Francesca Rapino et al. reported that the inhibition of CTU1 and proteins synergizing with it could kill melanoma cells[42]. Ming Zhang et al. identified the copy number amplifications of CTU1 in 25% of myxopapillary ependymomas by means of whole exome sequencing [43]. CTU1 has been identified as one of prognostic predictors for prostate cancer and bladder cancer[44, 45]. We also identified NUFIP1 as a risk factor of CC patients. The forming of ETV6-NUFIP1 fusion gene has been reported as a potential cause of acute lymphoblastic leukemia in Mexico [46]. Gaurang P Deshpande et al. reported that NUFIP1 had something to do with genome stability maintenance [47] which may help cancer cells survive the pressure from environment. Mutated genes NUFIP1 had a higher level of expression in metastasis tumor than primary tumor in neuroblastoma indicating its oncogenic driver role [48]. However, the potential role of NUFIP1 in the process of cervical cancer development remains to be revealed.

The risk score calculated by expression level of DEGs was demonstrated to be a risk indicator. Patients in high risk group shows a significant lower survival than low risk group. ROC curve for risk factors suggested that risk score predicted the prognosis better than other factors which may be valuable in clinical application. It suggested that more precise therapeutic strategy should be applied to CC patients with higher risk score. At last the expression of each DEGs were analysed with patients' clinical features. CTU1 and ZC3HAV1L were significantly associated with clinic FIGO stage and T stage. Their oncogenic role was exposed gradually with the progress of clinical stage. What is more, the expression level of CTU1 increased with the N stage, which showed that it might promote the lymph node metastasis of cervical cancer. The expression level of EEF1D increased with M stage, which showed that it might had something to do with organ metastasis. Both of WDR43, CTU1 and RBM38 were correlated with pathology class. The expression level of EEF1D, RBM38 and WDR43 ascended with the progression of cancer pathology grade. This information may be a clue for further research about correlation between RNA binding proteins and clinical feature in cervical cancer.

Thanks to public database such as TCGA and GTEx, the correlation between prognosis of CC patients and RBPs was analysed. A robust statistical support could be used to help the RBPs researchers in future. More clinical therapeutic schemes should be developed concentrating on RBPs genes in CC patients. There are some limits in this study. To begin with, the clinical stage, pathology grade and the treatment schemes downloaded from TCGA were incomplete. The HPV infection status of each patient was unknown. These

deficiencies affected the accuracy of prediction model we constructed at last. Moreover, the potential mechanisms of how RBPs regulate the development of CC and their interaction relationship were remained to be explained. Finally, the nomograph has to be validated in a larger cohort, that will be helpful for further epidemiological research. These deficiencies could be solved with a larger scale of clinical data appeared in future.

## Conclusion

DEGs (RNA binding proteins) with prognostic value in CC patients were identified by wilcoxon rank-sum test, cox regression, wald  $X^2$ , and AIC value. They have potential to be diagnose or treatment biomarks for CC patients. These DEGs and nomograph should be validated in future to confirm our study's result.

## Abbreviations

CC: Cervical Cancer; RBPs: RNA binding proteins; OS: overall survival; TCGA: The Cancer Genome Atlas; DEG: differently expressed genes; FDR: False discovery rate; GO: Gene Ontology; BP: biological process; CC: cellular component; MF: molecular function; KEGG: Kyoto Encyclopedia of Genes and Genomes; ROC: Receiver operating characteristic curve; AUC: area under curve; HR: Hazard ratio; DCs: Dendritic cells; iDCs: Immature DCs; pDCs: Plasmacytoid dendritic cells; TIL: Tumor infiltrating lymphocyte; CCR : Cytokine–cytokine receptor; APC: Antigen presenting cells; Tfh:Follicular helper T cells;

## Declarations

### ACKNOWLEDGMENTS

We would like to thank all the authors listed for their contribution to the present study. Thanks for the maintenance personnel from TCGA and GTEx database.

### Authors' contribution

ZYH contributed to conception of the work and write the article designing the analysis scheme, contributing to data acquisition and performing data analysis. FL and QCL reviewed and approved the final version of the manuscript. All authors read and approved the final manuscript.

### Funding

This study was funded by the National Natural Science Foundation of China Projects (Nos. 81771529) to F Li; Health Special Project of Pudong Health Bureau of Shanghai (Grant No. PW2020E-5) to QC Li; the National Natural Science Foundation of China (Grant No. 81773266) to QC Li.

### Availability of data and material

All data were available in TCGA (<https://portal.gdc.cancer.gov>) and GTEx database (<http://commonfund.nih.gov/GTEX/>).

### **Ethical statement**

The Internal Review Board of Shanghai East Hospital, School of Medicine, Tongji University has approved this study.

### **Consent for publication**

This manuscript has been consent for publication by all the authors.

### **Competing interests**

The authors declare that they have no conflicts of interest with the contents of this article.

## **References**

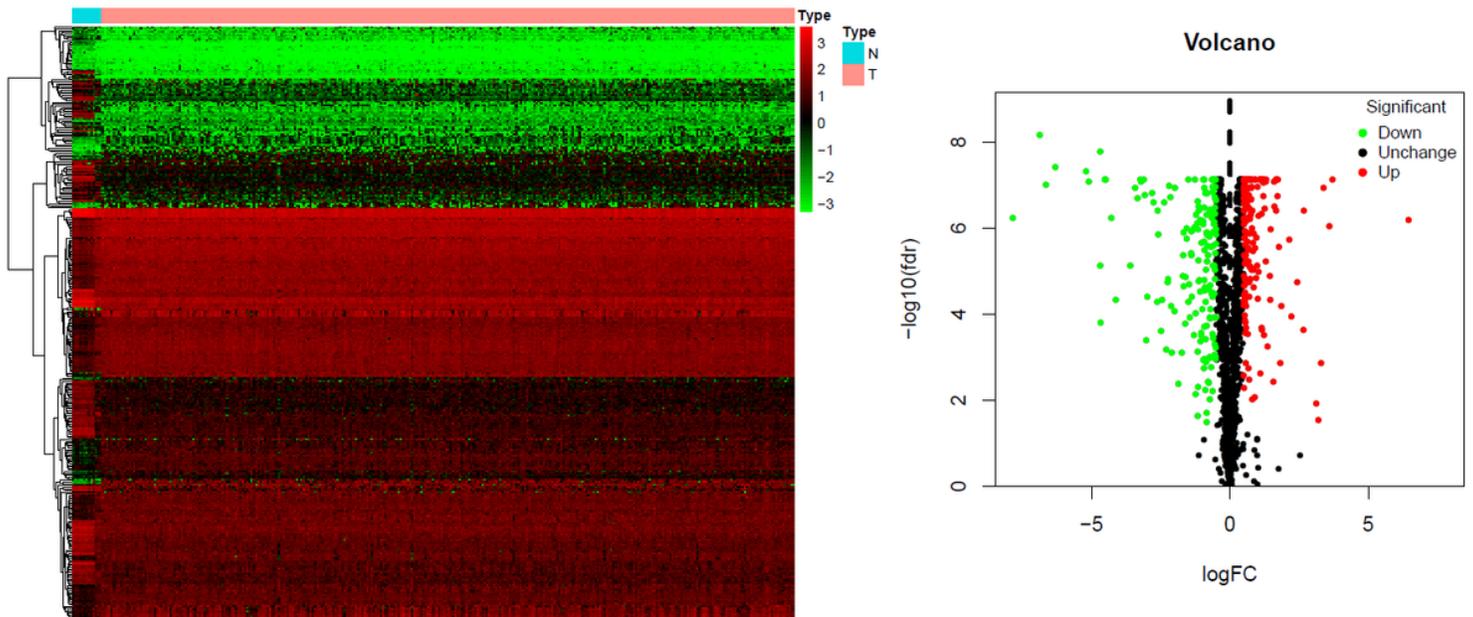
1. Olusola P, Banerjee HN, Philley JV, Dasgupta S. **Human Papilloma Virus-Associated Cervical Cancer and Health Disparities.** *Cells* 2019, 8(6).
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394–424.
3. Jin J. HPV Infection and Cancer. *JAMA.* 2018;319(10):1058.
4. Li H, Wu X, Cheng X. Advances in diagnosis and treatment of metastatic cervical cancer. *J Gynecol Oncol.* 2016;27(4):e43.
5. Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. *Lancet.* 2019;393(10167):169–82.
6. Boussios S, Seraj E, Zarkavelis G, Petrakis D, Kollas A, Kafantari A, Assi A, Tatsi K, Pavlidis N, Pentheroudakis G. Management of patients with recurrent/advanced cervical cancer beyond first line platinum regimens: Where do we stand? A literature review. *Crit Rev Oncol Hematol.* 2016;108:164–74.
7. Dizon DS, Mackay HJ, Thomas GM, Werner TL, Kohn EC, Hess D, Rose PG, Covens AL. State of the science in cervical cancer: Where we are today and where we need to go. *Cancer.* 2014;120(15):2282–8.
8. Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* 2008;582(14):1977–86.
9. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. *Nat Rev Genet.* 2014;15(12):829–45.
10. Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol.* 2007;8(6):479–90.

11. Lukong KE, Chang KW, Khandjian EW, Richard S. RNA-binding proteins in human genetic disease. *Trends Genet.* 2008;24(8):416–25.
12. Xie M, Ma T, Xue J, Ma H, Sun M, Zhang Z, Liu M, Liu Y, Ju S, Wang Z, et al. The long intergenic non-protein coding RNA 707 promotes proliferation and metastasis of gastric cancer by interacting with mRNA stabilizing protein HuR. *Cancer Lett.* 2019;443:67–79.
13. Zhang H, Wang Y, Dou J, Guo Y, He J, Li L, Liu X, Chen R, Deng R, Huang J, et al. Acetylation of AGO2 promotes cancer progression by increasing oncogenic miR-19b biogenesis. *Oncogene.* 2019;38(9):1410–31.
14. Zong FY, Fu X, Wei WJ, Luo YG, Heiner M, Cao LJ, Fang Z, Fang R, Lu D, Ji H, et al. The RNA-binding protein QKI suppresses cancer-associated aberrant splicing. *PLoS Genet.* 2014;10(4):e1004289.
15. Jeong HM, Han J, Lee SH, Park HJ, Lee HJ, Choi JS, Lee YM, Choi YL, Shin YK, Kwon MJ. ESRP1 is overexpressed in ovarian cancer and promotes switching from mesenchymal to epithelial phenotype in ovarian cancer cells. *Oncogenesis.* 2017;6(11):e391.
16. Matsuo K, Machida H, Mandelbaum RS, Konishi I, Mikami M. Validation of the 2018 FIGO cervical cancer staging system. *Gynecol Oncol.* 2019;152(1):87–93.
17. Balaya V, Guani B, Magaud L, Bonsang-Kitzis H, Ngô C, Mathevet P, Lécuru F. On Behalf Of The Senticol G: **Validation of the 2018 FIGO Classification for Cervical Cancer: Lymphovascular Space Invasion Should Be Considered in IB1 Stage.** *Cancers (Basel)* 2020, 12(12).
18. Fan X, Liu L, Shi Y, Guo F, Wang H, Zhao X, Zhong D, Li G. Integrated analysis of RNA-binding proteins in human colorectal cancer. *World J Surg Oncol.* 2020;18(1):222.
19. Huang Y, Chen S, Qin W, Wang Y, Li L, Li Q, Yuan X. A Novel RNA Binding Protein-Related Prognostic Signature for Hepatocellular Carcinoma. *Front Oncol.* 2020;10:580513.
20. Wang K, Li L, Fu L, Yuan Y, Dai H, Zhu T, Zhou Y, Yuan F. Integrated Bioinformatics Analysis the Function of RNA Binding Proteins (RBPs) and Their Prognostic Value in Breast Cancer. *Front Pharmacol.* 2019;10:140.
21. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. *Nat Rev Genet.* 2014;15(12):829–45.
22. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284–7.
23. Rooney MS, Shukla Sachet A, Wu Catherine J, Getz G, Hacohen N. Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. *Cell.* 2015;160(1):48–61.
24. Zhu L, Yang F, Wang L, Dong L, Huang Z, Wang G, Chen G, Li Q. Identification the ferroptosis-related gene signature in patients with esophageal adenocarcinoma. *Cancer Cell Int.* 2021;21(1):124.
25. Lin M, Bi H, Yan Y, Huang W, Zhang G, Zhang G, Tang S, Liu Y, Zhang L, Ma J, et al. Parthenolide suppresses non-small cell lung cancer GLC-82 cells growth via B-Raf/MAPK/Erk pathway. *Oncotarget.* 2017;8(14):23436–47.

26. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015;65(2):87–108.
27. Vaccarella S, Lortet-Tieulent J, Plummer M, Franceschi S, Bray F. Worldwide trends in cervical cancer incidence: impact of screening against changes in disease risk factors. *Eur J Cancer.* 2013;49(15):3262–73.
28. Muñoz N, Bosch FX, de Sanjosé S, Herrero R, Castellsagué X, Shah KV, Snijders PJ, Meijer CJ. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med.* 2003;348(6):518–27.
29. Ha M, Kim VN. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol.* 2014;15(8):509–24.
30. Hentze MW, Castello A, Schwarzl T, Preiss T. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol.* 2018;19(5):327–41.
31. Bass AJ, Thorsson V, Shmulevich I, Reynolds SM, Miller M, Bernard B, Hinoue T, Laird PW, Curtis C, Shen H, et al. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature.* 2014;513(7517):202–9.
32. Goudarzi KM, Lindström MS. Role of ribosomal protein mutations in tumor development (Review). *Int J Oncol.* 2016;48(4):1313–24.
33. Jain A, Brown SZ, Thomsett HL, Londin E, Brody JR. Evaluation of Post-transcriptional Gene Regulation in Pancreatic Cancer Cells: Studying RNA Binding Proteins and Their mRNA Targets. *Methods Mol Biol.* 2019;1882:239–52.
34. Siang DTC, Lim YC, Kyaw AMM, Win KN, Chia SY, Degirmenci U, Hu X, Tan BC, Walet ACE, Sun L, et al. The RNA-binding protein HuR is a negative regulator in adipogenesis. *Nat Commun.* 2020;11(1):213.
35. Kim TH, Tsang B, Vernon RM, Sonenberg N, Kay LE, Forman-Kay JD. Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. *Science.* 2019;365(6455):825–9.
36. Martínez-Terroba E, Ezponda T, Bértolo C, Sainz C, Remírez A, Agorreta J, Garmendia I, Behrens C, Pio R, Wistuba, II et al: **The oncogenic RNA-binding protein SRSF1 regulates LIG1 in non-small cell lung cancer.** *Lab Invest* 2018, **98**(12):1562–1574.
37. Pan J, Xu L, Pan H. Development and Validation of an m6A RNA Methylation Regulator-Based Signature for Prognostic Prediction in Cervical Squamous Cell Carcinoma. *Front Oncol.* 2020;10:1444.
38. Ji F, Du R, Chen T, Zhang M, Zhu Y, Luo X, Ding Y: **Circular RNA circSLC26A4 Accelerates Cervical Cancer Progression via miR-1287-5p/HOXA7 Axis.** *Mol Ther Nucleic Acids* 2020, **19**:413–420.
39. Tominaga K, Srikantan S, Lee EK, Subaran SS, Martindale JL, Abdelmohsen K, Gorospe M. Competitive regulation of nucleolin expression by HuR and miR-494. *Mol Cell Biol.* 2011;31(20):4219–31.
40. Gao C, Han C, Yu Q, Guan Y, Li N, Zhou J, Tian Y, Zhang Y. Downregulation of Msi1 suppresses the growth of human colon cancer by targeting p21cip1. *Int J Oncol.* 2015;46(2):732–40.

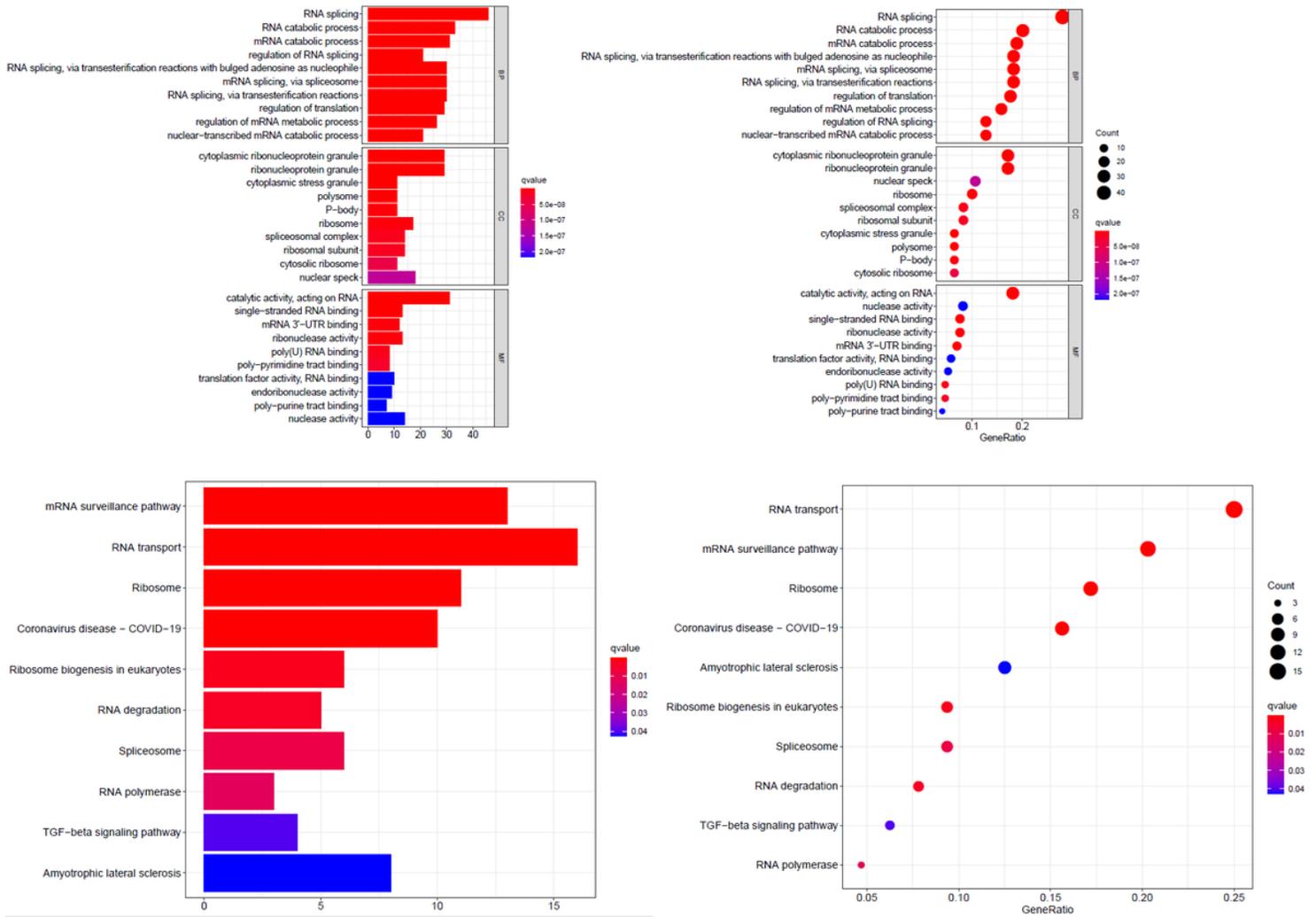
41. Delaunay S, Rapino F, Tharun L, Zhou Z, Heukamp L, Termathe M, Shostak K, Klevernic I, Florin A, Desmecht H, et al. Elp3 links tRNA modification to IRES-dependent translation of LEF1 to sustain metastasis in breast cancer. *J Exp Med*. 2016;213(11):2503–23.
42. Rapino F, Delaunay S, Rambow F, Zhou Z, Tharun L, De Tullio P, Sin O, Shostak K, Schmitz S, Piepers J, et al. Codon-specific translation reprogramming promotes resistance to targeted therapy. *Nature*. 2018;558(7711):605–9.
43. Zhang M, Iyer RR, Azad TD, Wang Q, Garzon-Muvdi T, Wang J, Liu A, Burger P, Eberhart C, Rodriguez FJ, et al. Genomic Landscape of Intramedullary Spinal Cord Gliomas. *Sci Rep*. 2019;9(1):18722.
44. 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. *Front Genet*. 2020;11:591667.
45. Guo C, Shao T, Jiang X, Wei D, Wang Z, Li M, Bao G. Comprehensive analysis of the functions and prognostic significance of RNA-binding proteins in bladder urothelial carcinoma. *Am J Transl Res*. 2020;12(11):7160–73.
46. Mata-Rocha M, Rangel-López A, Jiménez-Hernández E, Morales-Castillo BA, González-Torres C, Gaytan-Cervantes J, Álvarez-Olmos E, Núñez-Enríquez JC, Fajardo-Gutiérrez A, Martín-Trejo JA, et al: **Identification and Characterization of Novel Fusion Genes with Potential Clinical Applications in Mexican Children with Acute Lymphoblastic Leukemia.** *Int J Mol Sci* 2019, 20(10).
47. Deshpande GP, Hayles J, Hoe KL, Kim DU, Park HO, Hartsuiker E. Screening a genome-wide *S. pombe* deletion library identifies novel genes and pathways involved in genome stability maintenance. *DNA Repair*. 2009;8(5):672–9.
48. Wei JS, Johansson P, Chen L, Song YK, Tolman C, Li S, Hurd L, Patidar R, Wen X, Badgett TC, et al. Massively parallel sequencing reveals an accumulation of de novo mutations and an activating mutation of LPAR1 in a patient with metastatic neuroblastoma. *PLoS One*. 2013;8(10):e77731.

## Figures



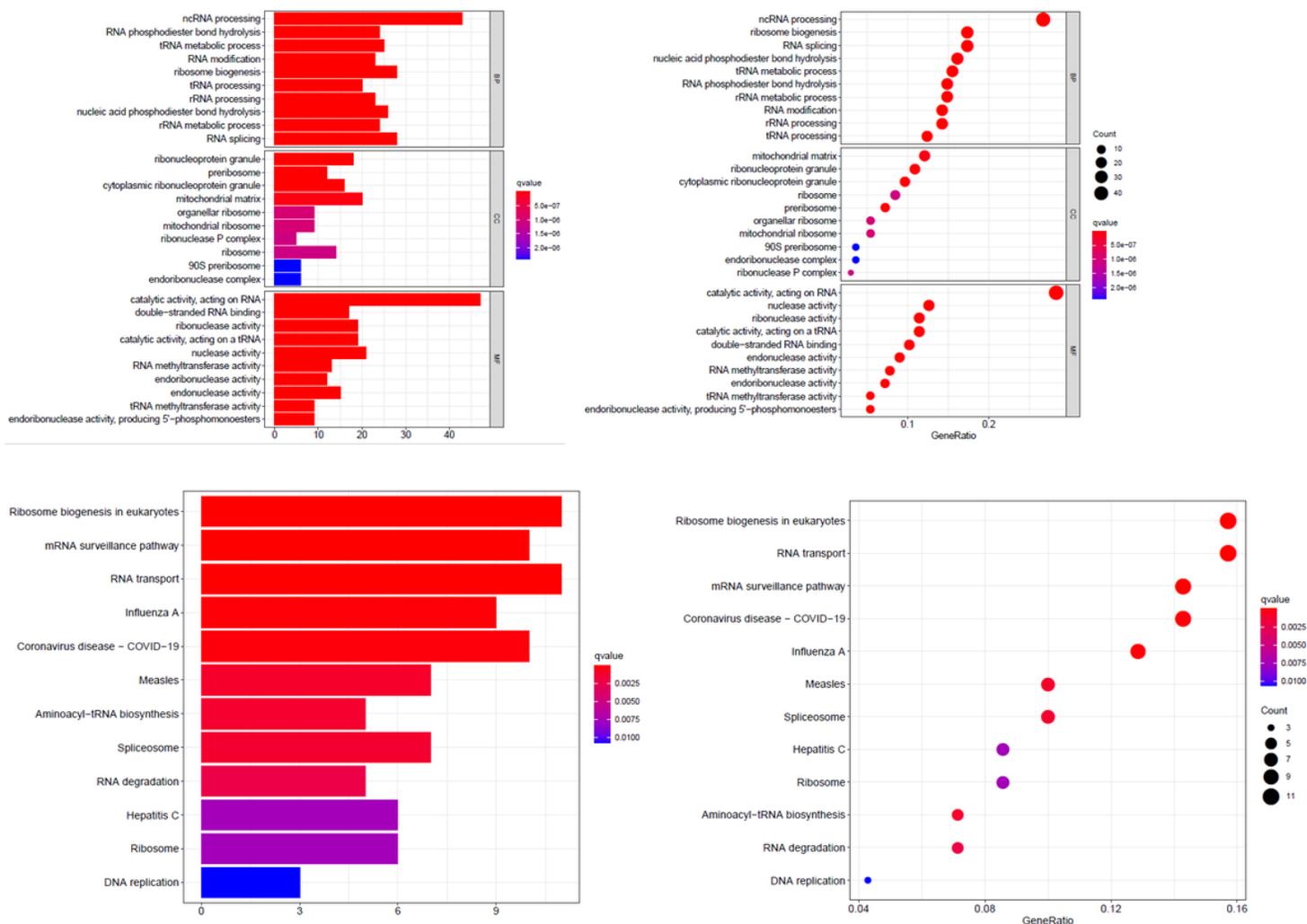
**Figure 1**

Expression level of DEGs (RNA binding proteins) between normal group and tumor group. (a). Expression levels of DEGs presented in heatmap. Down-regulated genes are presented in green and up-regulated genes are presented in red. (b). Expression levels and expression fold changes of DEGs presented in volcano plot. 176 down-regulated genes are presented as green dots; 171 up-regulated genes are presented as red dots.



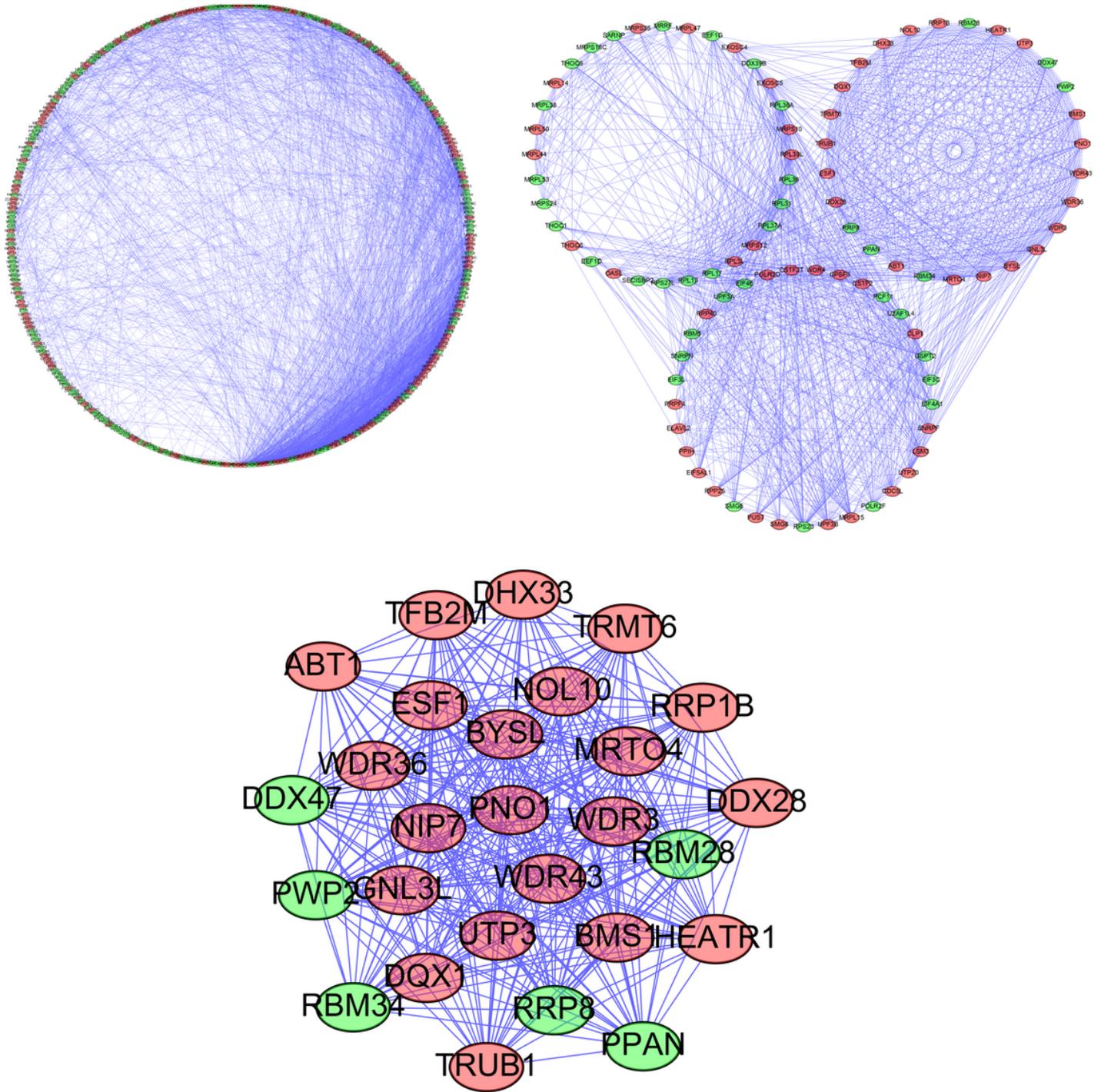
**Figure 2**

Down regulated DEGs were applied in GO and KEGG enrichment analysis. (a-b). GO enrichment analysis were shown in bubbles plot and bar blot respectively. (c-d). KEGG enrichment analysis were shown in bubbles plot and bar blot respectively. The significant degree of enrichment was measured by size of bubble and depth of color.



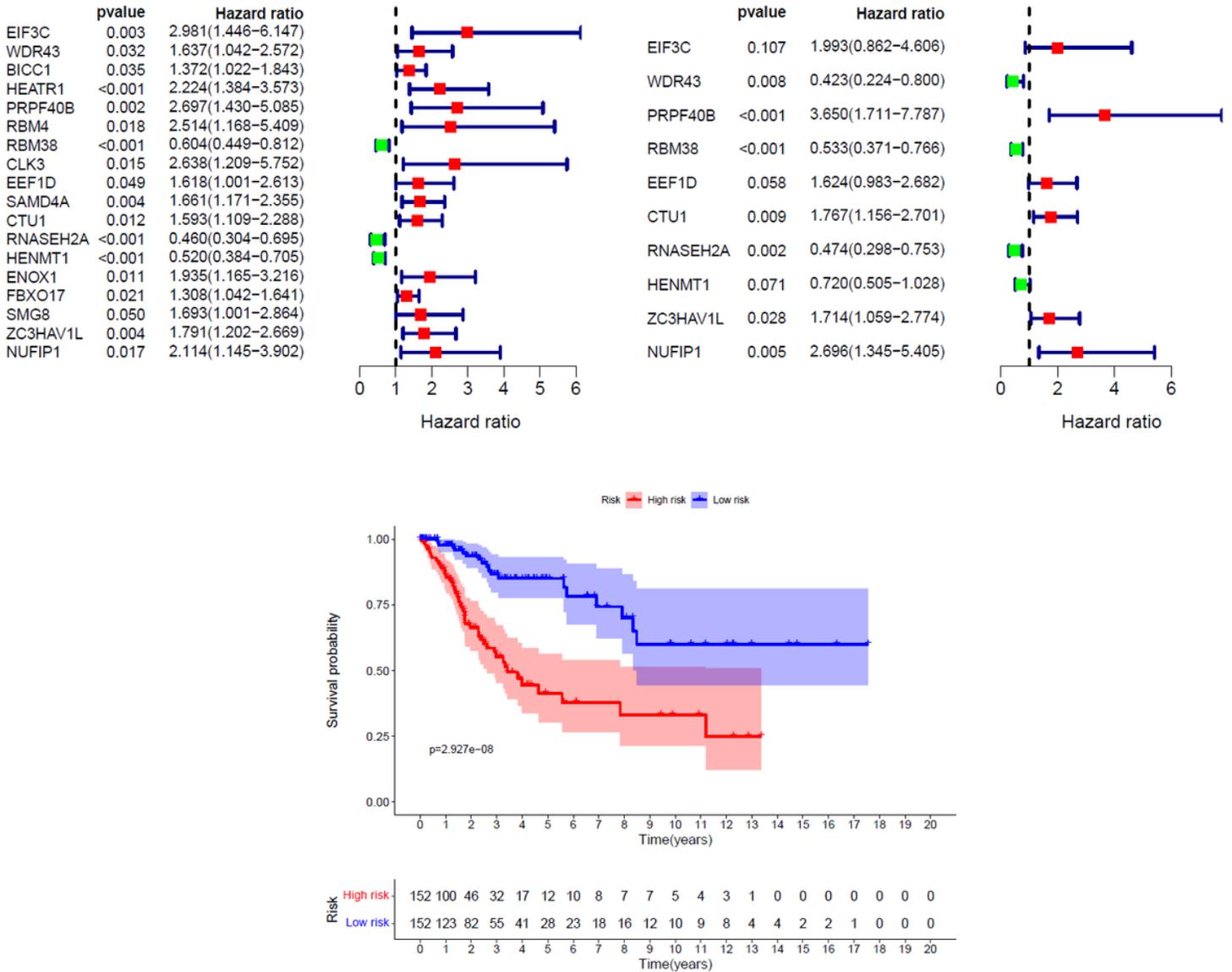
**Figure 3**

UP regulated DEGs were applied in GO and KEGG enrichment analysis. (a-b). GO enrichment analysis were shown in bubbles plot and bar blot respectively. (c-d). KEGG enrichment analysis were shown in bubbles plot and bar blot respectively. The significant degree of enrichment was measured by size of bubble and depth of color.



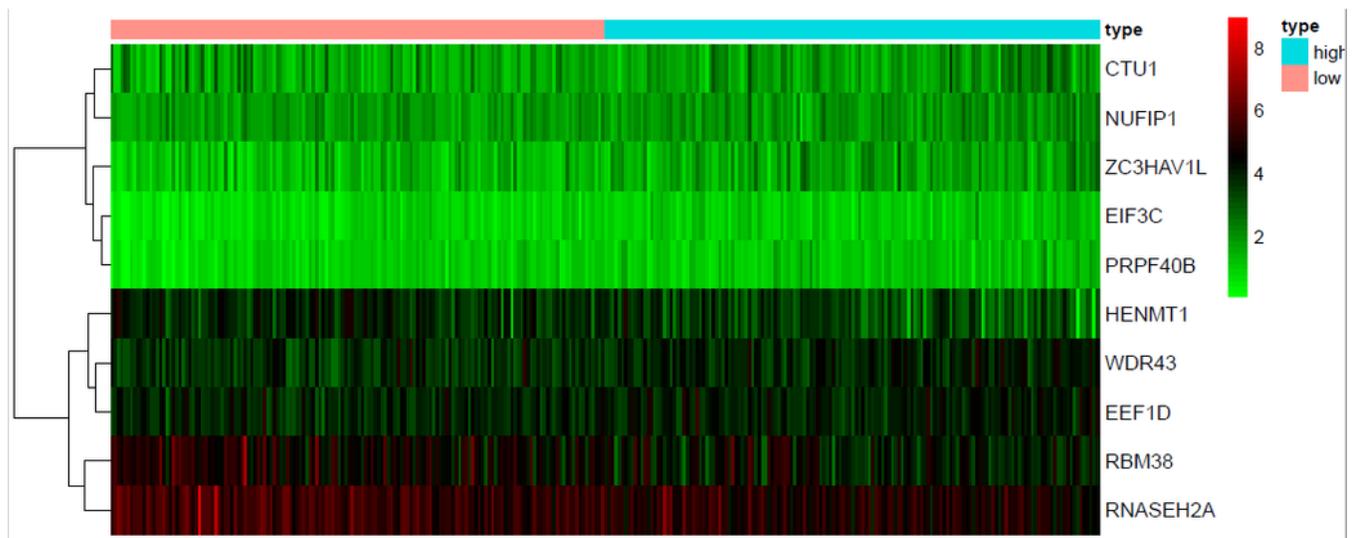
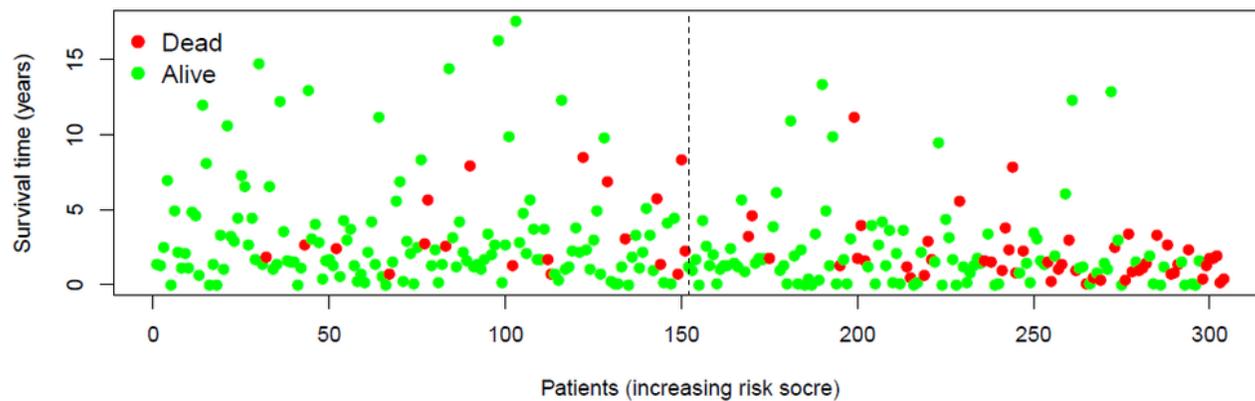
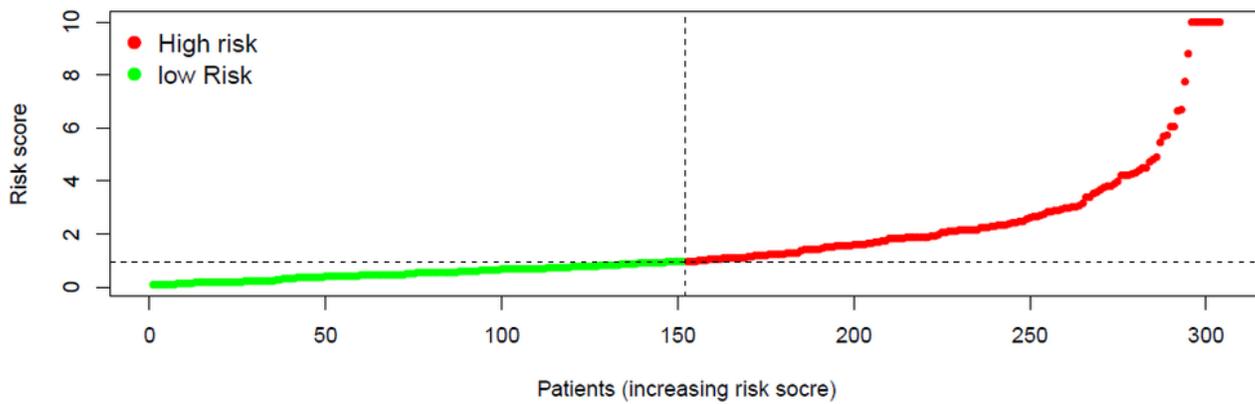
**Figure 4**

Construct PPI (Protein-protein interaction) network (a) PPI network of differentially expressed RBPs; RBPs were arranged in a circle. (b) MCODE plug-in sorted three most critical modules from PPI network; RBPs were arranged in three circle. Green circles stood for down-regulated RBPs in CC with a fold change of more than 1.41; Red circles stood for up-regulated RBPs with fold change more than 1.41.



**Figure 5**

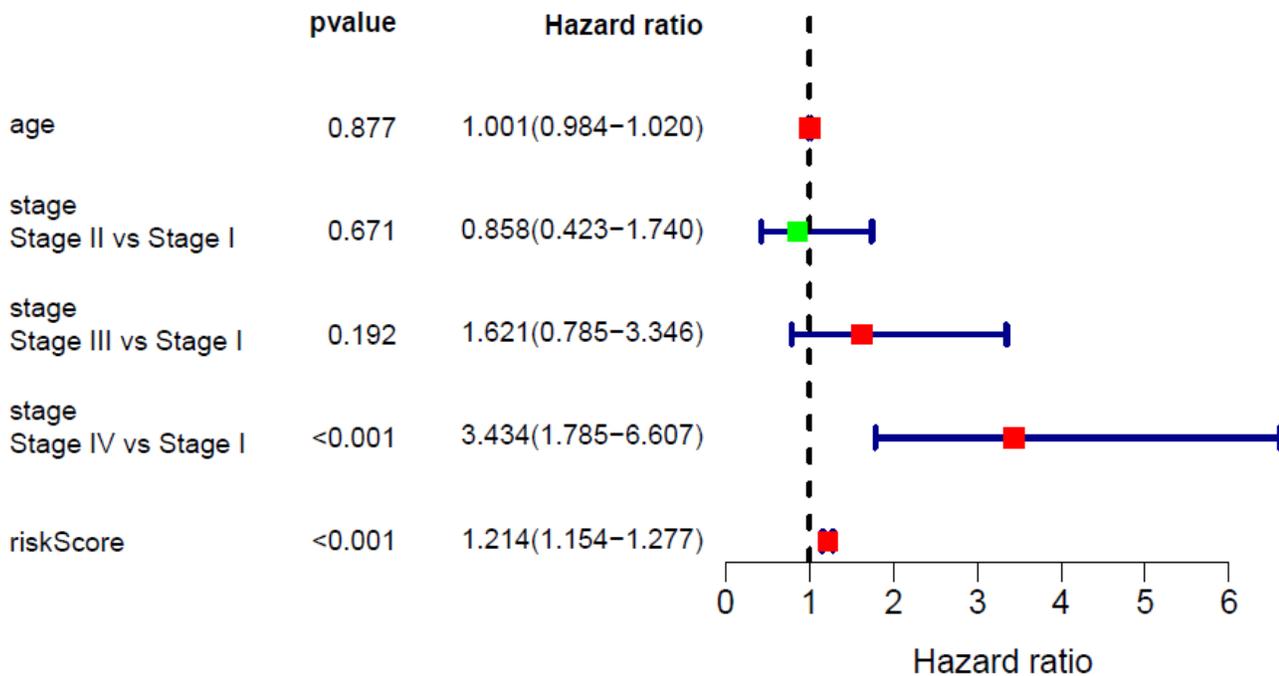
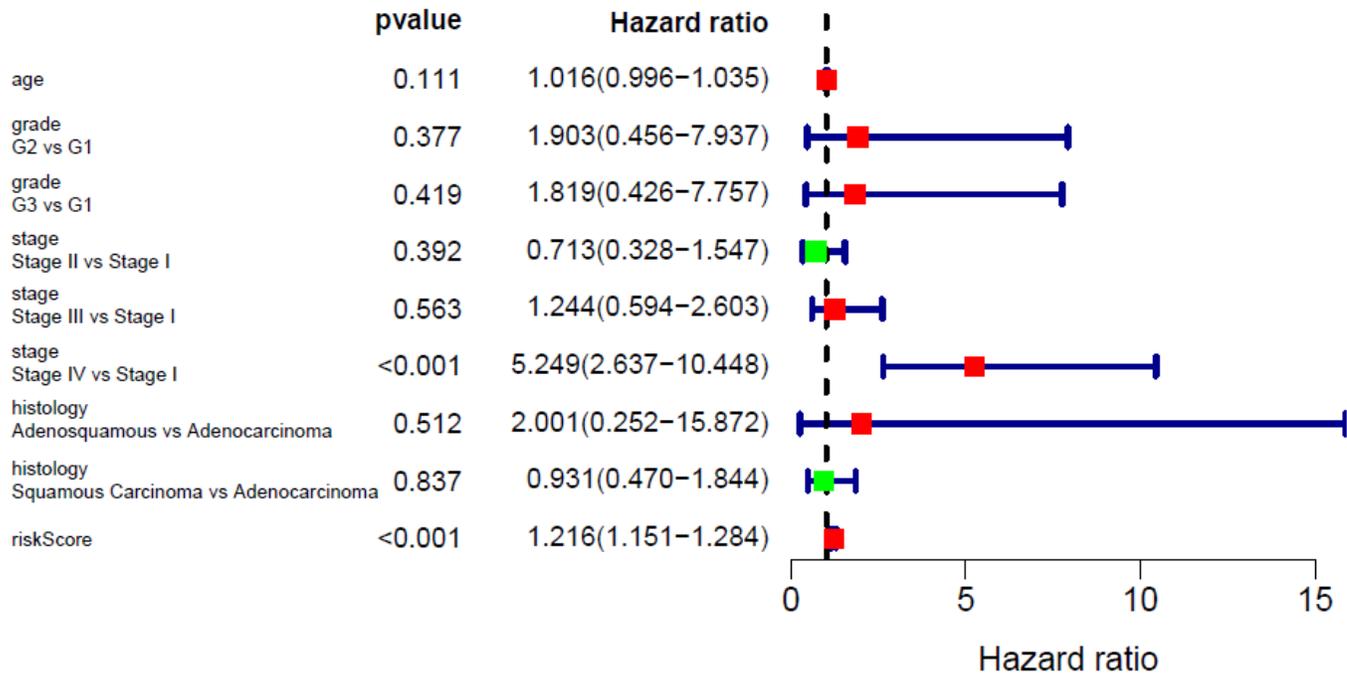
Forest plot of HR of DEGs and Kaplan-Meier curve for DEGs. (a). Forest plot of 18 prognosis-related DEGs retrieved by univariate cox regression. (b). Forest plot of 10 prognosis-related genes retrieved by multivariate cox regression model with AIC value. (c). KM curve for overall survival in the high-risk and the low-risk groups stratified by DEGs risk score.



**Figure 6**

CC patients in high and low risk groups are stratified by risk score counted by expression level of RBPs for analyse. (a). Risk score scatter plot of high risk group patients and low risk group patients. Dead patients were presented as red dots. Alive patients were presented as green dots. The survival time of dead patients decreased with the ascend of risk score. (b). The individual inflection point of the risk score curve was displayed by dotted line. It shows that patients were divided into low-risk and high-risk groups by

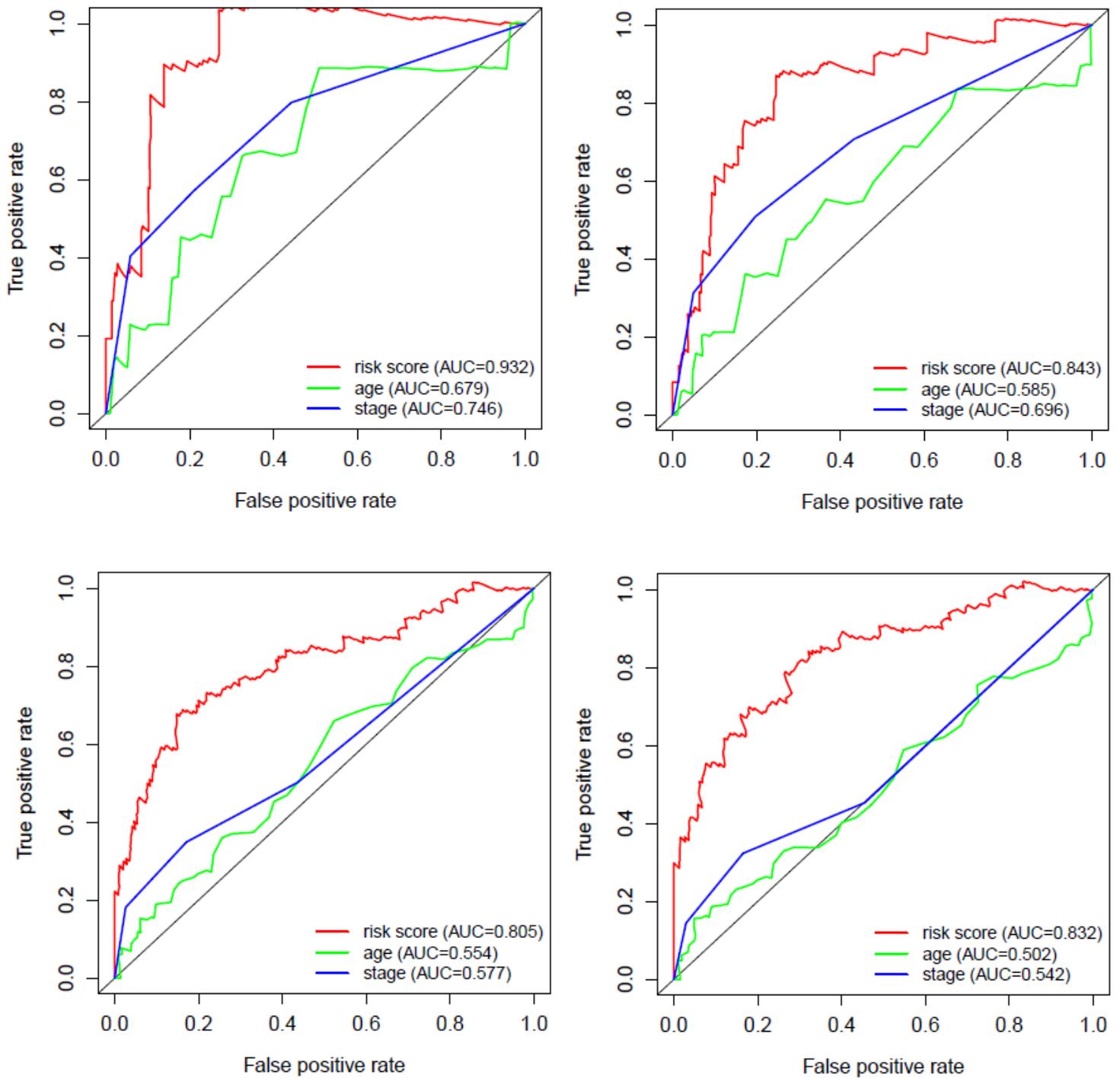
median of risk score. Red dots stood for patients with high risk. Green dots stood for patients with low risk. (c). Risk score heatmap of ten DEGs. The expression level of ten DEGs increased with color varied from green to red.



**Figure 7**

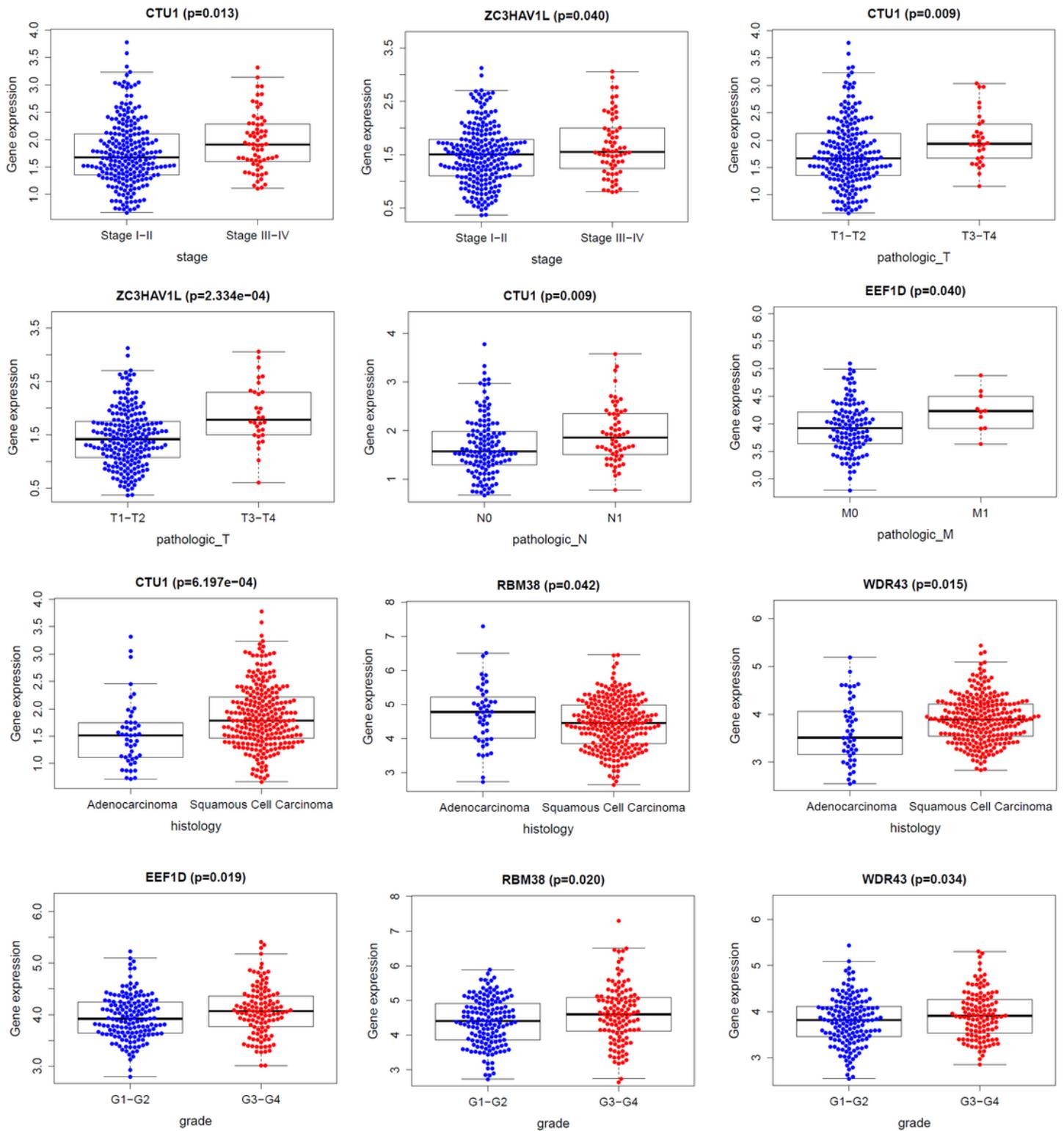
Forest plots of risk score and other clinical features. (a). forest plot for risk score and clinical features in univariate cox propotional risk regression model. (b). forest plot for risk score and clinical features in

multivariate cox propotional risk regression model.



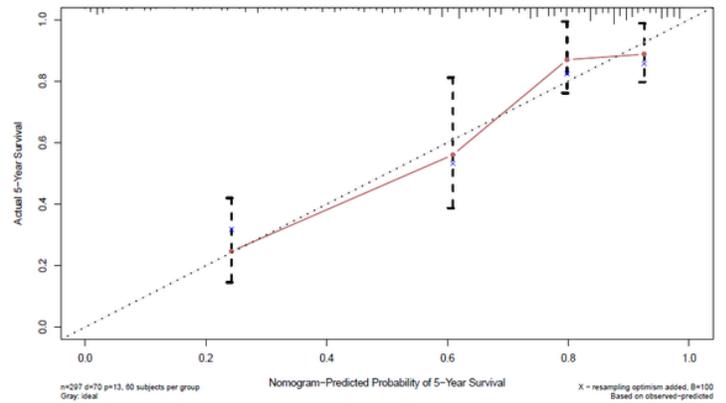
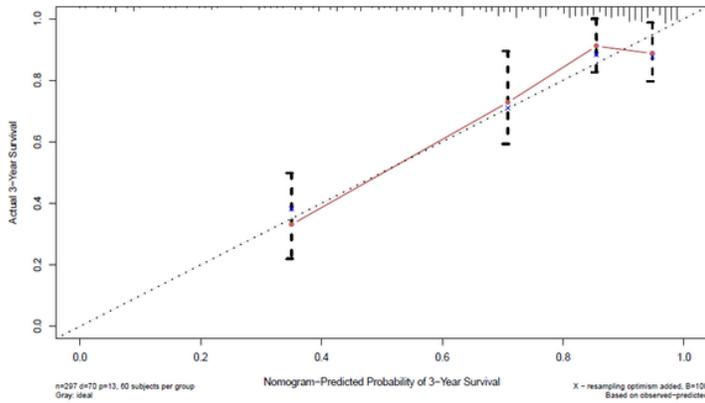
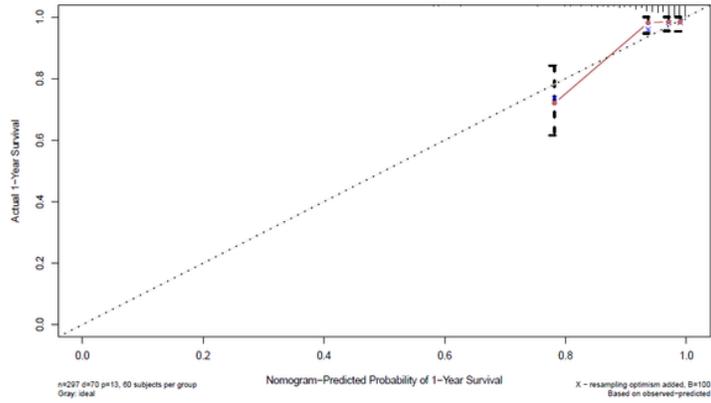
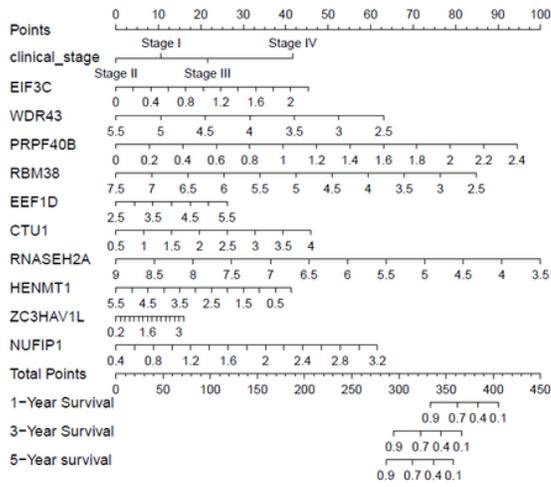
**Figure 8**

ROC curves for evaluating the discrimination of survival indicators (a). 0.5-year. (b). 1-year (c). 3-year. (d). 5-year. AUC: area under curve. The discrimination feasibility increased with the ascending of AUC.



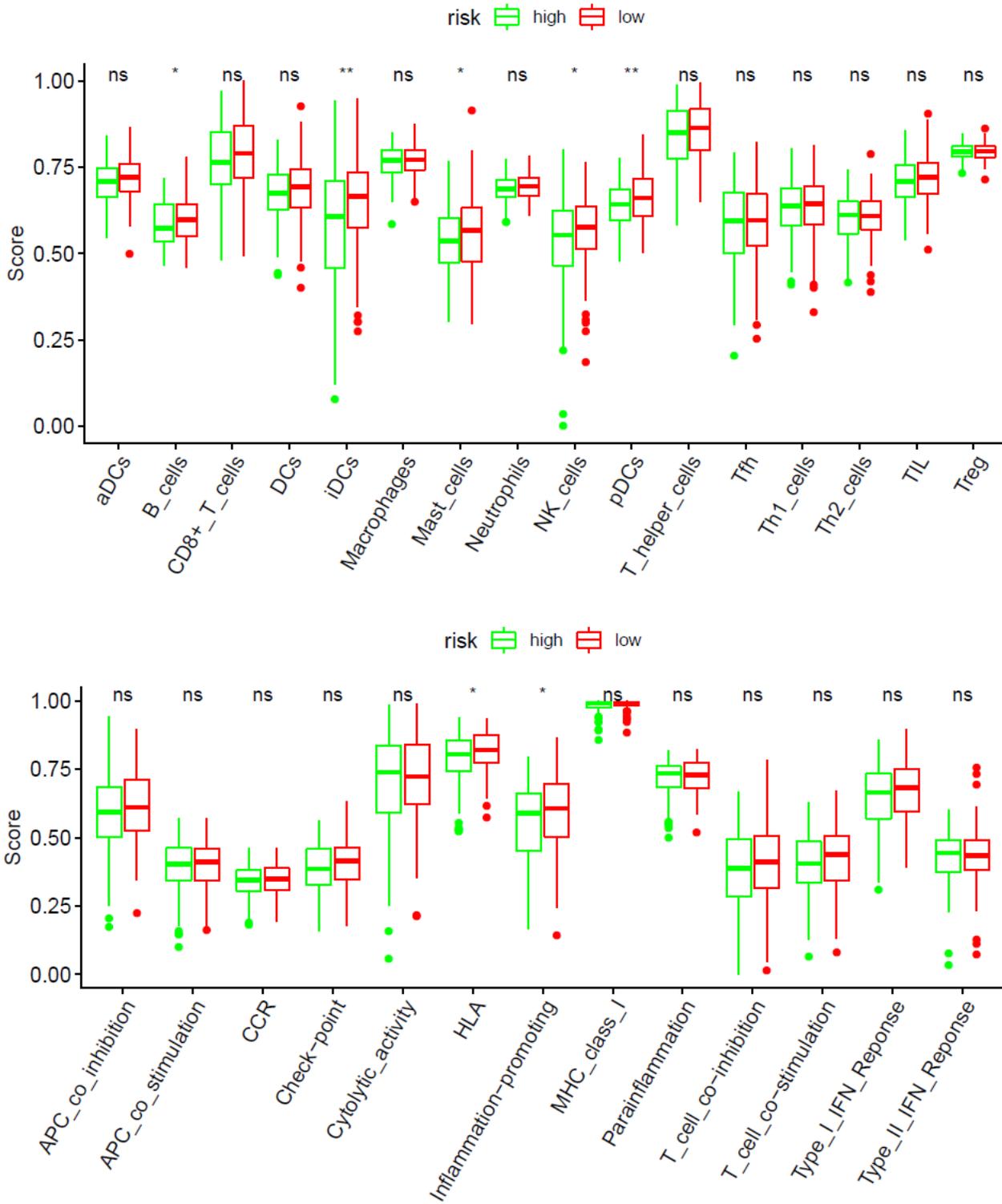
**Figure 9**

Box plot for displaying relationship between prognostic DEGs and clinical features. (a-j)



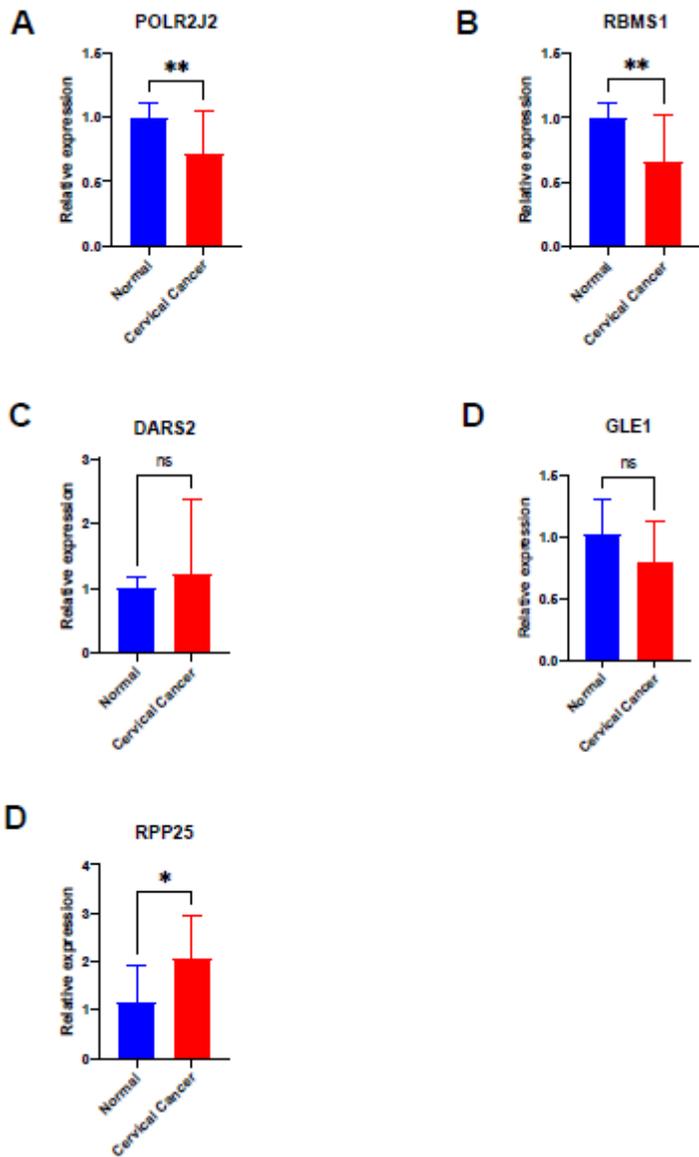
**Figure 10**

Prediction model constructed for CC patients. (a). Nomograph for CC with clinical stage and risk score which predict the survival of 1 year, 3 years and 5 years. (b). calibration curves of the prognostic nomogram prediction in the 1-year. (c). calibration curves of the prognostic nomogram prediction in the 3-year. (d). calibration curves of the prognostic nomogram prediction in the 5-year.



**Figure 11**

Box plot for ssGSEA immune score between the high and low risk groups categorised by median of risk score. (a) The scores of 16 immune cells (b) The scores of 13 immune-related functions. DCs dendritic cells, iDCs immature DCs, pDCs plasmacytoid dendritic cells, TIL tumor infiltrating lymphocyte, CCR cytokine-cytokine receptor, APC antigen presenting cells. Adjusted P values were shown as: ns, not significant; \*P < 0.05; \*\*P < 0.01; \*\*\*P < 0.001.



**Figure 12**

Relative expression level of differently expressed genes (RNA binding proteins) between normal tissues and cervical cancer tissues. (a) POLR2J2 (b) RBMS1 (c) DARS2 (d) GLE1 (e) RPP25

## Supplementary Files

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

- [table.S1.tcgapatientclinical.xls](#)
- [table.S2.xls](#)
- [table.S3.ownpatientclinical.xlsx](#)
- [table.S4.xls](#)
- [table.S5.txt](#)

- [table.S6.txt](#)
- [table.S7.txt](#)
- [table.S8.txt](#)
- [table.S9.qpcr1.xlsx](#)
- [table.S10.qpcr2.xlsx](#)
- [table.S11.qpcr3.xlsx](#)