Integrated analysis of RNA binding proteins in human colorectal cancer

Xuehui Fan
First Affiliated Hospital of Harbin Medical University
Lili Liu
First Affiliated Hospital of Harbin Medical University
Yue Shi
First Affiliated Hospital of Harbin Medical University
Fanghan Guo
First Affiliated Hospital of Harbin Medical University
Haining Wang
First Affiliated Hospital of Harbin Medical University
Xiuli Zhao
First Affiliated Hospital of Harbin Medical University
Di Zhong
First Affiliated Hospital of Harbin Medical University
Guozhong Li (lgzhyd1962@163.com)
First Affiliated Hospital of Harbin Medical University https://orcid.org/0000-0002-5482-393X

Research

Keywords: colorectal cancer(CRC), RNA-binding protein (RBP), Prognostic model construction, Survival analysis

Posted Date: June 9th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-34049/v1

Version of Record: A version of this preprint was published on August 22nd, 2020. See the published version at https://doi.org/10.1186/s12957-020-01995-5.

Abstract

Background: Though RNA-binding proteins play an essential role in a variety of different tumors, there is still limited effort made on the systematic analysis of RNA binding proteins (RBPs) in the survival of colorectal cancer(CRC).

Methods: An analysis was conducted of the CRC transcriptome data collected from the TCGA database and RBPs were extracted from CRC. R software was applied to analyze the differentially expressed genes (DEGs) of RBPs. To identify the pathways and functional annotation of RBPs DEGs, Gene Ontology (GO) function, Kyoto Encyclopedia of Genes and Genomes(KEGG) pathway enrichment analysis were carried out using the database for annotation, visualization and integrated discovery. The interactive gene database search tool was employed to conduct protein-protein interaction analysis, while Cytoscape software was applied for visual processing. Based on the COX regression analysis of the prognostic value of interacting RBPs and survival time, the RBPs related to survival were screened out, and a prognostic model was constructed. The data stored in TCGA database was taken as the Train group, while the chip data obtained from the GEO database was treated as the Test group to verify the model. Then, both survival analysis and ROC curve verification were conducted. Finally, the risk curves and nomograms of the two groups were drawn to predict the survival period.

Results: There were 314 genes up-regulated by RBPs DEGs and 155 genes down-regulated, from which twelve RBPs (NOP14, MRPS23, MAK16, TDRD6, POP1, TDRD5, TDRD7, PPARGC1A, LIN28B, CELF4, LRRFIP2, MSI2) with prognostic markers were obtained.

Conclusions: These twelve genes may be applicable as the predictor of CRC and play an essential role in the pathogenesis of CRC. In spite of this, it remains necessary to further explore the underlying mechanism.

Introduction

As a significant class of proteins in cells, RNA binding proteins (RBPs) can interact with RNA by recognizing the special RNA binding domains and are widely involved in multiple post-transcriptional regulatory processes such as RNA shearing, transport, sequence editing, intracellular localization and translation control[1]. It is estimated that there are up to 1500 proteins that have the potential to bind RNA in the human genome [2]. RBPs is characterized by the presence of an RNA binding domain (RBD) that contains 60-100 residues, which usually adopts αβ topology. Found in single or multiple copies, these domains usually bind to RNA depending on the exact sequence or structure [3]. Up to now, RBPs have been identified as associated with various human diseases, for example, spinal muscular atrophy and myotonic dystrophy [4]. There are various RBPs involved in tumorigenesis. SAM68 (SRC associated with 68 kDa mitosis) is classified into the STAR (signal transduction and RNA metabolism activation) family of RBPs. It is involved in several steps of mRNA metabolism, such as the transcription to alternative splicing and nuclear export. In addition, SAM68 is associated with the signal transduction pathways

required for the response of cells to stimuli, cell cycle transition and viral infection [5]. TARBP2 is over-expressed in metastatic cells and metastatic human breast tumours, the abnormal activation of which could cause the progression of breast carcinomas by affecting the stability of their target mRNA[6].

Colorectal cancer(CRC), which includes colon and rectal cancer, is known as a common digestive tract tumor. The molecular pathogenesis of CRC is a complex multistep process involving multiple acquired genetic and epigenetic abnormalities[7]. Some RBPs are known to have association with colorectal cancer. According to some research, Muscleblind-like 1 (MBNL1), an RBP implicated in developmental control, can suppress CRC cell metastasis significantly in vitro. MBNL1 destabilizes Snail transcripts and, thus inhibiting the Epithelial-Mesenchymal Transition (EMT) of CRC cells through the Snail/E-cadherin axis in vitro. RAS oncogene activation mutations are commonly seen in colon cancer[8].

In this study, an analysis was conducted of the RBP-related genes in CRC patients through differential gene expression and protein molecule interaction. Besides, the prognostic model was adopted to screen twelve genes in relation to the survial of CRC patients. We verified the model and performed survival analysis and risk assessment. These results are hoped to gain understanding as to the underlying mechanism related to the survival of CRC on a molecular level, thus providing a new direction for the prognosis of CRC and clinical treatment.

Methods

1. **Data source**

The FPKM transcriptome sequencing data of CRC was sourced from the TCGA database website (https://portal.gdc.cancer.gov/). The total number of samples is 521, of which the number of samples in the tumor group is 479 and that in the normal group is 42. Then, the RBP gene was obtained from the GOA database website (https://www.ebi.ac.uk/GOA/). Combined with the CRC transcriptome sequencing map, CRC RBPs were obtained. The data on gene expression (GSE17536) in colorectal patients was obtained from the GEO database website (https://www.ncbi.nlm.nih.gov/geo/), involving a total of 177 cases. All the data is publicly available online. This study requires no experiments to be conducted by any author on humans or animals. The flowchart of it is shown in Figure 1.

2. **Data processing of differentially expressed genes (DEGs)**

The RBPs were analyzed using R software to identify the difference between the tumor group and the sample group. Wilcox. Test was carried out to identify DEGs between the two groups, with the adjusted P□ 0.05 and |logFC| > 0.05

3. **GO and KEGG pathway analysis of DEGs**

GO analysis represents a common method applied for large-scale functional enrichment studies. Gene functions can be categorized into biological processes (BP), molecular functions (MF) and cellular

components (CC). KEGG is known as a commonly used database where a large amount of data on genomes, biological pathways, diseases, chemicals and drugs is stored. Through GO and KEGG analysis of DEGs, barplot and bubble were drawn respectively. All of the GO and pathway terms were ranked by their -log10 (q-value).

4. **Protein-protein interaction (PPI) network**

The Search Tool for the Retrieval of Interacting Genes (STRING) database (https://string-db.org/) is designed to analyze the PPI information. DEGs were inputted into the STRING database to obtain PPI information. Subsequently, the Cytoscape software was applied to visualize the PPI network, the cytoscape plug-in MCODE was used to obtain the most relevant sub-network module, and then the hub genes of the four modules were enriched for GO and KEGG analysis.

5. **Construction and analysis of prognostic models**

COX regression analysis was conducted of the prognostic value of 442 RBPs interacting with survival time, the RBPs related to survival were screened out, and a forest map was drawn. Then, the samples of the TCGA database were taken as the Train group and the samples of the GEO database were treated as the Test group, so as to construct the best prognostic model based on the Train group. There were twelve survival-related genes identified by the model, with which the correlation coefficient of each gene was obtained. Then, the risk score of each patient in the Train group and Test group was calculated according to gene expression. In the meantime, the patients were split into high-risk and low-risk groups by the median value of the risk score. The patients in the Train group and the Test group were categorized into either the high-risk group or low-risk group. The survival analysis was conducted, the ROC curve was verified, and then the risk curves of the Train and Test groups were drawn. Based on univariate and multivariate analyses, the nomograms based on the genes obtained from the prognostic model were drawn to predict the length of survival for the patients.

Results

1. **Identification of RBPs DEGs**

Transcriptome sequencing data of 1493 RBPs of CRC was obtained from the TCGA database. The differential expression analysis was conducted to find out that there were 314 up-regulated genes and 155 down-regulated genes, based on which volcano and heat maps were drawn as shown in Figure 2.

2. **Functional enrichment analyses of DEGs**

The up- and down-regulated genes of DEGS were analyzed for GO function and KEGG pathway enrichment, while both barplot and bubble were plotted. The enriched GO terms were divided into CC, BP, and MF ontologies. The top 10 most relevant items were selected, as shown in Figure 3. With regard to the up-regulated genome, the results of GO analysis indicated that DEGs were mainly enriched in BPs,

including ncRNA metabolic process, ncRNA processing, ribonucleoprotein complex biogenesis, and ribosome biogenesis and so on. CC analysis revealed that the DEGs were significantly enriched in preribosome, t-UTP complex, small-subunit processome, and cytoplasmic ribonucleoprotein granule and so on. As for the MF, the DEGs were enriched in catalytic activity, thus influencing RNA and ribonuclease activity. In the down-regulated genome, BP analysis demonstrated that the DEGs were significantly enriched, as reflected in the regulation of translation, RNA splicing, the regulation of cellular amide metabolic process and so on. CC analysis showed that the DEGs were significantly enriched in cytoplasmic ribonucleoprotein granule, ribonucleoprotein granule, cytoplasmic stress granule, etc. As for the MF, the DEGs were enriched in translation regulator activity, mRNA 3'-UTR binding and so on. Regarding the results of KEGG pathway analysis as shown in Figure 4, the DEGs in the up-regulated genome were primarily enriched in the pathways in Ribosome biogenesis in eukaryotes and RNA transport, etc. In the down-regulated genome, the DEGs were largely enriched in the pathways in Spliceosome and RNA transport, etc.

3. **PPI network construction**

The protein interactions among the DEGs were predicted using STRING tools. A total of 442 nodes and 6233 edges in the PPI network were obtained, as shown in Figure 5A. Then, Cytoscape software was applied to draw a network diagram of 442 genes, as shown in Figure 5B. Besides, four key sub-networks with the MCODE plug-in were extracted, GO was performed (Table 1) and KEGG enrichment analysis was conducted (Table 2) on the genes of the four sub-networks, respectively. Finally, the four sub-networks were visualized, as shown in Figure 5C-E. The number of hub genes in these 4 sub-networks is 61, 39, 6 and 6 respectively.

4. **Construction and analysis of prognostic models**

COX regression analysis was carried out of the prognostic value of 442 RBPs interacting with survival time, 19 RBPs related to survival were screened, and a forest map was drawn as shown in Figure 6A. Then, a prognostic model was constructed for the RBPs related to prognosis, and a prognostic marker gene comprised of 12 RBPs was established. These twelve genes are Nucleolar protein 14(NOP14), Mitochondrial ribosomal protein S23 (MRPS23), MAK16 homolog(MAK16), tudor domain containing 6 (TDRD6), processing of precursor 1(POP1), tudor domain containing 5(TDRD5), tudor domain containing 7(TDRD7), Peroxisome proliferator-activated receptor gamma coactivator 1-alpha(PPARGC1A), lin-28 homolog B (LIN28B), CUGBP Elav-like family member 4(CELF4), Leucine-rich repeat flightless-interacting protein 2(LRRFIP2), musashi RNA-binding protein 2 (MSI2). Then, the corresponding forest map was drawn for these twelve genes as shown in Figure 6B. Among them, TDRD5, ELF4 and LRRFIP2 are classed as high-risk genes, while the rest is classed as low-risk genes. Based on the established model, the risk value of each patient was calculated. According to the median value, the patients in the Train group and the Test group were divided into either a high-risk group or a low-risk group. Among them, the number of patients in the Train group as well as the high-risk group was 226. The number of patients in the low-risk group was 226. In the Test group, the number of patients in the high-risk group was 152 and

that of patients in the low-risk group was 25. The results indicated that the patients with high risk scores had a shorter survival time, as shown in Figure 6C, D. Finally, in terms of survival prediction, the ROC curve showed a relatively decent performance, as shown in Figure 6E, F. The AUC value in the Train group was 0.754 and the AUC value in the Test group was 0.553. Then, the risk curves were plotted for the Train and Test groups, as shown in Figure 7, which reveals that their abscissas are the same. They were divided into high and low risk groups by the median value. The patients were ranked by risk value in ascending order. The risk value of patients from left to right increased on a continued basis, as did the risk of fatality.

Then, independent prognostic analysis was conducted of univariate and multivariate for the Train and Test groups, as shown in Figure 8A-D. According to the results of single factor independent prognosis analysis, for the Train group and Test group, age and tumor stage can be treated as independent prognostic factors for the survival of colorectal patients ($p<0.05$). In the multivariate independent prognostic analysis, age and stage can be taken as independent prognostic factors for CRC in the Test group ($p<0.05$). For the Train group, however, only staging can be taken as independent prognostic factors for CRC ($p<0.01$), not age ($p=0.492$).

Finally, nomograms were plotted for these 12 RBP prognostic genes in the Train group to predict the survival time of the patients, as shown in Figure 8E. The RNA expression of 12 RBPs was applied as parameters to draw the point line in nomograms. The scores were added to obtain the total score, which can be used to predict the 1-year, 2-year and 3-year survival rates among CRC patients.

Discussion

As one of the most common malignant tumors, CRC is characterized by high recurrence rate and poor prognosis, especially in those developed countries. It is the third most common cancer among males and ranks second among females [9, 10]. So far, there have been various methods applied to predict the biomarkers of CRC prognosis [11]. RBPs are capable to regulate mRNA stability and contribute to cancer-associated pathways [12]. In this paper, the RBPs of CRC were analyzed. Through a series of analysis, there were 12 marker genes related to the prognosis of CRC obtained.

Tudor domain-containing (TDRD) refers to a family of evolutionarily conserved proteins. In general, PIWI and TDRD proteins are recognized as the major influencing factors in piRNA biogenesis and germ cell development [13]. In this study, it was found out that methyl lysine-bound TDRDs are primarily involved in histone modification and chromatin remodeling, while methyl arginine-bound TDRDs are usually associated with RNA metabolism, alternative splicing, small RNA pathways, and germ cell development [14, 15]. TDRDs have now been detected in various cancers. TDRD9 is highly expressed in a subset of non-small cell lung carcinomas and derived cell lines by hypomethylation of its CpG island [16]. TDRD1 is closely associated with ERG over-expression in primary prostate cancer [17]. According to the findings of Jiang et al. [18] , 7 TDRD genes (PHF20L1, ARIB4B, SETDB1, LBR, TDRKH, TDRD10, and TDRD5) show high levels of amplification in more than 10% of TCGA breast cancers. In liver cancer, TDRD5 has been

found to have a significant prognostic value. An early study revealed that TDRD5 is expressed in normal gastric and colonic mucosal tissues, suggesting the possibility that the TDRD5 gene is modified in CRC [19]. TDRD6 is capable to categorize irradiated prostate cancer patients into early and late relapse groups [20]. In addition, TDRD7 may play a certain role in the migration of tumor cells [21]. By analyzing CRCs, Mo et al. [22] have discovered not only frameshift mutations but also intratumoral heterogeneity of TDRD1, TDRD5 and TDRD9, which in combination might alter TDRD gene functions and make difference to the tumorigenesis of high microsatellite instability CRC. In our study, it was found out that TDRD5, TDRD6 and TDRD7 are differentially expressed in CRC, which makes it necessary to conduct a further study on the role of these three genes in colon cancer.

POP1 is referred to as a component of ribonuclease P, which is a ribonucleoprotein complex that generates mature tRNA molecules by cleaving their 5'-ends [23, 24]. In addition, it is a component of the MRP ribonuclease complex, which cleaves pre-rRNA sequences[25]. In the study, POP1 was found to be enriched in human prostate cancer cell lines [26], suggesting that it may be suitable as a potential marker for the diagnosis and prognosis of prostate cancer. Besides, it was found out that POP1 is up-regulated in CRC and applicable as a prognostic factor for CRC. Nevertheless, there is still no relevant research on the mechanism of POP1 in CRC, for which a further study is deemed necessary.

Kown as PGC1α, PPARGC1A is a transcriptional coactivator of genes encoding proteins responsible for the regulation of mitochondrial biogenesis and function[27]. D'Errico et al. [28] made a discovery that in the presence of Bax, the PGC1α-induced ROS accumulation represents one of the main apoptosis-driving factors in CRC cells. They also found out that PGC1α induces mitochondrial proliferation and activation in human intestinal cancer cells. [29]. Shin et al. [30] demonstrated that PGC-1α over-expression is effective in up-regulating the proliferation of HEK293 and CT26 cells. In addition, this expression exhibits a correlation with the enhancement of tumorigesis. In a case-control study, the heterozygous carriers of rs3774921 in PPARGC1A showed an increased risk of CRC [31]. PPARGC1A plays an essential role in the pathogenesis of colon cancer.

ZNF385A refers to a variety of RNA-binding protein that affects the localization and translation of a subset of mRNA. It binds the 3'-UTR of p53/TP53 mRNAs with ELAVL1 to control their nuclear export induced by CDKN2A. Thus, it has a potential to regulate p53/TP53 expression and mediate in part with the CDKN2A anti-proliferative activity[32]. The study led to a finding that ETV6-ZNF385A may be Acute lymphoblastic leukemia(ALL) new fusion gene [33]. Despite no ZNF385A involved in the pathogenesis of CRC, it may still play a certain role in CRC through the interaction with p53/TP53.

It is speculated that LRRFIP2 functions as activator of the canonical Wnt signaling pathway, which is associated with DVL3, the upstream of CTNNB1/beta-catenin. It regulates Toll-like receptor (TLR) signaling positively in response to agonist probably by competing with the negative FLII regulator for MYD88-binding, which plays a crucial role in the progression of colon cancer [34, 35]. In the study, it was found out that LRRFIP2 can be a candidate gene for the alternative splicing in colon and prostate cancer. There were three splice variants differing in their inclusion or skipping of exons 5 and/or 6 observed.

Containing five predicted putative serine phosphorylation sites and one putative O-glycosylation site, these exons could modulate LRRFIP2 protein function [36]. As a familial hereditary disease, hereditary nonpolyposis CRC (Lynch syndrome) is mainly caused by DNA mismatch (mismatch repair). In Lynch syndrome, Morak and colleagues discovered a paracentric inversion on chromosome 3p22.2 between the DNA mismatch repair gene, MLH1, and the downstream LRRFIP2 gene transcribed in antisense direction. This contributes to two new stable fusion transcripts, thus removing MLH1 gene and protein function [37]. Another study was conducted on a Lynch Syndrome family to find out that the MLH1.ITGA9 fusion allele cased loss of heterozygosity(LOH) occurred to five genes, with LRRFIP2 included, which resulted in the loss of mismatch repair capabilities [38]. It can be known from above that LRRFIP2 may play a critical role in the pathogenesis of CRC.

CELF4 is responsible for encoding a sort of protein with three domains to bind RNA-recognition motif and to regulate pre-mRNA alternative splicing. In some research, it was found out that CELF4 was hypermethylated in endometrial cancer. Methylated CELF4 may be suitable for endometrial cancer screening of cervical scrapings [39]. For CELF4, it remains necessary to conduct a further research to figure out its role in tumors.

As a member of the Musashi, MSI2 belongs to the family of Drosophila melanogaster RNA binding proteins. It has been identified as a critical regulator of haematopoietic stem cell (HSC) self-renewal and fate determination[40, 41]. In this study, MSI2 was found to be a central component in an unappreciated oncogenic pathway to promote intestinal transformation via the PDK–AKT–mTORC1 axis[42]. MSI2 is highly expressed in a variety of cancers, including hepatocellular carcinoma and lung cancer[43, 44]. As for colon cancer cell lines, some studies have been carried out recently to suggest that both USP10 and MSI2 proteins are up-regulated. Besides, USP10 could stabilize the oncogenic factor MSI2 through deubiquitination [45]. It was also found out that its expression is up-regulated in CRC, which makes it applicable as a prognostic marker gene for CRC.

NOP14 refers to a stress-responsive gene as required for 18S rRNA maturation and 40S ribosome production[46]. As indicated by Zhou et al. [47] , the NOP14 in pancreatic cancer cells is capable to promote motility, proliferation and metastatic capacity. According to the findings of Du et al. [48] , NOP14 primed tumor invasion and metastasis by improving the stability of mutp53 mRNA. By inhibiting the Wnt/β-catenin pathways, NOP14 suppresses breast cancer[49]. In addition, NOP14 can reduce melanoma cell proliferation and metastasis by regulating the Wnt/b-catenin signaling pathway30484495[50]. It was found out in this study that the expression of NOP14 was upregulated in CRC, which means its pathogenesis requires further research and confirmation.

MRPS23 gene, which is responsible for encoding a 28S subunit protein, was found to be over-expressed in breast cancer[51], uterine cervical cancer [52], hepatocellular carcinoma [53], colorectal[54] and uterine leiomyoma [55]. As revealed by Gao et al. [56] , inhibiting MRPS23 could lead to a significant reduction in breast cancer metastasis by inhibiting EMT phenotype. Though the expression of MRPS23 is increased in CRC, its specific pathogenesis remains unclear.

MAK16 encodes a ribosomal protein and plays an important role in ribosome biogenesis throughout the cell cycle [57]. In this study, it was found out that the mutations in MAK16 can induce the arrest of the cell cycle at G1 phase, during which the cell synthesizes mRNA and proteins as part of the preparation for cell division[58]. At present, there is still no role of MAK16 in the pathogenesis of tumors, which requires a further research to confirm.

LIN28, a heterochronic developmentally regulated RNA-binding protein, was originally identified in mutation studies of the nematode Caenorhabditis elegans [59]. LIN28B is a homologue of LIN28, was overexpressed in many cancer types [60, 61]. King et al. [62] found out that LIN28B over-expression is associated with the reduced survival time and increased probability of tumor recurrence for patients. The constitutive LIN28B expression promotes not only tumorigenesis, but also the induction of LGR5 and PROM1 in colonic epithelial cell[63]. In addition, LIN28B could promote the proliferation, clone formation and tumorigenesis of colon cancer cells by increasing BCL-2 expression [64]. It thus can be known that LIN28B is highly expressed in CRC and plays an important role in its pathogenesis. It was found suitable as a target gene for CRC prognosis.

In this paper, a discussion was conducted about the role of 12 genes in tumors. Though some genes were found irrelevant to the pathogenesis of CRC, their biological functions and changes in their expression in CRC suggest that they may play a role in CRC to some extent, which requires a lot of experiments to be conducted for verification. This is also the limitation that our study is subject to. It is hoped that more research can be done to explore the pathogenesis of CRC.

Conclusions

In summary, 12 prognostic RBPs were obtained through TCGA database analysis, including NOP14, MRPS23, MAK16, TDRD6, POP1, TDRD5, TDRD7, PPARGC1A, LIN28B, CELF4, LRRFIP2, and MSI2, which were then verified through the sample data obtained from the GEO database. In CRC, NOP14, MRPS23, MAK16, TDRD6, POP1, TDRD5, LIN28B, MSI2 were up-regulated, while TDRD7, PPARGC1A, CELF4, LRRFIP2 were down-regulated. These genes are related to the prognosis of CRC. More research is deemed necessary to verify the specific function of each gene, especially experimental studies. Our findings may improve the understanding of the incidence and prognosis of CRC, thus providing reference for the further exploration of the diagnosis and treatment of CRC.

Tables

Table 1. The GO function enrichment analysis of four most significant MCODE components

Sub-netwok 1

ONTOLOGY	ID	Description	Count	pvalue	p.adjust
BP	GO:0042254	ribosome biogenesis	46	1.57E-74	5.06E-72
BP	GO:0016072	rRNA metabolic process	43	7.30E-73	1.18E-70
BP	GO:0006364	rRNA processing	42	2.21E-71	2.38E-69
CC	GO:0030684	preribosome	27	5.03E-51	1.71E-49
CC	GO:0034455	t-UTP complex	18	2.51E-33	4.26E-32
CC	GO:0032040	small-subunit processome	15	5.16E-29	5.84E-28
MF	GO:0140098	catalytic activity, acting on RNA	20	8.42E-19	6.90E-17
MF	GO:0003724	RNA helicase activity	12	2.62E-17	1.08E-15
MF	GO:0030515	snoRNA binding	8	1.60E-14	4.38E-13

Sub-network 2

ONTOLOGY	ID	Description	Count	pvalue	p.adjust
BP	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	21	7.37E-26	1.03E-23
BP	GO:0000398	mRNA splicing, via spliceosome	21	7.37E-26	1.03E-23
BP	GO:0000375	RNA splicing, via transesterification reactions	21	8.71E-26	1.03E-23
CC	GO:0071013	catalytic step 2 spliceosome	13	6.97E-22	6.62E-20
CC	GO:0000974	Prp19 complex	13	2.03E-21	9.66E-20
CC	GO:0005682	U5 snRNP	13	3.10E-19	9.83E-18
MF	GO:0090079	translation regulator activity, nucleic acid binding	10	2.82E-14	2.23E-12
MF	GO:0003743	translation initiation factor activity	8	1.54E-13	4.37E-12
MF	GO:0008135	translation factor activity, RNA binding	9	1.66E-13	4.37E-12

Sub-network 3

ONTOLOGY	ID	Description	Count	pvalue	p.adjust
BP	GO:0000460	maturation of 5.8S rRNA	6	5.04E-18	5.24E-16
BP	GO:0034427	nuclear-transcribed mRNA catabolic process, exonucleolytic, 3'-5'	4	6.22E-13	3.23E-11
BP	GO:0043629	ncRNA polyadenylation	4	1.47E-12	3.67E-11
CC	GO:1905354	exoribonuclease complex	6	2.17E-18	3.69E-17
CC	GO:0000176	nuclear exosome (RNase complex)	5	1.50E-15	1.27E-14
CC	GO:0000178	exosome (RNase complex)	5	1.03E-14	5.82E-14
MF	GO:0017091	AU-rich element binding	3	7.07E-08	8.18E-07
MF	GO:0000175	3'-5'-exoribonuclease activity	3	1.29E-07	8.18E-07
MF	GO:0016896	exoribonuclease activity, producing 5'-phosphomonoesters	3	1.54E-07	8.18E-07

Sub-network 4

ONTOLOGY	ID	Description	Count	pvalue	p.adjust
BP	GO:0051028	mRNA transport	6	2.64E-13	1.90E-11
BP	GO:0050657	nucleic acid transport	6	1.13E-12	2.23E-11
BP	GO:0050658	RNA transport	6	1.13E-12	2.23E-11
CC	GO:0000346	transcription export complex	3	5.69E-09	9.11E-08
CC	GO:0016607	nuclear speck	4	2.35E-06	1.88E-05
CC	GO:0000784	nuclear chromosome, telomeric region	2	0.000588204	0.003137089

Table 2. The KEGG function enrichment analysis of four most significant MCODE components

List	ID	Description	Count	pvalue	p.adjust
Sub-network1	hsa03008	Ribosome biogenesis in eukaryotes	19	1.84E-32	3.68E-32
Sub-network2	hsa03040	Spliceosome	13	4.08E-13	2.85E-12
	hsa03013	RNA transport	12	1.00E-10	3.51E-10
	hsa03015	mRNA surveillance pathway	7	5.47E-07	1.28E-06
	hsa03010	Ribosome	5	0.001767902	0.003093829
Sub-network3	hsa03018	RNA degradation	5	4.75E-10	4.75E-10

Abbreviations

RBPs: RNA binding proteins; CRC: colorectal cancer; DEGs: differentially expressed genes; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; MBNL1: Muscleblind-like 1; EMT: Epithelial-Mesenchymal Transition; BP: biological processes; MF: molecular functions; CC: cellular components; NOP14: Nucleolar protein 14; MRPS23: Mitochondrial ribosomal protein S23; MAK16: MAK16 homolog; MAK16: TDRD6: tudor domain containing 6; POP1: processing of precursor 1; TDRD5: tudor domain containing 5; TDRD7: tudor domain containing 7; PPARGC1A: Peroxisome proliferator-activated receptor gamma coactivator 1-alpha; LIN28B: lin-28 homolog B; CELF4:CUGBP Elav-like family member 4; LRRFIP2: Leucine-rich repeat flightless-interacting protein 2; MSI2: musashi RNA-binding protein 2; TDRD:Tudor domain-containing; HSC: haematopoietic stem cell.

Declarations

Availability of data and materials

The datasets supporting the conclusion of this article are included within the article.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

No conflict or financial interests.

Author contributions

Xuehui Fan wrote the article; Lili Liu, Yue Shi and Fanghan Guo processed the data analysis; Haining Wang, Xiuli Zhao, Di Zhong conceived of this study; Guozhong Li revised the final manuscript

Funding

The National Natural Science Foundation (Grant No.81873746)

References

1. Glisovic T, Bachorik JL, Yong J, Dreyfuss G: RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett 2008, 582:1977-1986.
2. Gerstberger S, Hafner M, Tuschl T: A census of human RNA-binding proteins. Nat Rev Genet 2014, 15:829-845.
3. Lunde BM, Moore C, Varani G: RNA-binding proteins: modular design for efficient function. Nat Rev Mol Cell Biol 2007, 8:479-490.
4. Lukong KE, Chang KW, Khandjian EW, Richard S: RNA-binding proteins in human genetic disease. Trends Genet 2008, 24:416-425.
5. Frisone P, Pradella D, Di Matteo A, Belloni E, Ghigna C, Paronetto MP: SAM68: Signal Transduction and RNA Metabolism in Human Cancer. Biomed Res Int 2015, 2015:528954.
6. Goodarzi H, Zhang S, Buss CG, Fish L, Tavazoie S, Tavazoie SF: Metastasis-suppressor transcript destabilization through TARBP2 binding of mRNA hairpins. Nature 2014, 513:256-260.
7. Fearon ER, Vogelstein B: A genetic model for colorectal tumorigenesis. Cell 1990, 61:759-767.
8. Schubbert S, Shannon K, Bollag G: Hyperactive Ras in developmental disorders and cancer. Nat Rev Cancer 2007, 7:295-308.
9. Kraus S, Nabiochtchikov I, Shapira S, Arber N: Recent advances in personalized colorectal cancer research. Cancer Lett 2014, 347:15-21.
10. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM: Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010, 127:2893-2917.
11. Akagi Y, Kinugasa T, Adachi Y, Shirouzu K: Prognostic significance of isolated tumor cells in patients with colorectal cancer in recent 10-year studies. Mol Clin Oncol 2013, 1:582-592.
12. Perron G, Jandaghi P, Solanki S, Safisamghabadi M, Storoz C, Karimzadeh M, Papadakis AI, Arseneault M, Scelo G, Banks RE, et al: A General Framework for Interrogation of mRNA Stability Programs Identifies RNA-Binding Proteins that Govern Cancer Transcriptomes. Cell Rep 2018, 23:1639-1650.
13. Gan B, Chen S, Liu H, Min J, Liu K: Structure and function of eTudor domain containing TDRD proteins. Crit Rev Biochem Mol Biol 2019, 54:119-132.
14. Chen C, Nott TJ, Jin J, Pawson T: Deciphering arginine methylation: Tudor tells the tale. Nat Rev Mol Cell Biol 2011, 12:629-642.

15. Lu R, Wang GG: Tudor: a versatile family of histone methylation 'readers'. Trends Biochem Sci 2013, 38:546-555.
16. Guijo M, Ceballos-Chávez M, Gómez-Marín E, Basurto-Cayuela L, Reyes JC: Expression of TDRD9 in a subset of lung carcinomas by CpG island hypomethylation protects from DNA damage. Oncotarget 2018, 9:9618-9631.
17. Boormans JL, Korsten H, Ziel-van der Made AJ, van Leenders GJ, de Vos CV, Jenster G, Trapman J: Identification of TDRD1 as a direct target gene of ERG in primary prostate cancer. Int J Cancer 2013, 133:335-345.
18. Jiang Y, Liu L, Shan W, Yang ZQ: An integrated genomic analysis of Tudor domain-containing proteins identifies PHD finger protein 20-like 1 (PHF20L1) as a candidate oncogene in breast cancer. Mol Oncol 2016, 10:292-302.
19. Yoon H, Lee H, Kim HJ, You KT, Park YN, Kim H, Kim H: Tudor domain-containing protein 4 as a potential cancer/testis antigen in liver cancer. Tohoku J Exp Med 2011, 224:41-46.
20. Seifert M, Peitzsch C, Gorodetska I, Börner C, Klink B, Dubrovska A: Network-based analysis of prostate cancer cell lines reveals novel marker gene candidates associated with radioresistance and patient relapse. PLoS Comput Biol 2019, 15:e1007460.
21. Ito A, Mimae T, Yamamoto YS, Hagiyama M, Nakanishi J, Ito M, Hosokawa Y, Okada M, Murakami Y, Kondo T: Novel application for pseudopodia proteomics using excimer laser ablation and two-dimensional difference gel electrophoresis. Lab Invest 2012, 92:1374-1385.
22. Mo HY, Choi EJ, Yoo NJ, Lee SH: Mutational alterations of TDRD 1, 4 and 9 genes in colorectal cancers. Pathol Oncol Res 2020.
23. Lygerou Z, Pluk H, van Venrooij WJ, Séraphin B: hPop1: an autoantigenic protein subunit shared by the human RNase P and RNase MRP ribonucleoproteins. Embo j 1996, 15:5936-5948.
24. Wu J, Niu S, Tan M, Huang C, Li M, Song Y, Wang Q, Chen J, Shi S, Lan P, Lei M: Cryo-EM Structure of the Human Ribonuclease P Holoenzyme. Cell 2018, 175:1393-1404.e1311.
25. Goldfarb KC, Cech TR: Targeted CRISPR disruption reveals a role for RNase MRP RNA in human preribosomal RNA processing. Genes Dev 2017, 31:59-71.
26. Romanuik TL, Ueda T, Le N, Haile S, Yong TM, Thomson T, Vessella RL, Sadar MD: Novel biomarkers for prostate cancer including noncoding transcripts. Am J Pathol 2009, 175:2264-2276.
27. Mulinari S, Davis C: Why European and United States drug regulators are not speaking with one voice on anti-influenza drugs: regulatory review methodologies and the importance of 'deep' product reviews. Health Res Policy Syst 2017, 15:93.
28. D'Errico I, Lo Sasso G, Salvatore L, Murzilli S, Martelli N, Cristofaro M, Latorre D, Villani G, Moschetta A: Bax is necessary for PGC1α pro-apoptotic effect in colorectal cancer cells. Cell Cycle 2011, 10:2937-2945.
29. D'Errico I, Salvatore L, Murzilli S, Lo Sasso G, Latorre D, Martelli N, Egorova AV, Polishuck R, Madeyski-Bengtson K, Lelliott C, et al: Peroxisome proliferator-activated receptor-gamma coactivator

1-alpha (PGC1alpha) is a metabolic regulator of intestinal epithelial cell fate. Proc Natl Acad Sci U S A 2011, 108:6603-6608.

30. Shin SW, Yun SH, Park ES, Jeong JS, Kwak JY, Park JI: Overexpression of PGC-1α enhances cell proliferation and tumorigenesis of HEK293 cells through the upregulation of Sp1 and Acyl-CoA binding protein. Int J Oncol 2015, 46:1328-1342.
31. Cho YA, Lee J, Oh JH, Chang HJ, Sohn DK, Shin A, Kim J: Genetic variation in PPARGC1A may affect the role of diet-associated inflammation in colorectal carcinogenesis. Oncotarget 2017, 8:8550-8558.
32. Das S, Raj L, Zhao B, Kimura Y, Bernstein A, Aaronson SA, Lee SW: Hzf Determines cell survival upon genotoxic stress by modulating p53 transactivation. Cell 2007, 130:624-637.
33. Chabi B, Fouret G, Lecomte J, Cortade F, Pessemesse L, Baati N, Coudray C, Lin L, Tong Q, Wrutniak-Cabello C, et al: Skeletal muscle overexpression of short isoform Sirt3 altered mitochondrial cardiolipin content and fatty acid composition. J Bioenerg Biomembr 2018, 50:131-142.
34. Liu J, Bang AG, Kintner C, Orth AP, Chanda SK, Ding S, Schultz PG: Identification of the Wnt signaling activator leucine-rich repeat in Flightless interaction protein 2 by a genome-wide functional analysis. Proc Natl Acad Sci U S A 2005, 102:1927-1932.
35. Dai P, Jeong SY, Yu Y, Leng T, Wu W, Xie L, Chen X: Modulation of TLR signaling by multiple MyD88-interacting partners including leucine-rich repeat Fli-I-interacting proteins. J Immunol 2009, 182:3450-3460.
36. Thorsen K, Sørensen KD, Brems-Eskildsen AS, Modin C, Gaustadnes M, Hein AM, Kruhøffer M, Laurberg S, Borre M, Wang K, et al: Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. Mol Cell Proteomics 2008, 7:1214-1224.
37. Morak M, Koehler U, Schackert HK, Steinke V, Royer-Pokora B, Schulmann K, Kloor M, Höchter W, Weingart J, Keiling C, et al: Biallelic MLH1 SNP cDNA expression or constitutional promoter methylation can hide genomic rearrangements causing Lynch syndrome. J Med Genet 2011, 48:513-519.
38. Meyer C, Brieger A, Plotz G, Weber N, Passmann S, Dingermann T, Zeuzem S, Trojan J, Marschalek R: An interstitial deletion at 3p21.3 results in the genetic fusion of MLH1 and ITGA9 in a Lynch syndrome family. Clin Cancer Res 2009, 15:762-769.
39. Huang RL, Su PH, Liao YP, Wu TI, Hsu YT, Lin WY, Wang HC, Weng YC, Ou YC, Huang TH, Lai HC: Integrated Epigenomics Analysis Reveals a DNA Methylation Panel for Endometrial Cancer Detection Using Cervical Scrapings. Clin Cancer Res 2017, 23:263-272.
40. Ito T, Kwon HY, Zimdahl B, Congdon KL, Blum J, Lento WE, Zhao C, Lagoo A, Gerrard G, Foroni L, et al: Regulation of myeloid leukaemia by the cell-fate determinant Musashi. Nature 2010, 466:765-768.
41. Park SM, Deering RP, Lu Y, Tivnan P, Lianoglou S, Al-Shahrour F, Ebert BL, Hacohen N, Leslie C, Daley GQ, et al: Musashi-2 controls cell fate, lineage bias, and TGF-β signaling in HSCs. J Exp Med 2014, 211:71-87.
42. Wang S, Li N, Yousefi M, Nakauka-Ddamba A, Li F, Parada K, Rao S, Minuesa G, Katz Y, Gregory BD, et al: Transformation of the intestinal epithelium by the MSI2 RNA-binding protein. Nat Commun 2015,

6:6517.

43. He L, Zhou X, Qu C, Hu L, Tang Y, Zhang Q, Liang M, Hong J: Musashi2 predicts poor prognosis and invasion in hepatocellular carcinoma by driving epithelial-mesenchymal transition. J Cell Mol Med 2014, 18:49-58.
44. Li L, Yu H, Wang X, Zeng J, Li D, Lu J, Wang C, Wang J, Wei J, Jiang M, Mo B: Expression of seven stem-cell-associated markers in human airway biopsy specimens obtained via fiberoptic bronchoscopy. J Exp Clin Cancer Res 2013, 32:28.
45. Ouyang SW, Liu TT, Liu XS, Zhu FX, Zhu FM, Liu XN, Peng ZH: USP10 regulates Musashi-2 stability via deubiquitination and promotes tumour proliferation in colon cancer. FEBS Lett 2019, 593:406-413.
46. Liu PC, Thiele DJ: Novel stress-responsive genes EMG1 and NOP14 encode conserved, interacting proteins required for 40S ribosome biogenesis. Mol Biol Cell 2001, 12:3644-3657.
47. Zhou B, Wu Q, Chen G, Zhang TP, Zhao YP: NOP14 promotes proliferation and metastasis of pancreatic cancer cells. Cancer Lett 2012, 322:195-203.
48. Du Y, Liu Z, You L, Hou P, Ren X, Jiao T, Zhao W, Li Z, Shu H, Liu C, Zhao Y: Pancreatic Cancer Progression Relies upon Mutant p53-Induced Oncogenic Signaling Mediated by NOP14. Cancer Res 2017, 77:2661-2673.
49. Lei JJ, Peng RJ, Kuang BH, Yuan ZY, Qin T, Liu WS, Guo YM, Han HQ, Lian YF, Deng CC, et al: NOP14 suppresses breast cancer progression by inhibiting NRIP1/Wnt/β-catenin pathway. Oncotarget 2015, 6:25701-25714.
50. Li J, Fang R, Wang J, Deng L: NOP14 inhibits melanoma proliferation and metastasis by regulating Wnt/β-catenin signaling pathway. Braz J Med Biol Res 2018, 52:e7952.
51. Gatza ML, Silva GO, Parker JS, Fan C, Perou CM: An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer. Nat Genet 2014, 46:1051-1059.
52. Lyng H, Brøvig RS, Svendsrud DH, Holm R, Kaalhus O, Knutstad K, Oksefjell H, Sundfør K, Kristensen GB, Stokke T: Gene expressions and copy numbers associated with metastatic phenotypes of uterine cervical cancer. BMC Genomics 2006, 7:268.
53. Pu M, Wang J, Huang Q, Zhao G, Xia C, Shang R, Zhang Z, Bian Z, Yang X, Tao K: High MRPS23 expression contributes to hepatocellular carcinoma proliferation and indicates poor survival outcomes. Tumour Biol 2017, 39:1010428317709127.
54. Staub E, Gröne J, Mennerich D, Röpcke S, Klamann I, Hinzmann B, Castanos-Velez E, Mann B, Pilarsky C, Brümmendorf T, et al: A genome-wide map of aberrantly expressed chromosomal islands in colorectal cancer. Mol Cancer 2006, 5:37.
55. Li B, Zhang YL: Identification of up-regulated genes in human uterine leiomyoma by suppression subtractive hybridization. Cell Res 2002, 12:215-221.
56. Gao Y, Li F, Zhou H, Yang Y, Wu R, Chen Y, Li W, Li Y, Xu X, Ke C, Pei Z: Down-regulation of MRPS23 inhibits rat breast cancer proliferation and metastasis. Oncotarget 2017, 8:71772-71781.

57. Kater L, Thoms M, Barrio-Garcia C, Cheng J, Ismail S, Ahmed YL, Bange G, Kressler D, Berninghausen O, Sinning I, et al: Visualizing the Assembly Pathway of Nucleolar Pre-60S Ribosomes. Cell 2017, 171:1599-1610.e1514.
58. Vicuña L, Fernandez MI, Vial C, Valdebenito P, Chaparro E, Espinoza K, Ziegler A, Bustamante A, Eyheramendy S: Adaptation to Extreme Environments in an Admixed Human Population from the Atacama Desert. Genome Biol Evol 2019, 11:2468-2479.
59. Moss EG, Lee RC, Ambros V: The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. Cell 1997, 88:637-646.
60. Viswanathan SR, Powers JT, Einhorn W, Hoshida Y, Ng TL, Toffanin S, O'Sullivan M, Lu J, Phillips LA, Lockhart VL, et al: Lin28 promotes transformation and is associated with advanced human malignancies. Nat Genet 2009, 41:843-848.
61. Cao D, Allan RW, Cheng L, Peng Y, Guo CC, Dahiya N, Akhi S, Li J: RNA-binding protein LIN28 is a marker for testicular germ cell tumors. Hum Pathol 2011, 42:710-718.
62. King CE, Cuatrecasas M, Castells A, Sepulveda AR, Lee JS, Rustgi AK: LIN28B promotes colon cancer progression and metastasis. Cancer Res 2011, 71:4260-4268.
63. King CE, Wang L, Winograd R, Madison BB, Mongroo PS, Johnstone CN, Rustgi AK: LIN28B fosters colon cancer migration, invasion and transformation through let-7-dependent and -independent mechanisms. Oncogene 2011, 30:4185-4193.
64. Yuan L, Tian J: LIN28B promotes the progression of colon cancer by increasing B-cell lymphoma 2 expression. Biomed Pharmacother 2018, 103:355-361.

Figures

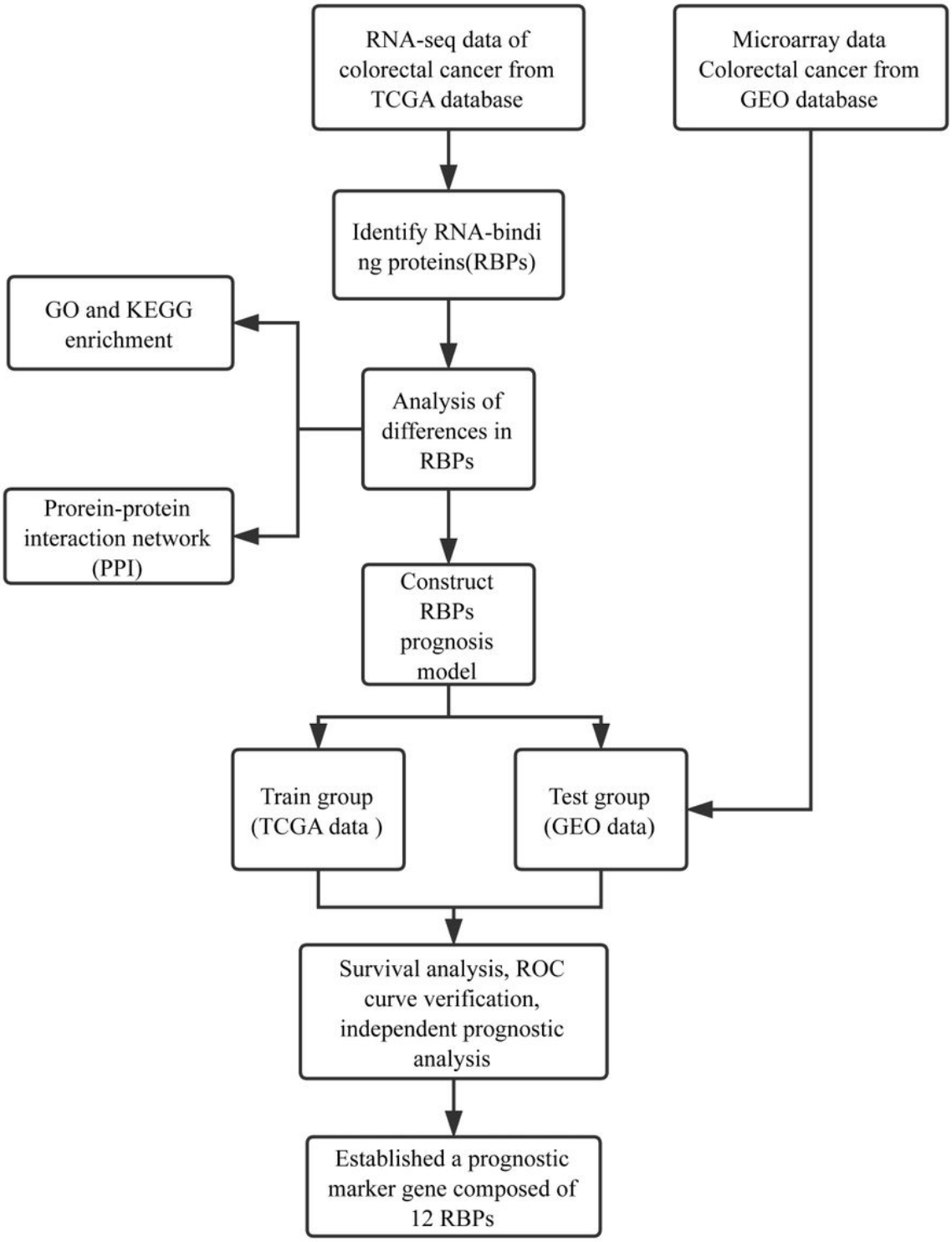

Figure 1

Flowchart of systemic analysis of RNA-binding protein in patients with CRC

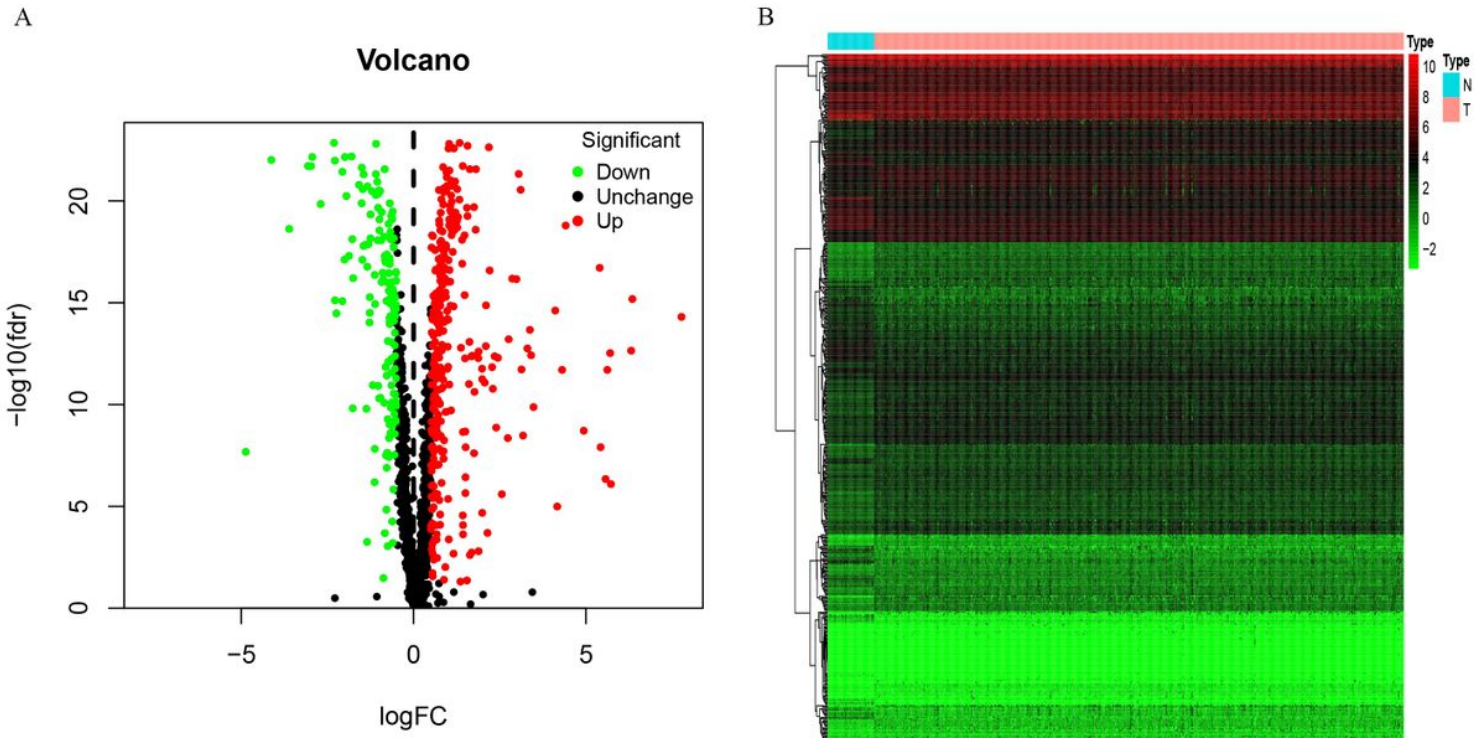

Figure 2

Volcano and heat map of RNA-binding protein DEGs A. Volcano map, B. Heat map. Red nodes represent upregulated genes, and green nodes represent downregulated genes.

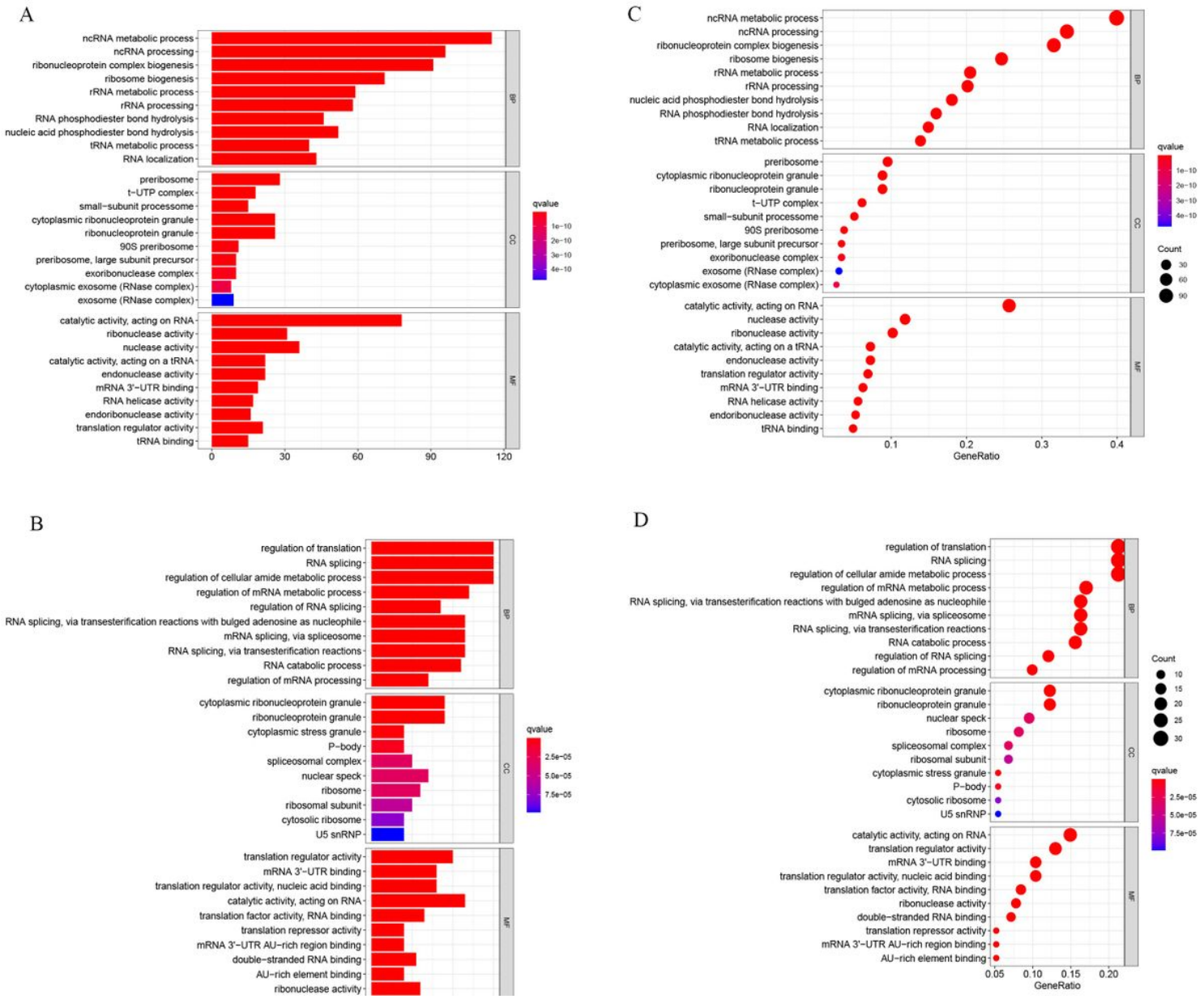

Figure 3

The gene ontology of analyses 477 RNA binding protein DEGs A. Barplot shows GO functional enrichment analysis predicted up-regulated DEGs, including biological process, cellular components and molecular functions. The color indicates the significance of the p-value. B. Bubble shows GO functional enrichment analysis predicted up-regulated DEGs. The size of the circle represents the number of genes enriched in the entry, and the color indicates the significance of the p-value. C. Barplot shows GO functional enrichment analysis predicted down-regulated DEGs. D. Bubble shows GO functional enrichment analysis predicted down-regulated DEGs.

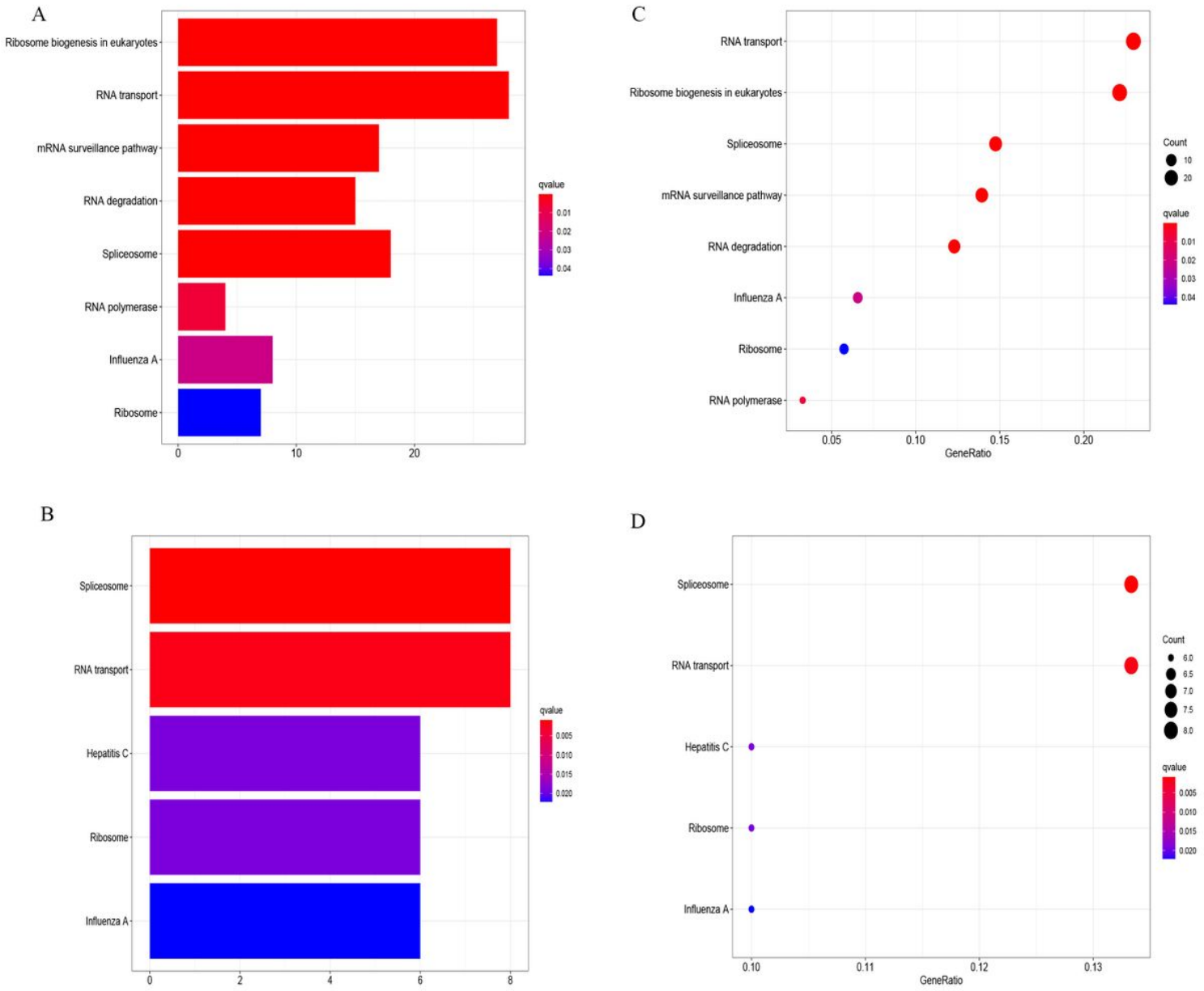

Figure 4

The KEGG pathway enrichment of analyses 477 RNA binding protein DEGs A. Barplot shows KEGG pathway analysis predicted up-regulated DEGs. The color indicates the significance of the p-value. B. Bubble shows KEGG pathway analysis predicted up-regulated DEGs. The size of the circle represents the number of genes enriched in the entry, and the color indicates the significance of the p-value. C. Barplot shows KEGG pathway analysis predicted down-regulated DEGs. D. Bubble shows KEGG pathway analysis predicted down-regulated DEGs.

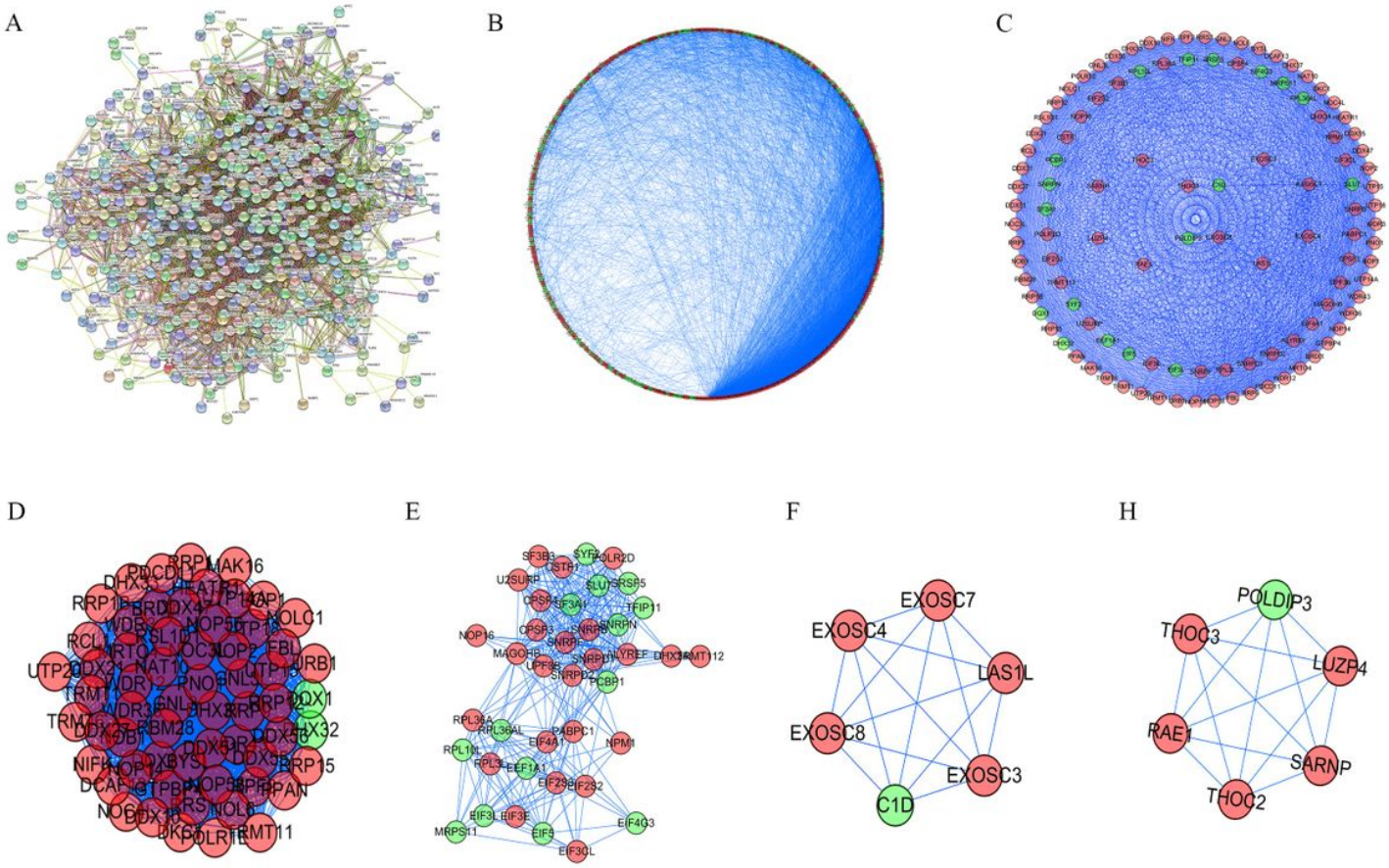

Figure 5

RNA-binding proteins DEGs are used to construct protein-protein interaction networks and subnetworks A. PPI interaction network map obtained from STRING website B. Cytoscape visualizes the genes of the interacting PPI network. Red nodes represent upregulated genes, while blue nodes refer to downregulated genes. C. Four MCODE modules visualization D-E. Four most significant MCODE components form the PPI network.

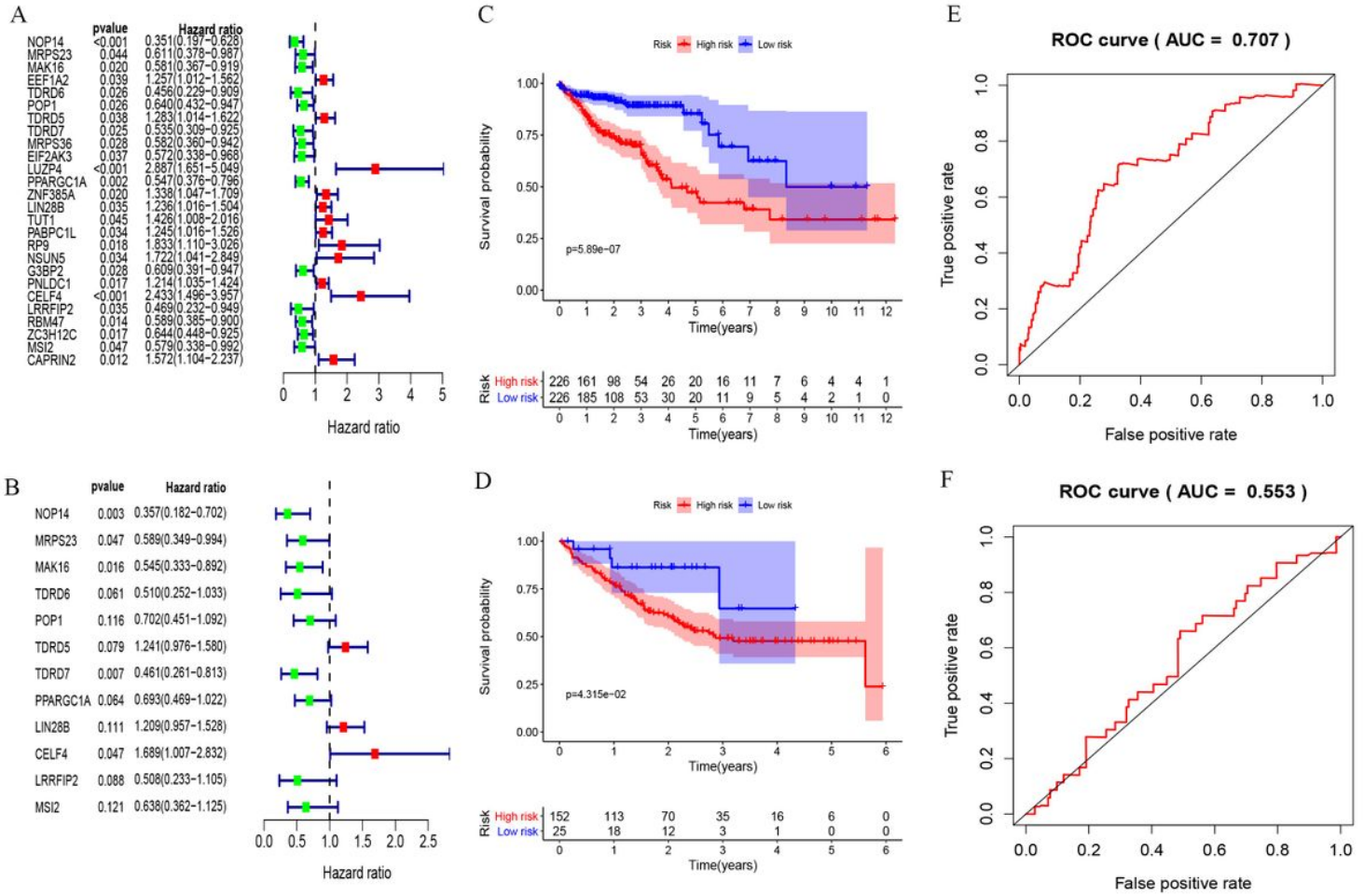

Figure 6

RNA-binding protein DGEs are used to construct prognostic models, survival analysis and verification of GEO data sets A. The 19 prognostic-related RBPs shown in the forest map, red indicates high-risk genes and green denotes low-risk genes. B. The 12 RBPs obtained by constructing the prognostic model shown in the forest map. C. Survival analysis curve of Train group, red indicates patients in high-risk group, blue denotes patients in low-risk group D. Survival analysis curve of Test group E. ROC curve of Train group F. ROC curve of Test group

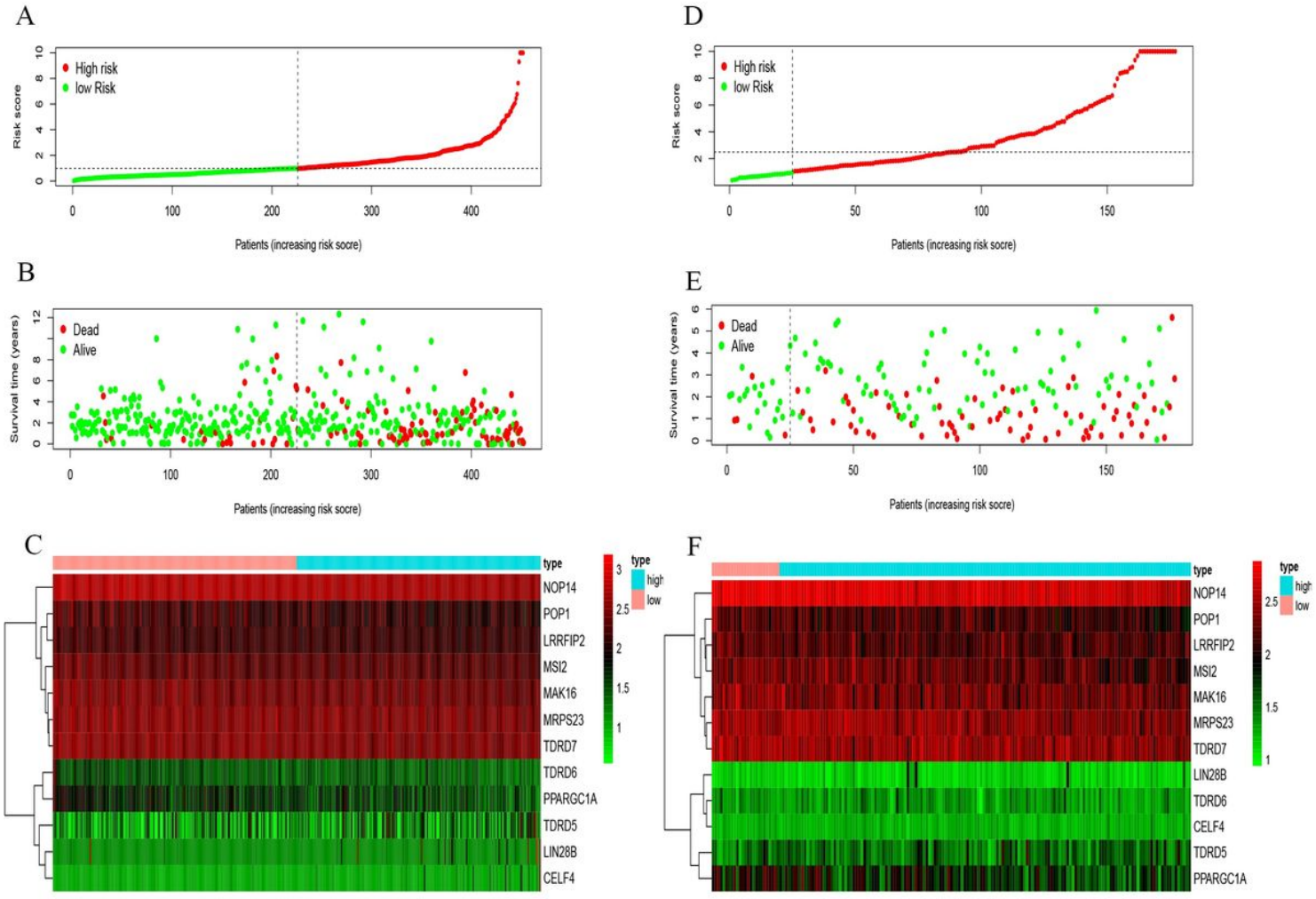

Figure 7

Risk curve of Train group and Test group A. the risk score distribution of Train group B. The distribution of survival status for Train group C. In train group,the heat map of 12 RBPs for the high- and low-risk groups D. the risk score distribution of Test group E. The survival status distribution for Test group F. In Test group, the heat map of 12 RBPs for the high- and low-risk groups

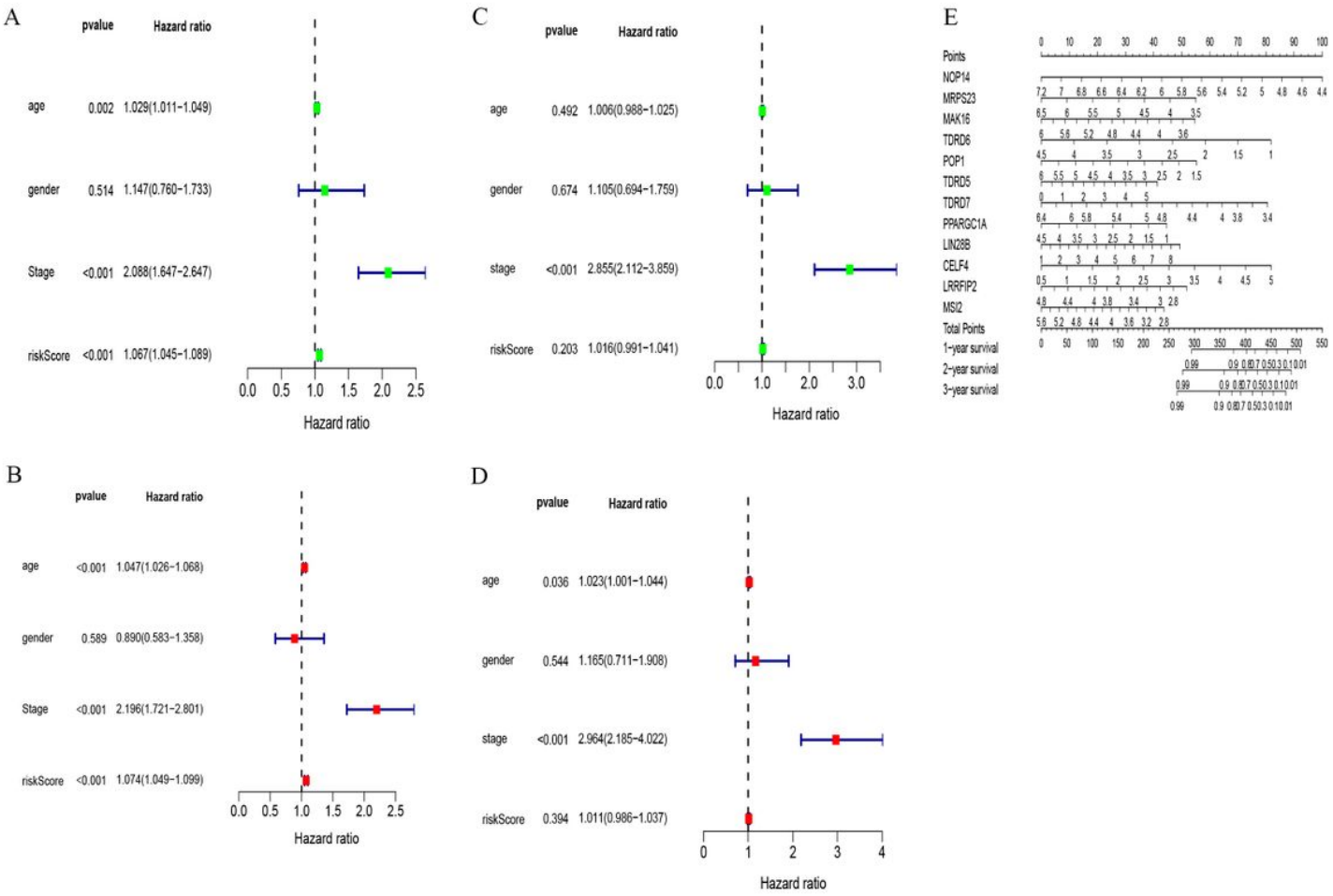

Figure 8

Independent prognosis analysis and prediction of 1, 2, and 3 years of nomograms of CRC patients in the Train and Test groups A. Single factor prognosis analysis of Train group B. Multi-factor prognosis analysis of Train group C. Single factor prognosis analysis of Test group D. Multi-factor prognostic analysis of Test group E. The nomograms for predicting 1-year, 2-year, and 3-year survival probability of patients with CRC for Train group.