Development and validation of a novel five-gene-based RNA binding protein associated prognostic model for human colon cancer

Haoling Liu

First Affiliated Hospital of Harbin Medical University

Qingquan Bai

First Affiliated Hospital of Harbin Medical University

Zhaoyang Lu

First Affiliated Hospital of Harbin Medical University

Xuan Song

First Affiliated Hospital of Harbin Medical University

Ye Jin

First Affiliated Hospital of Harbin Medical University

Yao Liu

First Affiliated Hospital of Science and Technology of China

Fanzheng Meng

First Affiliated Hospital of Science and Technology of China

Hongrui Guo

First Affiliated Hospital of Harbin Medical University

Tiemin Pei

First Affiliated Hospital of Harbin Medical University

Shuhang Liang

First Affiliated Hospital of Harbin Medical University

Ruipeng Song

First Affiliated Hospital of Science and Technology of Chian

Yingjian Liang

First Affiliated Hospital of Harbin Medical University

Jiabei Wang

First Affiliated Hospital of Science and Technology of China

Dalong Yin

First Affiliated Hospital of Science and Technology of China

Zhibo Qu

Fourth Affiliated Hospital of Harbin Medical University

Huaxin Guo

First Affiliated Hospital of Harbin Medical University

Hongchi Jiang

First Affiliated Hospital of Harbin Medical University

Haiyan Yang (✉ yanghaiyanhmu@163.com)

First Affiliated Hospital of Harbin Medical University <https://orcid.org/0000-0002-1870-4511>

Research article

Keywords: RNA binding protein, prognostic model, genome map database, bioinformatics analysis, colon cancer

Posted Date: August 19th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-56351/v1>

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

[Read Full License](#)

Abstract

Background

Dysregulation of RNA binding protein (RBP) expression has been reported in various malignant tumors, and it is related to the occurrence and development of cancer. However, the role of RBPs in colon cancer remains unclear.

Methods

We downloaded the RNA sequencing data of colon cancer from The Cancer Genome Atlas (TCGA) database, and determined the differently expressed RBPs between normal and cancer tissues. Then, through a series of bioinformatics analysis, we systematically studied the expression and prognostic value of these RBPs.

Result

A total of 490 different expression differently expressed RBPs were identified, including 323 up-regulated and 167 down. regulated RBPs. Five RBPs (PNLDC1, NSUN6, NOL3, PPARGC1A, LRRFIP2) were identified as prognosis related genes for the construction of prognostic model. Further analysis showed that the overall survival rate (OS) of patients in the high-risk subgroup was worse than that in the low-risk subgroup based on this model. The area under the characteristic curve of time-dependent receiver was 0.691 in TCGA and 0.624 in GEO, which confirmed the prognostic model to be a good one. We also established a nominal map based on the internal validation in 5 RBPs mRNAs and TCGA sequeues, showing a good ability to differentiate colon cancer.

Conclusions

We screened RBPs expression differences between colon cancer and adjacent non tumor colon tissues using the TCGA database to identify potential gene biomarkers. Besides, a very effective prediction model was constructed and tested based on the differential expression of RBPs using the TCGA and Gene Expression Omnibus (GEO) database. We also Validated of the relationship between the expression of five RBPs and prognosis

Background

Colon cancer is the third most common cancer and the second leading cause of cancer death worldwide. In 2018, there were over 1.8 million new cases and 881 thousand deaths globally. The global burden of colon cancer is expected to increase by 60% to more than 2.2 million new cases and 1.1 million deaths by 2030^[1]. In addition, although incidence rate and mortality rate of many high-income countries were stable or declining, but the age-specific incidence and mortality rates are still the highest in the world, especially the incidence among the young adults^[2, 3].

In this context, RNA binding proteins (RBPs) play an important role as regulatory factors of RNA life cycle, because they can bind to RNA sequences^[4]. In humans, more than 1,500 RBPs have been recently identified, accounting for 7.5% of the proteome. Each contemporary RBP binds to hundreds of RNAs, including both coding and non-coding RNAs, and precisely guides nuclear output, translation / degradation rates and intracellular localization of target transcripts. Thus, it plays an extensive regulatory role in almost all physiological processes. Accordingly, Alterations in RBP expression or function lead to abnormal RNA translation and may lead to the occurrence and development of various pathological conditions, including cancer^[5]. Recently, the next generation sequencing analysis of tumor samples showed that the expression of gene encoding RBP was significantly higher than that in the gene not encoding RBP^[6]. In addition, evidence showed that RBPs were closely related to the regulation of most cancer hallmarks, including cell proliferation, cell death resistance, stemness, cell dissemination and immune system evasion, which may act as a potential biomarker for tumor progression^[7]. In some cases, two RBPs can bind to the same RNA target to stabilize it, either enhancing or repressing translation^[8]. RBPs can also have dichotomous functions where they can both enhance^[9, 10] or repress^[11, 12] tumorigenesis depending upon the cellular context.

In recent years, advanced structural-analysis experiments provided evidence for complex protein-RNA interactions that do not require a typical RNA binding domain. RNA interactome capture (RIC) has identified "unconventional" RBPs in some organisms that have no recognizable RNA binding domains and have no known relationship with RNA biology^[13, 14]. Further studies have also shown that disordered protein regions can also promote the interactions between protein and RNA, which can be specific or non-specific. These unorthodox interactions regulate RNA metabolism and different RNA processes, including co- and post- transcriptionally^[15].

At least 16 families of RBPs are deregulated in cancer^[16]. Several specific RBPs have been published in the context of intestinal homeostasis and intestinal tumorigenesis including colon cancer including LIN-28, IGF2BPs/IMPs, musashi, HuR, Mex-3A, CELF1 and RBM3. .

Tumor is a kind of polygenic disease. Usually, each tumor is caused by the disorder of polygenic expression which form a complex network of signaling pathways, leading to the occurrence and development of tumor.

Numbers of projects began to explore methods to accurately predict the prognosis of cancer. Our study used mRNA- seq database from the Cancer Genome Atlas (TCGA). Univariate method was used to construct the gene marker of prognosis, significance analysis of RBPs expression and Cox regression survival analysis. The final regression model represents a potentially useful tool but needs to be verified in clinical practice. Nevertheless, our findings will help to elucidate the pathogenesis of colon cancer.

Methods

Data processing and screening for differentially expressed RBPS in colon cancer

the RNA-seq dataset of 39 normal colon tissue samples and 389 colon cancer

samples with matched clinical data were downloaded from The Cancer Genome Atlas database (TCGA, <https://portal.gdc.cancer.gov/>). The negative binomial distribution method was used to identify the differentially expressed RBPs between the two group. The Limma package (<http://www.bioconductor.org/packages/release/bioc/html/limma.html>) was based on the negative binomial distribution, applied to perform the analysis and preprocess the raw data, simultaneously. It fits a generalized linear model for each gene and uses empirical Bayes shrinkage for dispersion and fold-change estimation. All raw data were preprocessed by the Limma package. The Limma package was also used to identify the differently expressed RBPs in view of $|\log_2 \text{fold change}$

$(\text{FC}) \geq 1$ and false discovery rate (FDR) < 0.05 .

Visualization of Differently expressed RBPs with the method of KEGG and GO enrichment analysis

The biological functions of these differently expressed RBPs were comprehensively detected by GO (Gene Ontology) enrichment and Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis (Visualization and Integrated Discovery (DAVID) Bioinformatics Tool, version 6.8.). The GO analysis terms, including cellular component (CC), molecular function (MF), and biological process (BP). All enrichment analyses were carried out by utilizing online WEB-based Gene Set Analysis Toolkit (WebGestalt, <http://www.webgestalt.org/>)^[17]. Both P and FDR values were less than 0.05 as statistically significant.

Rbps Ppi Network Construction And Module Screening

The differently expressed RBPs were submitted to the STRING database (<http://www.stringdb.org/>) to identify protein-protein interaction (PPI) information^[18]. The Cytoscape 3.8.0 software was used to construct the PPI network further and visualized. The key modules and genes were selected in the PPI network by using Molecular Complex Detection (MCODE) plug-in with both MCODE score and node count number more than 5^[18]. All $P \leq 0.05$ was considered a significant difference.

Construction of a weighted OS predictive score model related with the RBPs expression

Univariate Cox regression analysis was performed on all differently expressed RBPs using the survival R package. A log-rank test was used to screen the significant candidate genes further. Subsequently, based on the above preliminary screened significant candidate genes, A multivariate Cox proportional hazards regression model was constructed and calculated a risk score to assess patient prognosis outcomes. The risk score formula for each sample was as follows:

$$\text{Risk score} = \beta_1 * \text{Exp1} + \beta_2 * \text{Exp2} + \beta_i * \text{Exp}_i$$

Where β is the coefficient value, and Exp was the gene expression level.

According to the median risk score survival analysis, colon cancer patients were divided into low- and high- risk groups. A log-rank test compared the difference of OS between the two subgroups. Additionally, a receiver operating characteristic (ROC) curve analysis was implemented by using the Survival ROC package to evaluate the prognostic capability of the above model^[19]. Besides, colon cancer patient samples with reliable prognostic information from the GSE75500 dataset (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE75500>) were used as a validation cohort to confirm the predictive capability of this prognostic model. Finally, the nomogram with calibration plots was conducted using the RMS R package to forecast the likelihood of OS. $P < 0.05$ was considered to be a significant difference.

Verification of express level and prognostic significance

The Human Protein Atlas (HPA) online database (<http://www.proteinatlas.org/>) was used to detect the expression of five RBPs at a translational level^[20]. The prognostic value of the five RBPs in colon cancer was verified by using the Kaplan Meier plotter (<https://kmplot.com/analysis/>) online tool^[21].

Results

Differently expressed RBPs between colon cancer and normal tissues

The databases of colon cancer were downloaded from TCGA contained 389 tumor and 39 normal colon tissue samples. Applying the R software packages3.64, the differently expressed RBPs were identified. A total of 1542 RBPs^[17] were included in the analysis, and 490 RBPs met the screening standard of this study ($P < 0.05$, $|\log_2 \text{FC}| > 0.5$), which consist of 323 upregulated and 167 downregulated RBPs. The expression distribution and relationship each other of these differently expressed RBPs was displayed in Figure 1.

GO and KEGG pathway enrichment analysis of the differently expressed RBPs and their downstream regulator

In order to study the function and mechanism of RBPs, we divided these differently expressed RBPs into two groups: up-regulated and down regulated. Then we use R software package to process them for functional enrichment analysis. The results showed that the upregulated RBPs were enriched in the biological processes(BP) related to RNA ribosome biogenesis, rRNA metabolic process, nucleic acid phosphodiester bond hydrolysis, RNA localization, RNA phosphodiester bond hydrolysis, tRNA metabolism, RNA 3' - terminal processing, RNA modification(Figure 1c). The down regulated RBPs were enriched in negative regulation of cellular amide metabolism, negative translation regulation, RNA splicing, mRNA catabolic process, tRNA metabolism, cytoplasmic ribonucleoprotein granule, ribonucleoprotein granule (Figure 1e). Through cellular component (CC) analysis, we found that the

upregulated RBPs with differential expression were enriched in preribosome, nuclear part, cytoplasmic ribonucleoprotein granule, ribonucleoprotein granule, spliceosomal complex, small-subunit processome, preribosome, large subunit precursor, 90s preribosome, exoribonuclease complex, cytoplasmic exosome(RNase complex) (Figure 1c). In addition, The cellular component (CC) analysis indicated that the upregulated differently expressed RBPs were enriched in cytoplasmic ribonucleoprotein granule, ribonucleoprotein granule, nuclear speck, P-body, spliceosomal complex, ribosome, cytoplasmic stress granule, ribosomal subunit, catalytic step 2 spliceosome, U2-type spliceosomal complex(Figure 1e). In terms of molecular function (MF), upregulated RBPs with different expressions are rich in RNA catalytic activity, nuclease activity, ribonuclease activity, tRNA catalytic activity, endonuclease activity, ribonucleoprotein complex binding, mRNA 3' - UTR binding, RNA helicase activity, tRNA binding, RNA methyltransferase activity(Figure 1c). In addition, the down regulated RBPs in MF have rich expression in RNA catalytic activity, mRNA 3' - UTR binding, translation regulator activity, double-stranded RNA binding, translation repressor activity, translation regulator activity, mRNA 3'-UTR AU-rich region binding, translation repressor activity, mRNA regulatory element binding(Figure 1e). With the method of the enrichment analysis of KEGG pathway, we found that the up regulated RBPs were enriched in ribosome biogenesis in eukaryotes, RNA transport, spliceosome, mRNA surveillance pathway, RNA degradation, ribosome, RNA polymerase, DNA replication, RNA degradation, ribosome, mRNA monitoring pathway, RNA transport and splicing (Figure 1d), while the downregulated RBPs were enriched in RNA transport, spliceosome, ribosome, TGF- β signaling pathway and RNA degradation(Figure 1f).

Protein-protein interaction (PPI) network construction and key modules selecting.

To further investigate the roles of differently expressed RNA binding proteins in colon cancer, we used the STRING database to create the PPI network (Figure 1a). Moreover, Cytoscape software was used to sort and analyze the data(Figure 2b). The expression network was processed using the MODE tool to identify possible key modules and the three key modules acquired(Figure 2c-f).

Prognosis-related RBPs selecting

With the method of univariate Cox regression analysis, the prognostic significance of these 5 differently expressed RBPs identified were investigated and 5 prognostic-associated candidate RBPs were validated (Figure 3a). This five-gene signature involved PNLDC1, NSUN6, NOL3, PPARGC1A and LRRFIP2, Of these, PPARGC1A and LRRFIP2, were protective factors while the rest were risk factors. This five-gene signature is of great importance for the prognostic evaluation of colon cancers. In addition, these 5 prognostic-associated candidate hub RBPs were analyzed by Multivariate Cox regression analysis and investigated their impact on patient survival time and clinical outcomes, and all the five RBPs were found to be independent predictors in the patients with colon cancer (Figure 3b).

Prognosis-related genetic risk score model construction and analysis

The five RBPs identified from the Multivariate Cox regression analysis were used to construct the predictive model. The predictive model was characterized by the linear combination of the expression

levels of the five genes weighted by their relative coefficient in the multivariate Cox regression as follows:

$$\text{Risk score} = (0.6708 \text{ExpPNLDC1}) + (1.3922 \text{ExpNSUN6}) + (0.7246 \text{ExpNOL3}) + (-0.6708 \text{ExpPPARGC1A}) + (-1.5486 \text{ExpLRRFIP2})$$

We then conducted a survival analysis to assess predictive ability of the 5 RBPs. A total of 337 patients with colon cancer were divided into low-risk and high-risk subgroups according to the risk score. The higher the risk score, the worse the clinical prognosis. The K-M OS curves of the two groups, based on the five genes, were significantly different ($p=8.432e-08$; Figure 3c). The area under the curve (AUC) of the time-dependent ROC curve was calculated to evaluate the prognostic capacity of the 5-gene features. The higher the AUC, the better the model performance. The five-gene-AUCs of biomarker prognostic model were 0.720 and 0.725 for the three and five-year survival time, respectively, indicating that the predict model had high sensitivity and specificity (Figure 3d;3e). In addition, The expression heat map, survival status of patients, and risk score of the signature consisting of five RBPs in the low- and high-risk subgroups are shown in Figure 3f~h.

Verifying the the prognostic model in the GEO dataset

The prognostic model was evaluated in other datasets to verify the reliability for the patients with colon cancer. The five-gene model was assessed in the GEO microarray data GSE75500 with same method in GSE75500 data. The OS of the colon cancer patients in the high-risk group was lower than that in the low-risk group ($p=3.17e-3$; Figure 3i). The time-dependent ROC analyses for the survival prediction of the prognostic model obtained AUCs of 0.691 at three years and 0.624 at five years. (Figure 3j;3k), and the risk score of the signature consisting of five RBPs in the low- and high-risk subgroups are displayed in Figure 3l~n. All these demonstrate that this five-RBP-gene prognostic model was capable of predicting OS in HCC patients.

A nomogram based on the five RBPs

In order to found a quantitative way for predicting the prognosis of the colon cancer patients, we integrated the five RBPs signature based on the multivariate Cox analysis to establish a nomogram that might help doctors to make the clinical decision for the patients (Figure 4a). We also assessed the prognostic significance of different clinical characteristics from TCGA by using COX regression analysis. However, The results showed that the tumor stage and risk score were independent prognostic factors correlated with OS through univariate and multivariate analysis (Figure 4b;4c).

Validation of the relationship between the expression of five RBPs and prognosis

To further explore the prognostic value of five RBPs in colon cancer, we drew Kaplan-Meier survival curves (PNLDC1, NSUN6, NOL3, PPARGC1A and LRRFIP2) to investigate the relationship between hub-RBPs and OS. The results of the log-rank test showed that the five RBPs were associated with OS in colon patients (Figure 4d~h). In order to determine the relationship between each gene expression and clinical characters, we analyzed the expression of proteins encoded by these five RBPs using clinical samples in

the human protein map database. NOL3 in colon tissue was significantly increased and LRRFIP2 was decreased compared with normal colon tissue (Figure 5). However, there was no significant difference in PNLDC1 and NSUN6 protein expression between tumor and normal colon tissues (Figure 5). However, PPARGC1A was not found on the website.

Discussion

We studied gene expression differences between colon cancer and adjacent non tumor colon tissues to identify potential gene biomarkers using the TCGA database. To establish a risk model and predict the prognosis of colon cancer, The differentially expressed genes were screened single variable, lasso and multivariate Cox analysis. Five RBPs: PNLDC1, NSUN6, NOL3, PPARGC1A and LRRFIP2 were identified to be related to the prognosis of colon cancer. Of these, PPARGC1A and LRRFIP2 were found to be protective factors. As an overall prognosis predictor, high levels of 5 RBPs were associated with poor prognosis of colon cancer. AUCs of ROC curve for the prognosis model for predicting the 3- and 5-year survival were 0.720 and 0.725, respectively, showed a good performance in survival prediction.

According to the RBPs- based risk scoring prognosis model, the patients with colon cancer were divided into high- and low- risk group. Clinicians can change patients' treatment plan to achieve personalized treatment. Strategies should be developed to prevent or detect metastasis of colon cancer in high-risk groups. Therefore, we should pay more attention to the high-risk groups,

Until now, the study of PNLDC1 in colon cancer and even all tumors has not been reported. Some results provided evidence that mammalian PNLDC1 is a regulator of piRNA biogenesis, transposon silencing and spermatogenesis, protecting the germline genome in mice^[22–24]. PIWI-interacting RNA (piRNA) is a small non-coding RNA (ncRNA) that interacts with PIWI proteins to form the piRNA silencing complex (piRISC). PIWI is a subfamily of Argonaute, and piRNA must bind to PIWI to exert its regulatory role in gene silencing and gene modification upstream or downstream of oncogenes in cancer cell lines or cancer tissues. Current studies indicated that piRNA and PIWI are significantly abnormally expressed in gastric, breast, kidney, colon, and lung cancers, and are involved in the initiation, progression, and metastasis of cancers, which may be the potential diagnostic tools, prognostic markers, and therapeutic targets for cancers. So PNLDC1 might affect the development and progress through regulating piRNA in colon cancer^[25].

Human NSUN6 is an RNA methyltransferase that catalyzes the transfer of the methyl group from S-adenosyl-L-methionine (SAM) to C72 of some types of tRNAs, it relies on a delicate network for RNA recognition, which involves both the primary sequence and tertiary structure of tRNA substrates. Until now, the biochemical properties and functions of NSUN6 homologs were not clear and no studies were reported in cancer until now^[26].

NOL3 gene encodes the apoptosis repressor with caspase recruitment domain (ARC) which is a potent and multifunctional death repressor that inhibits both death receptor and mitochondrial apoptotic

signaling. It has been previously reported to play a role in colon carcinogenesis^[27]. As a result of the increased ARC expression, NOL3 played a role in TRAIL-induced apoptosis reduced by hypoxia in colon cancer^[28].

Proteins of the peroxisome proliferator-activated receptor γ (PPAR γ) coactivator 1 (PGC-1) family of transcriptional coactivators coordinate physiological adaptations in many tissues, usually in response to demands for higher nutrient and energy supply. Of the founding members of the family, PGC-1 α (also known as PPARGC1A) is the most highly regulated gene, using multiple promoters and alternative splicing to produce a growing number of co-activator variants. PGC-1 α promoters are selectively active in distinct tissues in response to specific stimuli^[29]. Some results showed that PPARGC1A was significantly decreased in colon cancer. Taken together the data suggest that the transcriptional activity of PPAR γ may not only be decreased by mutation but also by down-regulation of the co-activator PGC-1 of PPAR γ in colon cancer^[30].

LRRFIP2 was identified as a modulator of Wnt signaling by a genome-wide functional screen. LRRFIP2 interacts with Dvl to increase the cellular abundance of β -catenin and activate TCF/LEF-dependent gene transcription^[31]. High Wnt signaling pathway activity promotes the proliferation, invasiveness, migration, progression chemoresistance, metastasis, stemness, ^[32–39] and also directs metabolic program of glycolysis and angiogenesis^[40].

The preoperative pathology of colon cancer is often only used to determine its pathological nature. In combination with the imaging data of patients such as MRI scan, clinicians determine the treatment plans for patients with colon cancer. Although the preoperative clinical stage of some colon cancer is not late, the survival time of some patients with colon cancer is greatly shortened because of peritoneal implant metastasis or distant metastasis post-operation and even though receiving postoperative chemotherapy without considering the biological characteristics of the tumor. A model to predict the prognosis of patients could help clinicians to develop individualized treatment plan for the patients. For example, the method of neoadjuvant chemotherapy, intraoperative hyperthermia and shorten the interval between postoperative reexamination can be used to improve the overall survival time of the high-risk patients.

However, several limitations of the current study should be considered. First, the population ethnicities in the TCGA database are mainly confined to people from western countries, and extrapolating the findings to other ethnic groups needs to be substantiated. Second, our data did not perform PPARGC1A expression because of lacking the experimental results.

A robust nomogram should be validated externally in different queues; therefore, it needs to be further validated in multicenter clinical trials and prospective studies. In the future, we will also explore the possibility of including more predictive variables to further improve performance. Other regression modeling methods will be used to determine whether the prediction accuracy can be further improved. To sum up, our results show that the 5-RBP-gene prognostic model is a reliable tool for predicting OS of

colon cancer, and the nomogram including the prognosis model can help clinicians choose personalized treatment for colon cancer patients.

Conclusions

We screened RBPs expression differences between colon cancer and adjacent non tumor colon tissues using the TCGA database to identify potential gene biomarkers. Besides, a very effective prediction model was constructed and tested based on the differential expression of RBPs using the TCGA and GEO database.

We also Validated of the relationship between the expression of five RBPs and prognosis.

Abbreviations

RBP:RNA binding protein; TCGA:The Cancer Genome Atlas; OS:survival rate; GEO:Gene Expression Omnibu; RIC:RNA interactome capture; GO:Gene Ontology; KEGG:Kyoto encyclopedia of genes and genomes; DAVID:Visualization and Integrated Discovery; CC:cellular component; MF:molecular function; BP:and biological process; PPI:protein-protein interaction; MCODE:Molecular Complex Detection; ROC:receiver operating characteristic; HPA:Human Protein Atlas; piRNA:PIWI-interacting RNA; ncRNA:non-coding RNA; piRISC:piRNA silencing complex; SAM:S-adenosyl-L-methionine; ARC:apoptosis repressor with caspase recruitment domain; PPAR γ :peroxisome proliferator-activated receptor γ ; PGC-1:peroxisome proliferator-activated receptor γ coactivator 1.

Declarations

Acknowledgments

We thank platforms of TCGA, and GEO datasets for data download.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Availability of data and materials

Publicly available datasets were analyzed in this study. This data can be found here: GSE75500, TCGA:The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conception and design: YHY and BQQ, Development of methodology: YHY, LHL and BQQ. Acquisition of data: MFZ and GHR. Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): LZY, SX, JY and LY. Writing, review, and/or revision of the manuscript: PTM, LSH, SRP. Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): LYJ and WJB. Study supervision: YDL, QZB, GHX and JHC

Ethics approval and consent to participate

All the analysis are under the approval by the ethics committee of the First Affiliated Hospital of Harbin Medical University Ethics Review Board.

Consent for publication

Not applicable.

Funding

No funding.

References

1. 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: 394-424.
2. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut.* 2017; 66: 683-691.
3. Siegel RL, Torre LA, Soerjomataram I, et al. Global patterns and trends in colorectal cancer incidence in young adults. *Gut.* 2019; 68: 2179-2185.
4. Mancarella C, Scotlandi K. IGF2BP3 From Physiology to Cancer: Novel Discoveries, Unsolved Issues, and Future Perspectives. *Front Cell Dev Biol.* 2020; 15(7):363-410. doi: 10.3389/fcell.
5. Coppin L, Leclerc J, Vincent A, et al. Messenger RNA life-cycle in cancer cells: emerging role of conventional and non-conventional RNA-Binding proteins? *Int. J. Mol. Sci.* 2018; 19:E650
6. Neelamraju Y, Gonzalez-Perez A, Bhat-Nakshatri P, Nakshatri H, and Janga, SC. Mutational landscape of RNA-binding proteins in human cancers. *RNA Biol.* 2018;15(1):115-129. doi: 10.1080/15476286
7. Pereira B, Billaud M, Almeida R. RNA-binding proteins in cancer: old players and new actors. *Trends Cancer.* 2017;3(7):506-528. doi: 10.1016
8. Sureban SM, Murmu N, Rodriguez P, May R, Maheshwari R, Dieckgraefe BK, Houchen CW, Anant S. Functional antagonism between RNA binding proteins HuR and CUGBP2 determines the fate of COX-2 mRNA translation. *Gastroenterology.* 2007;132(3):1055-65.
9. Hamilton KE, Noubissi FK, Katti PS, Hahn CM, Davey SR, Lundsmith ET, Klein-Szanto AJ, Rhim AD, Spiegelman VS, Rustgi AK. IMP1 promotes tumor growth, dissemination and a tumor-initiating

- cell phenotype in colorectal cancer cell xenografts. *Carcinogenesis*. 2013;34(11):2647-2654.
10. Gutschner T, et al. Insulin-like growth factor 2 mRNA-binding protein 1 (IGF2BP1) is an important protumorigenic factor in hepatocellular carcinoma. *Hepatology*. 2014;59(5):1900-11.
 11. Hamilton KE, Chatterji P, [Lundsmith ET](#), Andres SF, Giroux V, Hicks PD, Noubissi FK, [Spiegelman VS](#), [Rustgi AK](#). Loss of Stromal IMP1 Promotes a Tumorigenic Microenvironment in the Colon. *Mol Cancer Res*. 2015;13(11):1478-86.
 12. Wang G, Huang Z, Liu X, Huang W, Chen S, Zhou Y, Li D, Singer RH, Gu W. IMP1 suppresses breast tumor growth and metastasis through the regulation of its target mRNAs. *Oncotarget*. 2016;7(13):15690-702.
 13. Hentze MW, Castello A, Schwarzl T, Preiss T. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* 2018
 14. Castello A, Horos R, Strein C, Fischer B, Eichelbaum K, Steinmetz LM, Krijgsveld J, Hentze MW. System-wide identification of RNA-binding proteins by interactome capture. *Nat Protoc*. 2013; 8(3): 491-500.
 15. Järvelin AI, Noerenberg M, Davis I, Castello A. The new (dis)order in RNA regulation. *Cell Commun Signal*. 2016; 14:19.
 16. Pereira, B, Billaud, M, and Almeida, R. RNA-binding proteins in cancer: old players and new actors. *Trends Cancer*. 2017; 3: 506-528
 17. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein protein association networks with increased coverage, supporting functional discovery in genome wide experimental datasets. *Nucleic Acids Res*. 2019; 47:D607–13.
 18. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*. 2003; 4:2.
 19. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*. 2000; 56:337–44.
 20. Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, Alm T, Asplund A, Björk L, Breckels LM, Bäckström A, Danielsson F, Fagerberg L, et al. A subcellular map of the human proteome. *Science*. 2017; 356:356.
 21. Györfy B, Surowiak P, Budczies J, Lánczky A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in nonsmall-cell lung cancer. *PLoS One*. 2013; 8:e82241.
 22. Zhang Y, Guo R, Cui Y, Zhu Z, Zhang Y, Wu H, Zheng B, Yue Q, Bai S, Zeng W et al. An essential role for PNLDC1 in piRNA 3' end trimming and male fertility in mice. *Cell Res*. 2017; 27: 1392–1396
 23. Ding D, Liu J, Dong K, Midic U, Hess RA, Xie H, Demireva EY, Chen C PNLDC1 is essential for piRNA 3' end trimming and transposon silencing during spermatogenesis in mice. *Nat Commun*. 2017; 8: 819–828

24. Toru Nishimura, Ippei Nagamori, Tsunetoshi Nakatani, Natsuko Izumi, Yukihide Tomari, Satomi Kuramochi-Miyagawa, and Toru Nakano. [PNLDC1, mouse pre-piRNA Trimmer, is required for meiotic and post-meiotic male germ cell development.](#) *EMBO Rep.* 2018; 19(3): e44957
25. Guo B, Li D, Du L, Zhu X. [piRNAs: biogenesis and their potential roles in cancer.](#) *Cancer Metastasis Rev.* 2020 Jan 20. doi: 10.1007/s10555-020-09863-0.
26. Li J, Li H, Long T, Dong H, Wang ED, Liu RJ. [Archaeal NSUN6 catalyzes m5C72 modification on a wide-range of specific tRNAs.](#) *Nucleic Acids Res.* 2019 Feb 28;47(4):2041-2055.
27. Watanabe T, Kobunai T, Yamamoto Y, Ikeuchi H, Matsuda K, Ishihara S, Nozawa K, Iinuma H, Kanazawa T, Tanaka T, Yokoyama T, Konishi T, Eshima K, Ajioka Y, Hibi T, Watanabe M, Muto T, Nagawa H. Predicting ulcerative colitis-associated colorectal cancer using reverse-transcription polymerase chain reaction analysis. *Clin Colorectal Cancer.* 2011 Jun;10(2):134-141.
28. Ao JE, Kuang LH, Zhou Y, Zhao R, Yang CM. Hypoxia-inducible factor 1 regulated ARC expression mediated hypoxia induced inactivation of the intrinsic death pathway in p53 deficient human colon cancer cells. *Biochem Biophys Res Commun.* 2012 Apr 20;420(4):913-917
29. Martínez-Redondo V, Pettersson AT, Ruas JL. [The hitchhiker's guide to PGC-1 \$\alpha\$ isoform structure and biological functions.](#) 2015; 58(9):1969-1977.
30. Jonas Feilchenfeldt¹, Marie Anne Bründler, Claudio Soravia, Martin Tötsch, Christoph A Meier. Peroxisome Proliferator-Activated Receptors (PPARs) and Associated Transcription Factors in Colon Cancer: Reduced Expression of PPAR γ -coactivator 1 (PGC-1). *Cancer Lett.* 2004; 203(1):25-33.
31. Jun Liu, Anne G Bang, Chris Kintner, Anthony P Orth, Sumit K Chanda, Sheng Ding, Peter G Schultz. Identification of the Wnt Signaling Activator Leucine-Rich Repeat in Flightless Interaction Protein 2 by a Genome-Wide Functional Analysis. *PNAS.* 2005; 102(6): 1927-1932.
32. Yu J, Liu D, Sun X, Yang K, Yao J, Cheng C, Wang C, Zheng J. [CDX2 inhibits the proliferation and tumor formation of colon cancer cells by suppressing Wnt/ \$\beta\$ -catenin signaling via transactivation of GSK-3 \$\beta\$ and Axin2 expression.](#) *Cell Death Dis.* 2019;10;10(1):26
33. Vermeulen L, De Sousa E Melo F, van der Heijden M, Cameron K, de Jong JH, Borovski T, Tuynman JB, Todaro M, Merz C, Rodermond H, Sprick MR, Kemper K, Richel DJ, Stassi G, [Wnt activity defines colon cancer stem cells and is regulated by the microenvironment.](#) *Medema JP.Nat Cell Biol.* 2010;12(5): 468-476.
34. Piao SS, Shang B. [Pizotifen Inhibits the Proliferation and Migration of Colon Cancer HCT116 Cells by Down-regulating Wnt Signaling Pathway.](#) *Ann Clin Lab Sci.* 2019; 49(2):183-188.
35. Zhang Y, Tian S, Xiong J, Zhou Y, Song H, Liu C. [JQ-1 Inhibits Colon Cancer Proliferation via Suppressing Wnt/ \$\beta\$ -Catenin Signaling and miR-21.](#) *Chem Res Toxicol.* 2018; 31(5): 302-307.
36. Regan JL, Schumacher D, Staudte S, Steffen A, Haybaeck J, Keilholz U, Schweiger C, Golob-Schwarzl N, Mumberg D, Henderson D, Lehrach H, Regenbrecht CRA, Schäfer R, Lange M. [Non-Canonical Hedgehog Signaling Is a Positive Regulator of the WNT Pathway and Is Required for the Survival of Colon Cancer Stem Cells.](#) *Cell Rep.* 2017; 21(10):2813-2828.

37. Ba S, Xuan Y, Long ZW, Chen HY, Zheng SS. [MicroRNA-27a Promotes the Proliferation and Invasiveness of Colon Cancer Cells by Targeting SFRP1 through the Wnt/ \$\beta\$ -Catenin Signaling Pathway](#). *Cell Physiol Biochem*. 2017; 42(5): 1920-1933.
38. Yue B, Liu C, Sun H, Liu M, Song C, Cui R, Qiu S, Zhong M. [A Positive Feed-Forward Loop between LncRNA-CYTOR and Wnt/ \$\beta\$ -Catenin Signaling Promotes Metastasis of Colon Cancer](#). *Mol Ther*. 2018; 26(5): 1287-1298.
39. Uppada SB, Gowrikumar S, Ahmad R, Kumar B, Szeglin B, Chen X, Smith JJ, Batra SK, Singh AB, Dhawan P. [MASTL induces Colon Cancer progression and Chemoresistance by promoting Wnt/ \$\beta\$ -catenin signaling](#). *Mol Cancer*. 2018; 17(1): 111.
40. Pate KT, Stringari C, Sprowl-Tanio S, Wang K, TeSlaa T, Hoverter NP, McQuade MM, Garner C, Digman MA, Teitell MA, Edwards RA, Gratton E, Waterman ML. [Wnt signaling directs a metabolic program of glycolysis and angiogenesis in colon cancer](#). *EMBO J*. 2014; 33(13):1454-1473.

Figures

Figure 1

The differentially expressed RBPs and the relationship each other in HCC. (a) Heat map; In the transverse type, blue represents 50 normal tissues, red represents 100 colon cancer tissues, in the longitudinal thermogram, each point represents the expression dysregulation of RBP in this specimen, the color strip in the upper right corner represents the degree of RBP dysregulation, red represents the highest degree of up regulation, and green represents the highest degree of down regulation. (b) Volcano plot. The abscissa

$\log_2 FC$ represents the pair value of times for the difference of expression level between cancer and normal tissues, and the ordinate $-\log_{10} (FDR)$ represents the negative pair value of FDR. The larger the ordinate value is, the more significant the difference between the expression of RBP gene in the two samples is, and the more reliable the differential expression gene is, Red represents up-regulated expression of RBP in cancer, green represents down regulated expression, and black represents no difference or exclusion. KEGG pathway and GO enrichment analysis of aberrantly expressed RBPs, Abscissa represents the proportion of genes, The area of the black round in the upper right corner represents the number of genes (c) Upregulated differently expressed RBPs GO enrichment analysis;(d) Upregulated differently expressed RBPs KEGG pathway enrichment analysis;(e) Downregulated differently expressed RBPs GO enrichment analysis;(f) Downregulated differently expressed RBPs KEGG pathway enrichment analysis.

Figure 2

PPI network and modules analysis. (a) PPI network of differentially expressed RBPs; (b) PPI network of differentially expressed RBPs sort and analysis by Cytoscape; (c) PPI network of three key modules; (d) PPI network of the first key module; (e) PPI network of the second key module; (f) PPI network of the third key module; Green spheres: down-regulation; red spheres: upregulation(b-f).

Figure 3

(a) Univariate Cox regression analysis for identification of RBPs in the training dataset.; (b) Multivariate Cox regression analysis to identify prognosis related RBPs; (c) Survival curve for low- and high-risk subgroups in the TCGA cohort;(d) Time-dependent ROC curve analysis of 3 years survival prediction by the prognostic model in the TCGA cohort; (e) Time-dependent ROC curve analysis of 5 years survival prediction by the prognostic model in the TCGA cohort; (f) Expression heat map in the TCGA cohort; (g)

Risk score distribution, Taking the risk score equal to 1 as the dividing point, we divided the patients into two subgroups: high risk group and low risk group in the TCGA cohort; (h) Survival status distribution in the TCGA cohort; (i) Survival curve for low- and high-risk subgroups in the GSE75500 cohort; (j) Time-dependent ROC curve analysis of 3 years survival prediction by the prognostic model in the GSE75500 cohort; (k) Time-dependent ROC curve analysis of 5 years survival prediction by the prognostic model in the GSE75500 cohort; (l) Expression heat map in the GSE75500 cohort; (m) Risk score distribution in the GSE75500 cohort; (n) Survival status distribution in the GSE75500 cohort.

Figure 4

(a) Nomogram for predicting 1-, 3-, and 5-year OS of colon cancer patients in the TCGA cohort. In the nomogram, the parameters that affect the prognosis are not the same distribution, some are continuous data, some are binary or multivariate data, so after calculating the regression coefficient, the incomparable variables are uniformly calculated as counting variables, and finally the total score is calculated, and the prognosis is quantitatively determined by the total score comparison table; (b) Univariate analysis of the prognostic value of different clinical parameters; (c) Multivariate analysis of the prognostic value of different clinical parameters; (d) Prognostic value of PNLDC1 by Kaplan- Meier-plotter; (e) Prognostic value of NSUN6 by Kaplan- Meier-plotter; (f) Prognostic value of NOL3 by Kaplan- Meier-plotter; (g) Prognostic value of PPARGC1A by Kaplan- Meier-plotter; (h) Prognostic value of LRRFIP2 by Kaplan- Meier-plotter.

Figure 5

Verification of RBPs expression in HCC and normal liver tissue using the HPA database. (A) PNLDC1 in colon cancer; (B) PNLDC1 in normal colon tissue; (C) NSUN6 in colon cancer; (D) NSUN6 in normal colon tissue; (E) NOL3 in colon cancer; (F) NOL3 in normal colon tissue; (G) LRRFIP2 in colon cancer; (H) LRRFIP2 in normal colon tissue.