A novel seven-microRNA signature predicts prognosis in colorectal adenocarcinoma

Shengying Jiang

Second Hospital of Hebei Medical University

Xiaoli Xie

Second Hospital of Hebei Medical University

Hui-Qing Jiang (✉ jianghuiqing1959@sina.com)

Second Hospital of Hebei Medical University <https://orcid.org/0000-0002-5235-1457>

Research

Keywords: TCGA, Colorectal adenocarcinoma, MicroRNA, Prognosis

Posted Date: August 4th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-46147/v1>

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

[Read Full License](#)

Abstract

Background Colorectal carcinoma, primarily colorectal adenocarcinoma (CRA), is one of the most common malignancies, ranking third, while contributing the second cause of cancer death in the world. MicroRNAs, a type of non-coding RNA, play an important role in regulating cancer-related cell biology. To simply and accurately predict survival risk for CRA patients, we identified a novel seven-miRNA signature. The microRNA (miRNA) expression profiles along with clinical data of 512 CRA patients were downloaded from The Cancer Genome Atlas (TCGA) database, and 415 patients with complete clinical information were further divided into training set and test set randomly. To construct the prognostic signature in the training set, a series of Cox regression analyses were performed, including univariate regression, least absolute shrinkage and selectionator operator (LASSO) - Cox regression, and multivariate regression.

Results Seven predictive miRNAs (miR-153-2, miR-3199-2, miR-144, miR-887, miR-561, miR-3684, and miR-505) were ultimately screened. The ROC curves for 5-year survival in the training set, test set and entire set were 0.889, 0.742, and 0.816, respectively. Kaplan-Meier analysis results of the three sets all showed $p < 0.05$. Further analyses demonstrated that the signature was an independent prognostic risk factor for CRA patients, and its predictive ability was superior to age and TNM stage. Functional enrichment analysis revealed that p53 and ErbB pathways were related to the prognostic regulation of miRNAs in the signature in CRA patients.

Conclusion Our study demonstrated the potential of this novel seven-miRNA signature to independently predict overall survival in patients with CRA. Functional enrichment analysis further revealed the possible regulatory role of miRNAs in the signature in CRA-related cell biological processes and signaling pathways.

Background

Colorectal adenocarcinoma (CRA) is the main form of colorectal cancer (CRC), accounting for more than 95% of CRCs, and has high morbidity and mortality.¹ Due to its high heterogeneity, traditional predictors such as age and TNM stage are not sufficient to accurately predict the survival risk of CRA patients. Exploring novel biomarkers is necessary to provide effective and personalized predictive tools. For the past few years, investigators have carried out a series of explorations in this field, and several prognostic gene signatures,^{2,3} transcriptional signatures and noncoding RNA signatures have been published.⁴⁻⁶ However, there is still no recognized prognostic prediction model, and further research is required.

MicroRNAs (miRNAs) are a class of endogenous non-coding single-stranded RNA molecules about 20~25 nucleotides with regulatory functions, and participates in a series of important processes in life, including early development, cell proliferation, apoptosis, cell death, fat metabolism and cell differentiation. Many miRNA expression profiles related to particular malignancies have been found to have tumor-suppressive or oncogenic roles in different cancer types and further affect the prognosis of

patients. Moreover, the functions of miRNAs are involved in the occurrence, development and metastasis of tumors.⁷ For instance, Mirzaei et al reported that miR-29b has significant tumor-suppressive effects on chronic lymphocytic leukemia (CLL).⁸ In addition, Zhou Y et al discovered that miR-130a acts as an oncogenic miRNA in gastric cancer.⁹

In our study, a seven-miRNA prognostic signature was developed through data mining of The Cancer Genome Atlas (TCGA) database. The prognostic model can identify high-risk patients with lower survival rates to allow more effective targeted therapy to be offered. Functional enrichment analysis clarified that the screened miRNAs participated in many CRA-related cell biological processes and signaling pathways. Our findings not only demonstrated the potential power of the seven-miRNA prognostic model for predicting the survival of CRA patients, but also provide new ideas for targeted therapy in patients with CRA..

Methods

Acquisition and processing of the data sets

MiRNA expression data along with clinical information for 521 CRA patients were downloaded from the TCGA database at <http://cancergenome.nih.gov/> (up to March 2020). From the miRNA expression data, the miRNAs with minimal expression levels (<1) were first filtered out, and the data were then processed via a standard procedure using the “edgeR” package (version 3.6.2). Finally, 385 differentially expressed miRNAs were identified between cancer tissues and normal tissues in CRA patients; a $|\log_2FC|$ of >1 and an adjusted P -value of <0.05 were used as cutoff criteria. The heatmap of the top significantly differentially expressed miRNAs was generated with the “ggplot2” package (version 3.6.3). Patients with incomplete clinical information (including sex, age, TNM stage, survival status and survival time) were excluded from further analysis. After the above exclusion steps, 415 CRA patients were enrolled in and divided into training set ($n = 208$) and test set ($n = 207$) randomly.

Screening the prognosis-related miRNAs and modeling in the training set

First, the regression analysis of univariate Cox proportional hazards was performed in the training set to analyze the correlation between the expression level of each miRNA and the patient’s survival time and status via the R package “survival”. Second, least absolute shrinkage and selection operator (LASSO) - Cox regression was conducted on the miRNAs with a P -value of <0.05 in the first analysis. LASSO is a compression estimate method that can produce a more refined model by constructing a penalty function. In this study, we used the R package “glmnet”. To determine the value of penalty parameter λ , cross-validation was carried out first, and the value of λ resulting in the minimum mean cross-validated error was selected.¹⁰ Third, the selected miRNAs were included in the multivariate Cox regression analysis to calculate their respective coefficients. Last, the following seven-miRNA prognostic signature was constructed: seven-miRNA signature = $\sum(\beta_i * \text{Expi})$, where i is the name of the prognostic miRNA and β is

the coefficient of that miRNA in the third analysis. MiRNAs with a positive coefficient were considered as protective miRNA, while those with a negative coefficient were considered as oncogenic miRNA.

Efficacy verification of the seven-miRNA signature in the three data sets

To further assess the efficacy of the seven-miRNA signature, we calculated the score of each patient in the three data sets (the training set, the test set and the entire set) based on this seven-miRNA signature. According to the median, each set was divided into a high-risk group and a low-risk group, and then the Kaplan-Meier survival analysis was applied to compare the overall survival difference between the high-risk group and the low-risk group using the R package “survival”. To verify the efficacy of the seven-miRNA signature, time-dependent receiver operating characteristic (ROC) curves and the areas under the ROC curves (AUCs) were generated via the R package “timeROC”.

Combination of clinical factors to evaluate the efficacy of the seven-miRNA signature

In the entire data set, the seven-miRNA signature combining clinical factors (including age, gender, race, tumor site and TNM stage) was analyzed by univariate Cox regression to identify associations between these miRNAs and overall patient survival. The variables with a *P*-value of <0.05 were included in further horizontal and vertical comparisons. ROC curve analysis was performed with the R packages “plotROC” and “ggplot2” to horizontally compare the seven-miRNA signature with clinical factors related to the prognosis of CRA. Kaplan-Meier survival curves were used for stratified longitudinal analysis.

MiRNA target gene prediction and functional enrichment analysis

Target genes of the seven selected miRNAs were predicted via the following three miRNA databases: miRTarBase (<http://mirtarbase.mbc.nctu.edu.tw/>, version: 7.0), TargetScan (<http://www.targetscan.org/>, version: Human 7.2) and miRDB (<http://mirdb.org/>). The intersection of the results obtained from the three databases was considered the set of miRNA-related genes. Functional enrichment analyses of these miRNA-related genes were performed on the Database for Annotation, Visualization, and Integrated Discovery (DAVID, version 6.8) at <https://david.ncifcrf.gov/>, which could conduct Gene Ontology (GO) biological process (BP) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. The results with a *P*-value of < 0.05 were visualized using the R packages “Cairo” (version 3.6.3) and “ggplot2” (version 3.6.3).

Results

Patient data

MiRNA expression files and clinical information for 521 CRA patients (comprising 529 tumor samples and 11 normal tissues) were downloaded from the TCGA database. A total of 415 CRA patients with complete clinical data were enrolled in further analysis. All enrolled patients had primary adenocarcinoma, did not have a previous or concurrent malignancy, and received no chemotherapy or

radiotherapy before surgery. After differential miRNA expression analysis, the 415 CRA patients were randomly divided into two sets (Table 1).

Differentially expressed miRNAs between cancer tissues and normal tissues in CRA patients

Before analysis, data normalization and log2 transformation were conducted in the miRNA expression files firstly using R software (version 3.6.1). We finally identified 385 differentially expressed (231 upregulated and 154 downregulated) miRNAs ($|\log_2\text{FC}| > 1$ and the adjusted P -value < 0.05) using the “DESeq” package. The heatmap of the expression of differentially expressed miRNAs in Figure 1 could clearly distinguish the CRA samples from the controlling normal samples.

Signature identification of the prognosis-related miRNA in the training set

Univariate Cox regression analysis with the LASSO method was used in the training set for selection of CRA prognosis-related miRNAs among the 385 differentially expressed miRNAs (Figure 2). Then, through a stepwise multivariate Cox regression analysis to maximize accuracy and validity, seven miRNAs were finally screened, and the prognosis-related miRNA signature was constructed based on the expression levels of these seven miRNAs. The signature formula was as follows: seven-miRNA signature = $\text{hsa-miR-144} \times -0.255 + \text{hsa-miR-153-2} \times 0.472 + \text{hsa-miR-505} \times 0.475 + \text{hsa-miR-887} \times 0.565 + \text{hsa-miR-3199-2} \times 0.509 + \text{hsa-miR-561} \times 0.452 + \text{hsa-miR-3684} \times 0.516$. The higher concordance index (C index), the better model performance, and the C index of the seven-miRNA signature was 0.772.

Among these miRNAs, hsa-miR-144 had a negative coefficient in the risk formula, indicating that it might be considered a protective miRNA; that is, the higher the expression level of this miRNA, the longer was the survival time of the patient. The remaining six miRNAs in this risk formula had positive coefficients, pointing that higher expression levels of these miRNAs were associated with shorter patient survival times.

Evaluation of the survival prediction capability of the seven-miRNA signature in the training set

The 208 samples in the training set were scored according to the risk scoring formula and divided into the high-risk group ($n=104$) and the low-risk group ($n=104$), with a median value of 1.009 as the cutoff point. The Kaplan-Meier overall survival curves of the two groups showed statistically significant differences between patient survival (log-rank test $P=5e-08$) (Figure 3A). ROC curves along with AUC values were further generated to evaluate the prognostic ability of the seven-miRNA signature (Figure 3B). The AUCs for 3- and 5-year survival in the training set were 0.769 and 0.889, respectively, indicating that the signature had good performance in predicting CRA patients' survival risk. In addition, the distributions of the risk score, survival status and expression profile of the seven miRNAs in the training set were illustrated to visualize the analysis results (Figure 4). Patients with higher risk scores had poorer overall survival than patients with lower risk scores, as shown in the figure. The expression levels of the protective miRNA hsa-miR-144 were significantly decreased in patients with high risk scores, whereas

those of the remaining six oncogenic miRNAs (miR-153-2, miR-505, miR-887, miR-3199-2, miR-561, and miR-3684) were increased.

Performance verification of the seven-miRNA signature in the test set and entire set

To evaluate the performance of the seven-miRNA signature for predicting the survival of CRA patients, Kaplan-Meier survival analysis was conducted and ROC curves with AUC values were generated in the test set (n= 207) and the entire set (n= 415). Patients in the two validation sets were classified into a high-risk group and a low-risk group according to the constructed seven-miRNA signature. As shown in the Kaplan-Meier survival curve for the test set, patients in the high-risk group (n= 103) had significantly shorter overall survival times than those in the low-risk group (n= 104) ($P = 6 \times 10^{-3}$, Figure 5A). In the test set, the AUC for 3-year survival was 0.666, and that for 5-year survival was 0.742 (Figure 5B). Similar results were obtained in Kaplan-Meier survival analysis of the entire TCGA set, in which the patients in the high-risk group (n= 207) had significantly shorter overall survival times than those in the low-risk group (n= 208) ($P = 0$, Figure 6A). In the entire TCGA set, the AUC for 3-year survival was 0.722, while that for 5-year survival was 0.816 (Figure 6B). These results confirmed the power of the seven-miRNA signature for predicting the prognosis of CRA patients.

To better demonstrate differences in the survival outcomes and miRNA expression levels between CRA patients in the high-risk and low-risk groups, we visualized the distributions of the risk scores, survival statuses and miRNA expression levels of patients in the test set and the entire TCGA set. The overall survival times of patients in the high-risk group was significantly shorter than those of patients in the low-risk group; moreover, one protective miRNA, hsa-mi-144, was highly expressed in the low-risk group, while the remaining 6 oncogenic miRNAs were highly expressed in the high-risk group, reconfirming the results in the training set. (Figure 5C, Figure 6C).

Combination of routine clinical factors for efficacy evaluation of the seven-miRNA signature

Clinical factors play an important role in tumor prognosis. To construct a more stable and effective prognostic model, the seven-miRNA signature combined with conventional clinical parameters (including age, race, sex, tumor site and TNM stage) was analyzed by univariate and multivariate Cox regression along with stratification analysis to evaluate the associations of these parameters with the overall survival in the entire data set. Variables with a P -value of <0.05 —i.e., the seven-miRNA signature, age and TNM stage—were considered to be associated with prognosis, as shown in Table 2. In addition, the conclusion can be made that the seven-miRNA signature is an independent risk factor for the prognosis of CRA patients.

Then, ROC curves and AUC values were used to horizontally compare the efficacy of the seven-miRNA signature with those of patient age and TNM stage in predicting the prognosis of patients with CRA. First, the seven-miRNA signature, age, and a new variable combining both of these variables were included in the analysis. As shown in Figure 7A, the seven-miRNA signature showed a higher predictive power than the age. Then, the seven-miRNA signature, TNM stage, and a new variable combining both of these

variables were included; similarly, the predictive power of the seven-miRNA signature was superior to that of the TNM stage (Figure 7B).

Finally, analysis was performed on the 415 patients stratified by age and TNM stage to compare the risk score of the seven-miRNA signature. The patients were divided by the average age in years into the younger group (<65, n=189) and the aged group (≥ 65 , n=226). In the younger group, the seven-miRNA signature divided the patients into a high-risk group (n=95) and a low-risk group (n=94) according to the median risk score. Kaplan-Meier survival curves showed that the difference between the two risk groups was statistically significant ($P=5.891\text{e-}05$, Figure 8A&B). The corresponding values were also statistically significant ($P=1\text{e-}07$) in the aged group, which was also divided into the high- and low-risk groups (n=112 and 114, respectively). Then, the patients were classified by TNM stage into the early-stage group (stages I+II, n=229) and the advanced-stage group (stages III+IV, n=186). In both stage groups, the seven-miRNA signature separated the patients into a high-risk group and a low-risk group independently. Similarly, the Kaplan-Meier survival curves for both stage groups indicated that the overall survival times of patients in the high-risk groups were significantly shorter than those of patients in the low-risk groups ($P=5.891\text{e-}05$, 0.00011855, Figure 8C&D).

Target gene prediction and functional enrichment analysis of the prognostic miRNAs

To investigate the cellular processes and biological pathways associated with the seven prognostic miRNAs, miRNA target prediction was separately performed with miRDB, miRTarBase7.0 and TargetScan. MiRDB identified 9297 miRNA-related genes,¹¹ while TargetScan,¹² the most comprehensive miRNA target gene database, identified 38678 mRNAs. MiRTarBase,¹³ the experimentally validated miRNA-target interaction database, identified 152 mRNAs associated with the seven prognostic miRNAs. The intersection of the results from these three databases was identified as the set of miRNA-related genes (n=67, Figure 9B) for further functional enrichment analyses. First, GO BP and KEGG pathway enrichment analyses were performed with these miRNA-related genes via DAVID.^{14,15} Then, the genes with a P -value of < 0.05 (Table 3) were visualized using the R packages “Cairo” and “ggplot2” (Figure 9A&C). The GO and KEGG analysis results showed that these genes are involved in cellular processes associated with cancer, such as regulation and negative regulation of apoptosis, and in biological pathways such as the p53 signaling pathway. The functional enrichment analysis results indicated the potential roles of the seven prognostic miRNAs in the regulation of CRA tumorigenesis and progression via targeting of their associated mRNAs.

Discussion

CRC is the world's third most common cancer,¹⁶ accounting for approximately 10% of all new cancer cases annually worldwide.¹⁷ CRA is the most common histological type of CRC. Although CRC screening for earlier detection has somewhat reduced morbidity and mortality in recent years, the five-year survival rate of CRA patients remains unsatisfactory,¹⁸ and the molecular mechanisms underlying CRA progression are still elusive. CRC is a highly heterogeneous tumor, and few studies have focused

specifically on the prognosis of CRA. Developing novel predictive biomarkers is important and needed for obtaining a better understanding of CRA and providing diagnostic and therapeutic strategies. In this study, we conducted a series of analyses on miRNA expression and corresponding clinical data for training and validation set generated from 415 CRA patients in the TCGA database, with an aim to develop a meaningful predictive model related to CRA prognosis.

MiRNAs, an important subset of noncoding RNAs, have been recognized to play multifaceted roles in controlling cellular functions by repressing their target genes.¹⁹ For example, let-7 family members can regulate cancer stem cells by targeting HRAS and HMGA2,²⁰ and miR-21, which is highly expressed in breast cancer, is correlated with poor patient survival.²¹ Virtually all CRCs exhibit altered miRNA expression;²² for instance, miR-22 contributes to tumorigenesis in CRC,²³ and downregulation of miR-34 results in CRC metastasis via an increase in IL6 signaling pathway activity.²⁴ These findings suggest the potential of miRNAs as excellent prognostic biomarkers for CRA survival prediction. The heat map generated in our study intuitively indicates that compared with the normal tissues, the CRA tissues exhibited obvious miRNA expression abnormalities, proving the previous statement by Okugawa that almost all CRCs exhibit altered miRNA expression.

Through a comprehensive analysis including Cox regression with the LASSO method on the training set, seven prognostic miRNAs were finally screened. Model evaluation is dependent on the values of the C index and AUC. The higher the C index and AUC values, the better is the model.²⁵ Our risk score formula, termed the seven-miRNA signature, was established with a C index of 0.772. The AUC for 5-year survival was 0.889 in the training set and 0.816 in the entire TCGA set, indicating that the signature had good predictive power. Several types of CRC prognostic score models have recently been developed and verified, including an immunoscore with a C index of 0.612 and AUC of 0.630 for 5-year survival,²⁶ a lncRNA score with an AUC of 0.733 for 5-year survival,²⁷ and so on. However, prognostic models comprising miRNAs for use with CRA patients have not been reported, and our signature filled this gap. Moreover, the performance of the seven-miRNA signatures was evaluated in the training set and verified in the test set and the entire set using ROC curve and Kaplan-Meier survival analysis. The seven-miRNA signature performed well in distinguishing whether patients with CRA had a high or low prognostic risk.

The heat map for the three data set clearly shows the expression patterns of the seven miRNAs in CRA samples. The expression of miR-144 was high in the low-risk group but low in the high-risk group, and the remaining 6 miRNAs exhibited the opposite pattern. These results indicate that miR-144 is the only negatively regulated miRNA with an antitumor effect; thus, miR-144 may be a protective factor for CRA and may become a new therapeutic target. This possibility needs further study. Subsequent GO analysis showed that several of the cancer-related cellular processes and biological pathways related to these genes, such as regulation and negative regulation of apoptosis, positive regulation of cell proliferation, and regulation of protein amino acid phosphorylation, are involved in tumorigenesis and tumor progression. KEGG analysis revealed that the p53 and ErbB signaling pathways were enriched. The p53 signaling has frequently been shown to be associated with poor prognosis in CRC,²⁸ and a significant

proportion of CRCs have been reported to exhibit changes in ErbB signaling.²⁹ These results suggest the participation of the seven miRNAs in CRA development and prognosis through regulation of their target genes.

Furthermore, among the factors related to the prognosis of CRA patients, clinical parameters, such as TNM stage,³⁰ which is currently a prognostic indicator favored by clinicians, are essential. Cox regression analysis revealed three prognostic factors—age, TNM stage and the seven-miRNA signature, among which the seven-miRNA signature was an independent prognostic risk factor. To gain insight into the prognosis of patients with CRA, we performed a transverse ROC analysis and a stratified longitudinal analysis on the entire TCGA set. The AUC of the seven-miRNA signature was superior to that of TNM stage and age for predicting prognosis. Then, the stratification analysis results suggested that the seven-miRNA signature could accurately distinguish high-risk patients from low-risk patients stratified by TNM stage and age, which again demonstrated the predictive power of the seven-miRNA signature.

To the best of our knowledge, this study is the first to report the relationship between a seven-miRNA signature and survival in CRA. However, this study has some limitations. First, although the biological functions of the prognostic miRNAs were annotated through functional enrichment analysis, in vivo and in vitro studies are still needed to further reveal the mechanism by which these prognostic miRNAs mediate CRA tumor development and prognosis. Second, in the sample data obtained from the TCGA database, white ethnic groups were the most common, and data for Asian ethnic groups were scarce. Although our study showed that the prognosis of CRA patients was not correlated with the ethnic group, whether differences exist within ethnic groups is unknown. More data from Asian ethnic groups are needed for further investigation.

Conclusion

In conclusion, we performed scientific data mining for miRNAs from a TCGA dataset of patients with CRA and identified a seven-miRNA signature with good predictive performance. The seven-miRNA signature was successfully validated in the test set and the entire set and was demonstrated to be an independent factor for survival prediction and superior to other conventional clinical variables. Functional enrichment analysis results suggested that the miRNAs in the seven-miRNA signature may be related to a variety of processes and pathways related to tumorigenesis and progression, further supporting the power of the seven-miRNA signature for predicting the prognosis of patients with CRA.

Abbreviations

CRA: colorectal adenocarcinoma; CRC: colorectal cancer; miRNA: MicroRNA; TCGA: The Cancer Genome Atlas; LASSO: least absolute shrinkage and selectionator operator; CLL: lymphocytic leukemia; ROC: receiver operating characteristic; AUCs: ROC curves; DAVID: Database for Annotation, Visualization, and Integrated Discovery; GO: Gene Ontology; BP: biological process; KEGG: Kyoto Encyclopedia of Genes and Genomes; C index: concordance index.

Declarations

Acknowledgements

Not applicable.

Authors' contributions

Shengying Jiang was involved in study designing, writing of this article, data processing, and preparing of figures. Xiaoli Xie was involved in data processing and writing of this article. Huiqing Jiang was involved in developing the idea and proofreading the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (81702324 and 81602529), Natural Science Foundation of Hebei Province (H2018206176 and H2017206141), and the Medical Science Research Project of Hebei Province (20190510).

Availability of data and materials

The datasets analyzed during the current study are available in The Cancer Genome Atlas (TCGA) repository, <http://cancergenome.nih.gov/>.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Thrumurthy SG, Thrumurthy SS, Gilbert CE, Ross P, Haji A. Colorectal adenocarcinoma: risks, prevention and diagnosis. *BMJ*. 2016;354:i3590.
2. Okugawa Y, Toiyama Y, Toden S, Mitoma H, Nagasaka T, Tanaka K, et al. Clinical significance of SNORA42 as an oncogene and a prognostic biomarker in colorectal cancer. *Gut*. 2017;66(1):107–117.

3. Hayama T, Hashiguchi Y, Okamoto K, Okada Y, Ono K, Shimada R, et al. G12V and G12C mutations in the gene KRAS are associated with a poorer prognosis in primary colorectal cancer. *Int J Colorectal Dis.* 2019;34(8):1491–1496.
4. Zheng H, Song K, Fu Y, You T, Yang J, Guo W, Wang K, Jin L, Gu Y, Qi L, Zhao W, Guo Z. A qualitative transcriptional signature for determining the grade of colorectal adenocarcinoma. *Cancer Gene Ther.* 2019;10.1038/s41417-019-0139-1.
5. Kamal Y, Schmit SL, Hoehn HJ, Amos CI, Frost HR. Transcriptomic Differences between Primary Colorectal Adenocarcinomas and Distant Metastases Reveal Metastatic Colorectal Cancer Subtypes. *Cancer Res.* 2019;79(16):4227–4241.
6. He Z, Dang J, Song A, Cui X, Ma Z, Zhang Z. Identification of LINC01234 and MIR210HG as novel prognostic signature for colorectal adenocarcinoma. *J Cell Physiol.* 2019;234(5):6769–6777.
7. Mollaei H, Safaralizadeh R, Rostami Z. MicroRNA replacement therapy in cancer. *J Cell Physiol.* 2019;234(8):12369–12384.
8. Mirzaei H, Fathollahzadeh S, Khanmohammadi R, Darijani M, Momeni F, Masoudifar A, et al. State of the art in microRNA as diagnostic and therapeutic biomarkers in chronic lymphocytic leukemia. *Journal of Cellular Physiology*, 2018;233(2), 888–900.
9. Zhou Y, Li R, Yu H, Wang R, Shen Z. microRNA-130a is an oncomir suppressing the expression of CRMP4 in gastric cancer. *Onco Targets Ther.* 2017;10:3893-3905.
10. Birnbaum DJ, Finetti P, Lopresti A, Gilabert M, Poizat F, Raoul JL, et al. A 25-gene classifier predicts overall survival in resectable pancreatic cancer. *BMC Medicine*, 2017,15(1), 170.
11. Weijun Liu and Xiaowei Wang. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biology.* 2019;20(1):18.
12. Agarwal V, Bell GW, Nam J, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *eLife*, 2015;4:e05005.
13. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* 2018;46(D1):D296-D302.
14. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1):44–57.
15. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009;37(1):1–13.
16. Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat Rev Gastroenterol Hepatol.* 2019;16(12):713–732.
17. Wong SH, Yu J. Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nat Rev Gastroenterol Hepatol.* 2019;16(11):690–704.
18. Katona BW, Weiss JM. Chemoprevention of Colorectal Cancer. *Gastroenterology.* 2020;158(2):368–388.

19. Rupaimoole R, Calin GA, Lopez-Berestein G, Sood AK. miRNA Deregulation in Cancer Cells and the Tumor Microenvironment. *Cancer Discov.* 2016;6(3):235–246.
20. Sung SY, Liao CH, Wu HP, Hsiao WC, Wu IH, Yu J, Lin SH, Hsieh CL. Loss of let-7 microRNA upregulates IL-6 in bone marrow-derived mesenchymal stem cells triggering a reactive stromal response to prostate cancer. *PloS One.* 2013;8:e71637.
21. Frankel LB, Christoffersen NR, Jacobsen A, Lindow M, Krogh A, Lund AH. Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells. *J Biol Chem.* 2008;283:1026–33.
22. Okugawa Y, Grady WM, Goel A. Epigenetic Alterations in Colorectal Cancer: Emerging Biomarkers. *Gastroenterology.* 2015;149(5):1204–1225.e12.
23. Liu Y, Chen X, Cheng R, Yang F, Yu M, Wang C, et al. The Jun/miR-22/HuR regulatory axis contributes to tumorigenesis in colorectal cancer. *Mol Cancer.* 2018;17(1):11.
24. Rokavec M, Oner MG, Li H, Jackstadt R, Jiang L, Lodygin D, et al. IL-6R/STAT3/miR-34a feedback loop promotes EMT-mediated colorectal cancer invasion and metastasis. *J Clin Invest.* 2014;124:1853–67.
25. Li B, Cui Y, Diehn M, Li R. Development and Validation of an Individualized Immune Prognostic Signature in Early-Stage Nonsquamous Non-Small Cell Lung Cancer. *JAMA Oncol.* 2017;3(11):1529–1537.
26. Zhao X, Liu J, Liu S, Yang F, Chen E. Construction and Validation of an Immune-Related Prognostic Model Based on TP53 Status in Colorectal Cancer. *Cancers (Basel).* 2019;11(11):1722.
27. Fan Q, Liu B. Discovery of a novel six-long non-coding RNA signature predicting survival of colorectal cancer patients. *J Cell Biochem.* 2018;119(4):3574-3585.
28. Li XL, Zhou J, Chen ZR, Chng WJ. P53 mutations in colorectal cancer - molecular pathogenesis and pharmacological reactivation. *World J Gastroenterol.* 2015;21(1):84–93.
29. Ross JS, Fakih M, Ali SM, Elvin JA, Schrock AB, Suh J, et al. Targeting HER2 in colorectal cancer: The landscape of amplification and short variant mutations in ERBB2 and ERBB3. *Cancer.* 2018;124(7):1358–1373.
30. Lea D, Håland S, Hagland HR, Søreide K. Accuracy of TNM staging in colorectal cancer: a review of current culprits, the modern role of morphology and stepping-stones for improvements in the molecular era. *Scand J Gastroenterol.* 2014;49(10):1153–1163.

Tables

Table 1 Clinical data for the CRA patients in each data set

Clinical information	Training set n (%)	Testing set n (%)	Entire set n (%)	<i>P-value</i>
N	208	207	415	
Age(year)				
<65	97 (46.6)	92 (44.4)	189 (45.5)	0.905
>=65	111 (53.4)	115 (55.6)	226 (54.5)	
Gender				
Female	105 (50.5)	90 (43.5)	195 (47.0)	0.36
Male	103 (49.5)	117 (56.5)	220 (53.0)	
Race				
Asian	7 (3.4)	2 (1.0)	9 (2.2)	0.771
Black	108 (51.9)	105 (50.7)	213 (51.3)	
White	66 (31.7)	74 (35.7)	140 (33.7)	
Not reported	27 (13.0)	26 (12.6)	53 (12.8)	
Tumor site				
Rectum	27 (13.0)	36 (17.4)	63 (15.2)	0.814
Rectosigmoid junction	28 (13.5)	26 (12.6)	54 (13.0)	
Colon	153 (73.6)	145 (70.0)	298 (71.8)	
TNM stage				
I	39 (18.8)	40 (19.3)	79 (19.0)	0.999
II	74 (35.6)	76 (36.7)	150 (36.1)	
III	63 (30.3)	57 (27.5)	120 (28.9)	
IV	32 (15.4)	34 (16.4)	66 (15.9)	
Survival status				
Alive	164 (78.8)	168 (81.2)	332 (80.0)	0.841
Dead	44 (21.2)	39 (18.8)	83 (20.0)	

Table 2 Univariate and multivariate Cox regression analysis results for the survival risk of CRA patients across the entire set.

Variables	Univariate Cox			Multivariate Cox		
	HR	95% CI of HR	P	HR	95% CI of HR	P
Gender(Female vs. Male)	0.995	0.646-1.533	0.983	-0.279	0.608-1.454	0.780
Race(Asian vs. Black vs. White vs. Not reported)	1.003	0.732-1.375	0.983	-0.874	0.602-1.215	0.382
Age(<65 vs. >=65)	2.101	1.338-3.299	0.001	4.017	1.634-4.164	5.89E-05
Tumor site(Rectum vs. Rectosigmoid junction vs. Colon)	1.054	0.775-1.435	0.736	0.580	0.799-1.514	0.562
TNM stage(I+II vs. III+IV)	3.296	2.060-5.273	6.56E-07	4.751	1.983-5.185	2.02E-06
Seven-miRNA signature(Low vs. High)	4.365	2.557-7.453	6.66E-08	4.838	2.218-6.565	1.31E-06

Table 3 Main GO BP terms enriched with the seven miRNAs in the signature

GO id	Term	Count	PValue	FDR
GO:0043066	negative regulation of apoptosis	9	4.84E-05	0.077
GO:0043069	negative regulation of programmed cell death	9	5.34E-05	0.085
GO:0060548	negative regulation of cell death	9	5.45E-05	0.087
GO:0042981	regulation of apoptosis	12	1.63E-04	0.259
GO:0043067	regulation of programmed cell death	12	1.77E-04	0.283
GO:0010941	regulation of cell death	12	1.83E-04	0.292
GO:0010033	response to organic substance	11	3.04E-04	0.484
GO:0042127	regulation of cell proliferation	11	6.07E-04	0.966
GO:0008284	positive regulation of cell proliferation	8	8.65E-04	1.372
GO:0006916	anti-apoptosis	6	0.001	1.644
GO:0010608	posttranscriptional regulation of gene expression	6	0.001	1.828
GO:0051412	response to corticosterone stimulus	3	0.002	2.551
GO:0042325	regulation of phosphorylation	8	0.002	2.704
GO:0018107	peptidyl-threonine phosphorylation	3	0.002	3.226
GO:0000902	cell morphogenesis	7	0.002	3.320
GO:0051174	regulation of phosphorus metabolic process	8	0.002	3.382
GO:0019220	regulation of phosphate metabolic process	8	0.002	3.382
GO:0031099	regeneration	4	0.002	3.487
GO:0009725	response to hormone stimulus	7	0.002	3.857
GO:0051385	response to mineralocorticoid stimulus	3	0.003	3.973
GO:0016071	mRNA metabolic process	7	0.003	4.014
GO:0043434	response to peptide hormone stimulus	5	0.003	4.170
GO:0018210	peptidyl-threonine modification	3	0.003	4.372
GO:0032989	cellular component morphogenesis	7	0.004	5.642
GO:0009719	response to endogenous stimulus	7	0.004	6.203

GO:0001932 regulation of protein amino acid phosphorylation	5	0.004	6.265
GO:0022403 cell cycle phase	7	0.004	6.881
GO:0048729 tissue morphogenesis	5	0.005	7.179
GO:0031399 regulation of protein modification process	6	0.005	7.593
GO:0022402 cell cycle process	8	0.005	7.711
GO:0043065 positive regulation of apoptosis	7	0.005	8.210
GO:0043068 positive regulation of programmed cell death	7	0.006	8.477
GO:0010942 positive regulation of cell death	7	0.006	8.659
GO:0007568 aging	4	0.008	12.331
GO:0032268 regulation of cellular protein metabolic process	7	0.008	12.738
GO:0045859 regulation of protein kinase activity	6	0.009	14.031
GO:0043549 regulation of kinase activity	6	0.011	15.947
GO:0051338 regulation of transferase activity	6	0.013	18.539
GO:0032990 cell part morphogenesis	5	0.016	22.221
GO:0048762 mesenchymal cell differentiation	3	0.016	22.448
GO:0016477 cell migration	5	0.020	27.582
GO:0008380 RNA splicing	5	0.022	29.851
GO:0001775 cell activation	5	0.023	30.717
GO:0008283 cell proliferation	6	0.024	31.796
GO:0045471 response to ethanol	3	0.024	32.387
GO:0009057 macromolecule catabolic process	8	0.026	34.658
GO:0006917 induction of apoptosis	5	0.032	40.668
GO:0006397 mRNA processing	5	0.032	40.978
GO:0012502 induction of programmed cell death	5	0.032	40.978
GO:0008285 negative regulation of cell proliferation	5	0.047	53.408

Figures

Figure 1

Heatmap generated by unsupervised hierarchical clustering of the differentially expressed miRNAs between cancer tissues and normal tissues in CRA patients.

Figure 2

Construction of the seven-miRNA signature. (A) LASSO coefficient profiles of the miRNAs associated with the prognosis of patients with CRA in the training set; (B) The λ value was determined by lambda.min.

Figure 3

Capability evaluation in the training set. (A) The time-dependent ROC curves for 3-year survival and 5-year survival; (B) Kaplan-Meier curves showing that the overall survival rate in the low-risk group was significantly higher than that in the high-risk group.

Figure 4

Risk score analysis in the training set. (A) Distribution of risk scores; (B) Patient survival time and status. The red dotted line indicates the optimum cutoff value for dividing patients into the low-risk and high-risk groups. (C) Heat map of the miRNAs in the prognostic signature.

Figure 5

Performance verification in the test set. (A) ROC curve analysis of the seven-miRNA signature; (B) Kaplan-Meier overall survival analysis of the CRA patients using the seven-miRNA signature; (C) Distributions of the risk scores, survival statuses and expression profiles of the seven miRNAs in the CRA patients.

Figure 6

Performance verification in the entire set. (A) ROC curve analysis of the seven-miRNA signature; (B) Kaplan-Meier overall survival analysis of CRA patients using the seven-miRNA signature; (C) Distributions of the risk scores, survival statuses and expression profiles of the seven miRNAs in the CRA patients.

Figure 7

Comparison of the prognostic value of the seven-miRNA signature and clinical parameters. (A) The time-dependent ROC curves for follow-up were plotted to assess the prognostic efficacy of age, miRNA signature, and a new variable combining both of these variables. (B) Time-dependent ROC curves for follow-up were plotted to assess the prognostic efficacy of the TNM stage, miRNA signature, and a new variable combining both of these variables.

Figure 8

Stratification analysis by age and TNM visualized by Kaplan-Meier overall survival curves in the entire set. (A) the younger group; (B) the aged group; (C) the early-stage group; (D) the advanced-stage group.

Figure 9

Prediction of target genes of the miRNAs in the seven-miRNA signature and functional enrichment analysis of these genes. (A) GO BP enrichment analysis; (B) The predicted target genes of the miRNAs in the seven-miRNA signature were defined as the intersection of the genes predicted by miRTarBase, TargetScan and miRDB; (C) KEGG signaling pathway analysis.