

# Identifying autophagy gene-associated module biomarkers through construction and analysis of an autophagy-mediated ceRNA-ceRNA interaction network in colorectal cancer

KUN QIAN<sup>1\*</sup>, HUIYING HUANG<sup>1\*</sup>, JING JIANG<sup>2\*</sup>, DAHUA XU<sup>1</sup>, SHENGNAN GUO<sup>1</sup>,  
YING CUI<sup>1</sup>, HAO WANG<sup>2</sup>, LIQIANG WANG<sup>1</sup> and KONGNING LI<sup>1</sup>

<sup>1</sup>College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang 150081; <sup>2</sup>Obstetrics and Gynecology Department, The Second Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang 150001, P.R. China

Received December 27, 2017; Accepted April 10, 2018

DOI: 10.3892/ijo.2018.4443

**Abstract.** Autophagy is crucial in cellular homeostasis and has been implicated in the development of malignant tumors. However, the regulatory function of autophagy in cancer remains to be fully elucidated. In the present study, the autophagy-mediated competing endogenous RNA (ceRNA)-ceRNA interaction networks in colorectal cancer (CRC) were constructed by integrating systematically expression profiles of long non-coding RNAs and mRNAs. It was found that a large proportion of autophagy genes were inclined to target hub nodes, including a fraction of autophagy genes, by comparing with other genes within ceRNA networks, and showed preferential interaction with themselves. The present study also revealed that autophagy genes may be used as prognostic markers for cancer therapy. A risk score model based on multivariable Cox regression analysis was then used to capture novel biomarkers in connection with lncRNA for the prognosis of CRC. These biomarkers were confirmed in the test dataset and an additional independent dataset. Furthermore, the prognostic value of biomarkers is independent of conventional clinical factors. These results provide improved understanding of autophagy-mediated ceRNA regulatory mechanisms in CRC and provide novel potential molecular therapeutic targets for the diagnosis and treatment of CRC.

## Introduction

The role of autophagy in cancer has been the subject of numerous studies, which have attempted to dissect the complex functions in the process of tumor development. In previous years, increasing evidence has indicated that autophagy is an evolutionarily conserved mechanism of adaptation to sustain homeostasis against stress in adverse microenvironmental conditions (1). Autophagy was originally considered to be a tumor suppression mechanism in human breast cancer, ovarian cancer and colorectal cancer (CRC) (2,3). However, emerging data revealed that the dysregulation of autophagy function not only led to the disruption of physiological processes, but also enabled the creation of an environment for cancer initiation (4). For example, activation of cancer pathways within tumor cells can increase the requirement for autophagy to promote tumor growth and survival (5). These studies provide key insights into the potential role of autophagy in tumor biology. A previous study by Mizushima *et al* showed that autophagy genes controlled the process of autophagy in cells (6). Autophagy genes, including mRNAs or long non-coding RNAs (lncRNAs) may serve for competing endogenous RNAs (ceRNAs) as the basic component of the ceRNA regulatory network and be critical in tumor development. However, the biomedical significance of autophagy genes in the ceRNA-ceRNA interaction network remains to be fully elucidated. In particular, it remains unclear whether autophagy genes effectively act as novel potential biomarkers or therapeutic targets to combat cancer, including CRC. Therefore, it is essential to perform a systematic analysis of the autophagy-mediated ceRNA-ceRNA interaction network to evaluate the potential clinical use of autophagy genes.

Taking advantage of RNA-sequencing transcriptomic data of The Cancer Genome Atlas (TCGA) and information on autophagy genes collected from autophagy-associated databases, the present study constructed the autophagy-mediated ceRNA-ceRNA interaction networks in CRC [including colon adenocarcinoma (COAD) and rectal adenocarcinoma (READ)] by systematically analyzing the molecular profiles of CRC

---

*Correspondence to:* Professor Kongning Li or Mr. Liqiang Wang, College of Bioinformatics Science and Technology, Harbin Medical University, 157 Baojian Road, Harbin, Heilongjiang 150081, P.R. China  
E-mail: likongning@ems.hrbmu.edu.cn  
E-mail: wangliqiang0619@163.com

\*Contributed equally

**Key words:** autophagy, competing endogenous RNA, interaction network, colorectal cancer, prognostic biomarkers

and microRNA (miRNA) regulations, respectively. Through in depth analysis of the ceRNA networks, it was found that autophagy genes have a major role in the ceRNA networks of CRC, and autophagy-related triplets encompassing ceRNAs and miRNAs have a direct association with the survival rate of patients with colon cancer and rectal cancer. The present study also assessed the potential clinical utility of modules consisting of autophagy-related triplets for predicting patient survival rates. The analyses of autophagy gene-associated ceRNA network provided an improved understanding of the mechanism of action of autophagy in CRC and may provide a novel outlook into clinical therapeutic modality development for CRC.

## Materials and methods

**Data resources.** The gene expression and clinical information of CRC were derived from TCGA 2015 (<http://cancergenome.nih.gov/>) (7), which provided the miRNASeq and RNASeqV2 data, including 246 patients with COAD and 84 patients with rectal cancer READ. The patients were filtered by age, stage and chemotherapy information separately for further analysis (Table I). A total of 239 and 221 patients of the total of 246 had information on stage and chemotherapy, respectively (Table I).

Information of exons mapping to coding/lncRNA genes were obtained from GENCODE Release 19 (8), and those lncRNA exons that overlapped with any known coding genes were filtered out. The lncRNA expression profile of CRC was recalculated using the method of Wang *et al* (9), and only types of 'antisense', 'non-coding', 'processed transcript' and 'lincRNA' with lengths of >200 nt were considered as lncRNAs. The reads per kilobase per million reads (RPKM) were then recalculated to quantify the expression levels of lncRNAs (9). All expression profiles in the present study were preprocessed by zero line-deleted, adding 0.05 for zero and log<sub>2</sub>-transformed.

A cohort of 63 patients with COAD from the Gene Expression Omnibus (GEO) database (accession no. GSE29623) (10) was used as an independent external test set. Replicate genes were combined by averaging their expression values. We were not able to find suitable external verification data with both mRNA/miRNA expression and clinical data on the GEO database for READ. Therefore, there was no GEO validation set for READ in our study.

A total of 433,617 pairs of experimental interactions between miRNAs and lncRNAs/mRNAs were downloaded from starBase v2.0 (hg19) (11). The autophagy genes were collected from the Cancer Proteomics database (12) (<http://celldeathproteomics.uio.no/>), ncRDeathDB (13) (<http://www.rna-society.org/ncrdeathdb/>), and miRDeathDB (14) (<http://rna-world.org/mirdeathdb/>). Following the union of the three databases and removal of the redundant or unidentified genes, 1,973 RNAs associated with autophagy were retained, which included 194 miRNAs, 1,775 mRNAs, and four lncRNAs.

**Construction and analysis of the CRC ceRNA-ceRNA network.** For construction of the network, the following two principles were mainly followed to identify ceRNA pairs in CRC.

i) Predicting co-regulated gene pairs: For a given gene pair (g1 and g2), miRNAs that regulate the two genes were first identified, followed by measurement of whether these two genes

significantly shared miRNAs using a hypergeometric test. The P-value was calculated according to the following formula:

$$p = 1 - F(x | N, K, M) = 1 - \sum_{i=0}^{x-1} \frac{c(K, i) \times c(N - K, M - i)}{c(N, M)}$$

N represents the number of all human miRNAs, which is also termed background distribution, K represents the whole number of miRNAs regulating g1, M represents the whole number of miRNAs regulating g2, and x represents the number of shared miRNAs between g1 and g2. All P-values were adjusted by the Bonferroni correction and only the pairs with adjusted P<0.005 were considered as significantly co-regulated gene pairs.

ii) Predicting co-expression gene pairs: The Pearson's correlation coefficients were calculated for each pair of genes in the co-regulated network. A pair with a correlation coefficient >0 and a false discovery rate (FDR)<0.005 was considered as co-expressed. Using the two steps above, the CRC ceRNA-ceRNA network was constructed.

**Construction of the human CRC autophagy-mediated ceRNA-ceRNA interaction network.** At first, an miRNA was added into the ceRNA network if the miRNA was shared by any pair of ceRNA genes in the ceRNA network. Furthermore, if at least two of any triplet were autophagy genes in the miRNA-mediated ceRNA-ceRNA interaction network, the triplets were added to the autophagy-mediated ceRNA-ceRNA interaction network. Eventually, the autophagy-mediated ceRNA-ceRNA interaction network was constructed.

**Prognostic triplets.** Univariate Cox regression analysis was used to evaluate the association between the expression level of each gene and survival rate. Triplets in the autophagy-mediated ceRNA-ceRNA interaction network were defined as prognostic triplets when all three genes of the triplet had a P-value of P<0.05 in TCGA.

**Prognostic modules.** The multivariate Cox proportional hazards model was used to determine the association between the expression level of prognostic triplets and survival rate. The corresponding P-values of the log-rank tests were respectively calculated on the TCGA training set, TCGA test set and GEO datasets.

The prognostic triplets with log-rank P<0.05 in the training, test and GEO datasets were preserved. The GEO verification for READ was not considered due to the lack of GEO data. The linked gene block of the preserved prognostic triplets with the highest number of genes was considered as a gene module, and the module was visualized by Cytoscape (15) and Rcirco 1.2.0 (16).

**Grouping of specimens into different risk groups using the multivariate Cox proportional hazards regression model.** The multivariate Cox proportional hazards regression model was used to calculate the respective risk coefficients based on the gene expression of the input gene set and survival rates, and a risk score was assigned to each patient. The risk score for each patient was calculated as follows:

$$Risk\_score = \sum_{j=1}^n \beta_j * Expression_{gene(j)}$$

Table I. Clinical characteristics of the patients with colorectal cancer from The Cancer Genome Atlas.

Parameter	COAD (246)		READ (84)	
	Patients (n)	Mean $\pm$ SD survival (days)	Patients (n)	Mean $\pm$ SD survival (days)
Age (years)				
≤65	121	848.37 $\pm$ 80.0	45	958.8 $\pm$ 12.1
>65	125	975.99 $\pm$ 75.8	39	781.94 $\pm$ 89.0
Stage				
I	37	808.77 $\pm$ 42.7	-	-
II	97	1,071.79 $\pm$ 91.1	-	-
I&II	-	-	35	1,064.58 $\pm$ 71.3
III	74	908.39 $\pm$ 04.5	-	-
IV	31	624.16 $\pm$ 27.0	-	-
III&IV	-	-	45	750.34 $\pm$ 80.1
Chemotherapy				
Yes	96	840.26 $\pm$ 50.3	43	908.67 $\pm$ 81.1
No	115	786.16 $\pm$ 69.0	36	792.75 $\pm$ 74.7

COAD, colon adenocarcinoma; READ, rectal adenocarcinoma.

$\beta_j$  is the multivariate Cox regression coefficient of gene  $j$  in the input gene set  $G$ ,  $n$  is the number of genes in  $G$ , and  $Expression_{gene(i)}$  is the expression of gene  $i$  in this sample. All patients were segregated into high- and low-risk groups using the median risk score as the cut-off point.

## Results

### Importance of autophagy genes in the ceRNA CRC network.

The expression profiles from TCGA and interaction information from StarBase v2.0 were integrated to identify the ceRNA-ceRNA interaction network, and the lncRNA expression profile of CRC was recalculated using method of Wang *et al* (9). As described above, ceRNA-ceRNA interaction networks of the patients with CRC were constructed, which included 91,185 edges (88,388 mRNA-mRNAs; 2,722 mRNA-lncRNAs; 75 lncRNA-lncRNAs) and 5,812 nodes (5,696 mRNAs, 116 lncRNAs) for COAD (Fig. 1A, left panel), and 50,141 edges (48,845 mRNA-mRNAs; 1,266 mRNA-lncRNAs; 30 lncRNA-lncRNAs) and 4,956 nodes (4,865 mRNAs, 91 lncRNAs) for READ (Fig. 1B, left panel). The topological analysis suggested that the CRC ceRNA network exhibited scale-free characteristics typical of biological networks with power-law distribution (Fig. 1A, top right panel, COAD, 0.8938; Fig. 1B, top right panel, READ, 0.9125) and the clustering coefficient of the CRC ceRNA network was significantly higher, compared with that of the random network (empirical  $P < 0.001$ ) (Fig. 1A, lower right panel and Fig. 1B, lower right panel). In addition, the degrees of genes in the ceRNA networks were analyzed, and it was found that the degrees of autophagy genes were significantly lower, compared with those of the non-autophagy genes (Fig. 2A, left panel and Fig. 2B, left panel using the Wilcoxon

Table II. Summary of the nodes in colorectal cancer ceRNA-ceRNA interaction networks.

Node	COAD (n)	READ (n)
Hub and autophagy	104	96
Non-hub and autophagy	640	515
Hub and non-autophagy	1,059	892
Non-hub and non-autophagy	4,009	3,453
Total	5,812	4,956

COAD, colon adenocarcinoma; READ, rectal adenocarcinoma; ceRNA, competing endogenous RNA.

Table III. Summary of the autophagy-mediated ceRNA-ceRNA interaction networks in colorectal cancer.

Attribute	COAD (n)	READ (n)
Node		
Auto-miRNA	101	100
miRNA	228	220
Auto-mRNA	738	606
mRNA	3,118	2,337
Auto-lncRNA	0	0
lncRNA	55	38
Total	4,240	3,301
Edge		
miRNA-mRNA	86,286	68,075
miRNA-lncRNA	600	422
mRNA-mRNA	14,986	8,578
lncRNA-lncRNA	0	0
mRNA-lncRNA	247	111
Total	102,119	77,186

COAD, colon adenocarcinoma; READ, rectal adenocarcinoma; miRNA, microRNA; lncRNA, long non-coding RNA.

rank sum test (COAD,  $P = 1.261e-06$ ; READ,  $P = 0.03379$ ;  $P < 0.05$ ), and autophagy genes were inclined to interact with autophagy genes (Fig. 2A, middle panel and Fig. 2B, middle panel, Wilcoxon rank sum test). It was also found that only a small proportion (COAD, 13.98%; READ, 15.71%) of autophagy genes were hubs, with the majority of autophagy genes being target hub nodes (Fig. 2A, right panel, Fig. 2B, right panel, Table II, Kruskal-Wallis test) by comparing with other genes within the ceRNA network. The top 20% of genes were defined as hub genes. It was demonstrated that autophagy genes function mainly through regulating hubs or acting as a hub itself. The present study also constructed autophagy-mediated ceRNA-ceRNA interaction networks for CRC, comprising 4,240 nodes and 102,119 edges for COAD (256,659 triplets), and 3,301 nodes and 77,186 edges for READ (162,576 triplets), respectively. The details are shown in Table III.

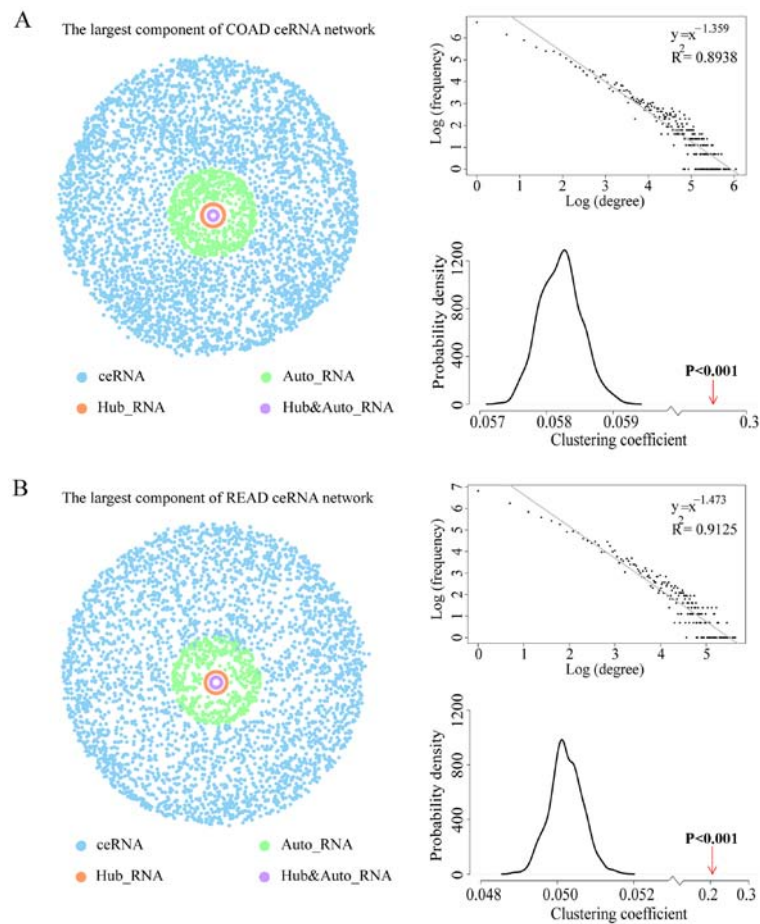


Figure 1. Layouts and analyses of colorectal cancer ceRNA-ceRNA interaction networks. Layouts of the largest component ceRNA network in (A) COAD and (B) READ are shown on the left. A node colored blue represents a ceRNA, a node colored green represents an autophagy gene, a node colored orange represents a hub gene, and a node colored purple represents a hub node related to autophagy. The upper graphs show the degree distribution of the COAD and READ ceRNA network, respectively. The lower graphs show that the clustering coefficient of the COAD and READ ceRNA network is significantly higher than that of the random network, respectively. COAD, colon adenocarcinoma; READ, rectal adenocarcinoma; ceRNA, competing endogenous RNA.

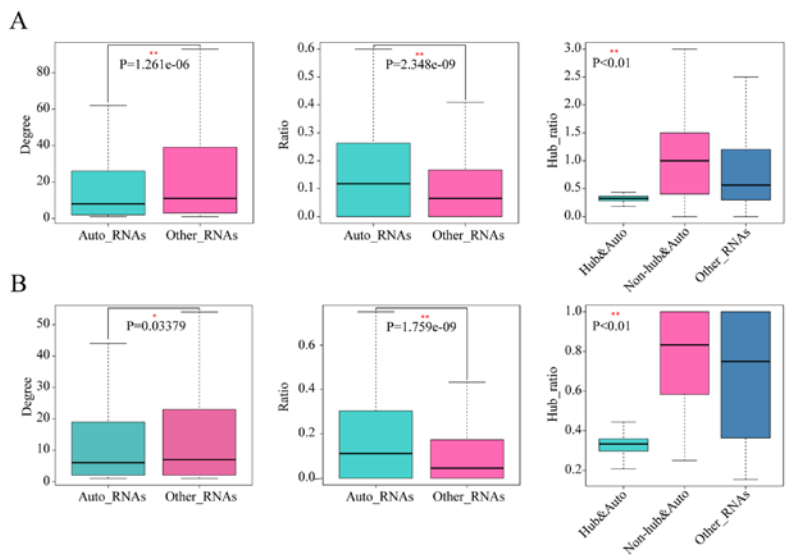


Figure 2. Analyses of autophagy genes and other genes within colorectal cancer ceRNA networks. Results for (A) COAD and (B) READ show the degree of autophagy genes and non-autophagy genes (left panel), the ratios of autophagy genes to non-autophagy genes in the neighbors of autophagy genes and non-autophagy genes (ratio = number of autophagy genes in neighbors / number of non-autophagy genes in neighbors) (middle panel), and the comparison of ratios of non-autophagy and hub genes to non-autophagy and non-hub genes in the neighbors among hub and autophagy genes, non-hub and autophagy genes and other genes in ceRNA networks (ratio = number of non-autophagy and hub genes in neighbors / number of non-autophagy and non-hub genes in neighbors) (right panel). The P-value was calculated by the Wilcoxon rank sum test (left and middle panels) or Kruskal-Wallis test (right panels). \*P<0.05; \*\*P<0.01. Auto\_RNAs, autophagy genes; Other\_RNAs, non-autophagy genes; Hub&Auto\_RNAs, autophagy and hub genes; Non-hub&Auto\_RNAs, autophagy and non-hub genes; COAD, colon adenocarcinoma; READ, rectal adenocarcinoma; ceRNA, competing endogenous RNA.

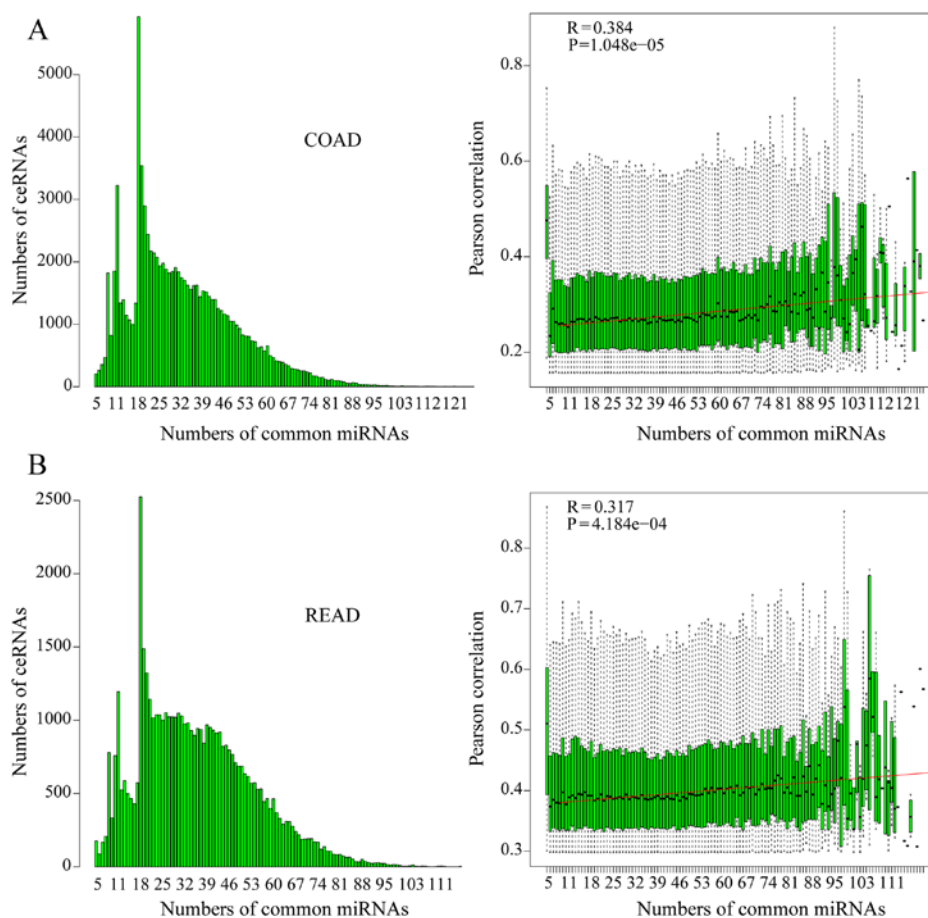


Figure 3. Analyses of colorectal cancer autophagy-mediated ceRNA-ceRNA interaction networks. Results for (A) COAD and (B) READ. Associations between the numbers of shared miRNAs and the numbers of ceRNA pairs are shown on the left. Associations between the numbers of shared miRNAs and Pearson correlation coefficients of ceRNA pairs are shown on the right. The red line shows fitting of the medians of Pearson correlation coefficients. The P-value and correlation coefficient R of fitting were calculated using Pearson correlation analysis. COAD, colon adenocarcinoma; READ, rectal adenocarcinoma; ceRNA, competing endogenous RNA; miRNAs, microRNAs.

The associations between the numbers of shared miRNAs and the numbers of ceRNA pairs, and the Pearson correlation coefficients in the autophagy-mediated ceRNA-ceRNA interaction networks, are shown in Fig. 3. There existed a common finding that, as the number of common miRNAs increased, the number of ceRNAs increased at first and then decreased in COAD and READ (Fig. 3A, left panel, and Fig. 3B, left panel). It was also found that the co-expression of ceRNAs in the network increased with the number of common miRNAs (Fig. 3A, right panel, and Fig. 3B, right panel for COAD ( $P=1.048e-05$ ,  $R=0.384$ ) and READ ( $P=4.184e-04$ ,  $R=0.317$ ).

*Identification of triplets significantly associated with survival rates in CRC.* Previous studies have suggested that complete triplets encompassing ceRNAs and miRNAs are associated with critical biological functions and the survival rates of patients with cancer (9). Univariate Cox regression analysis was used to calculate the risk coefficient and P-value of each gene in the autophagy-mediated ceRNA-ceRNA interaction networks, and all triplets in the networks contained at least two autophagy genes. Certain triplets were found to be associated with the prognosis of CRC (17).

To further verify the prognostic ability of triplets, the patients were divided into two subgroups (training group and

test group) containing equal numbers. No significant differences were observed between the two subgroups using the Wilcoxon rank sum test for gender, age, weight or stages (all  $P>0.05$ ). The multivariate Cox proportional hazards regression model was used to calculate the respective risk coefficients based on the gene expression of the triplet. All patients in the TCGA training data set were then segregated into high- and low-risk groups using the median risk score as the cut-off point. The overall survival time for the high- and low-risk groups were estimated using the Kaplan-Meier method (18-21). The log-rank test was used to analyze the differences in the survival times.

A similar multivariate Cox regression analysis was performed in the TCGA test dataset, and in GSE29623 for COAD. Only the prognostic triplets with log-rank  $P<0.05$  in the TCGA training set, TCGA test set and GEO dataset were preserved (TCGA training set and TCGA test set for READ). As there was no lncRNA expression data for GSE29623, multivariate Cox regression analysis was not required in GSE29623 for the triplets containing lncRNA. Finally, 37 (COAD) and nine (READ) triplets with prognostic abilities were obtained.

*Construction and analysis of prognostic modules in CRC, and validation in additional independent test cohorts.* It was

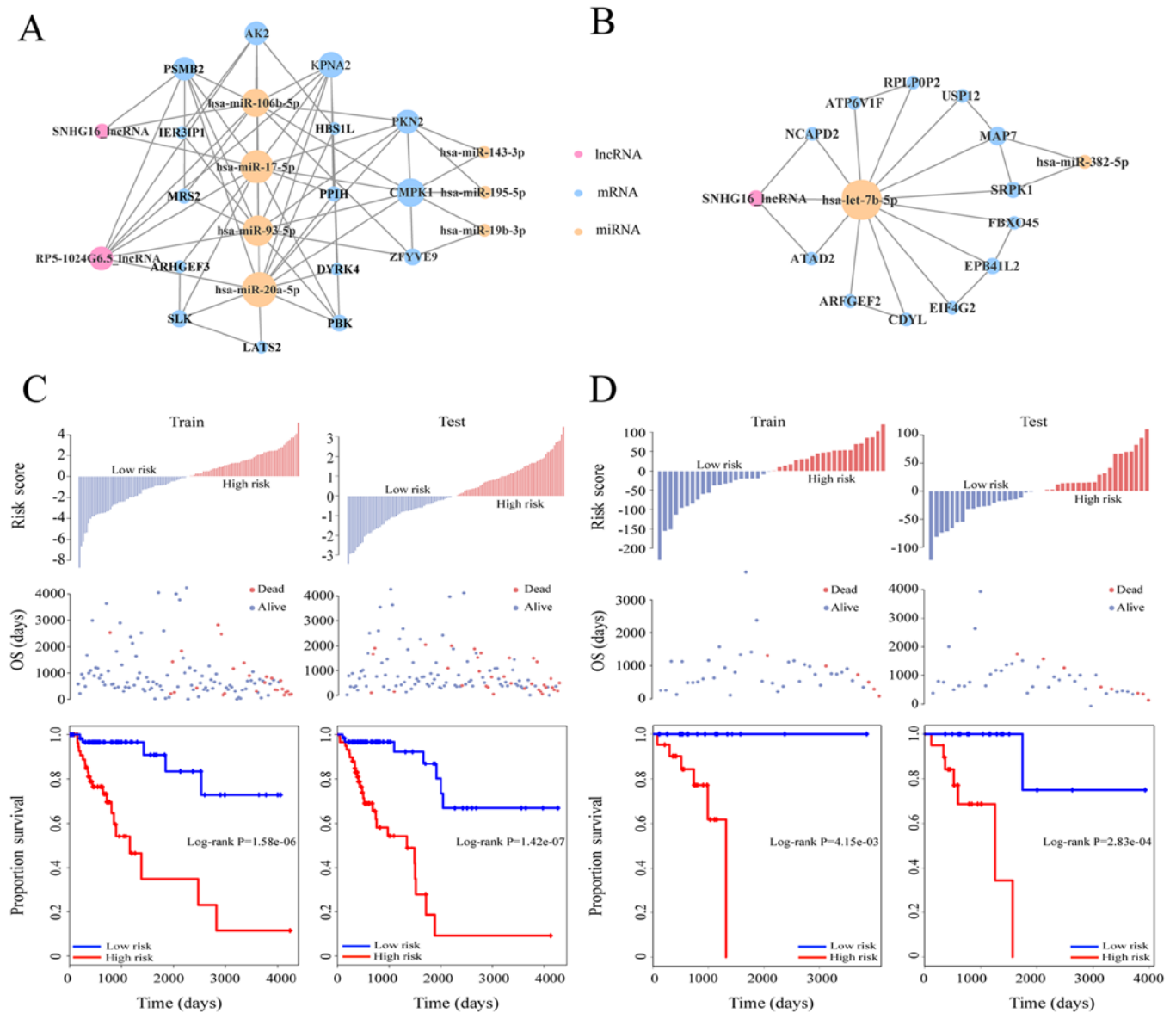


Figure 4. Layouts and analyses of the 24-gene module for COAD and 15-gene module for READ. (A) Layout of the COAD module. (B) Layout of the READ module. A node colored pink represents an lncRNA, a node colored blue represents an mRNA, and a node colored orange represents a miRNA. For the (C) 24-gene COAD module and (D) 15-gene READ module, the module based risk score distribution and patients' survival status in the training and test sets are shown above, and Kaplan-Meier estimates of the OS of patients with COAD according to the module signature are shown below. COAD, colon adenocarcinoma; READ, rectal adenocarcinoma; miRNA, microRNA; lncRNA, long non-coding RNA; OS, overall survival.

found that 34 of the 37 triplets in COAD and nine triplets in READ exhibited interaction wiring, and they were considered as a 24-gene module and 15-gene module. It was found that both modules had higher clustering coefficients than random conditions (random 1,000 times,  $P < 0.01$ ). The interaction wiring of the module genes was visualized for COAD (Fig. 4A) and READ (Fig. 4B) by Cytoscape (15). The two clustering modules were analyzed in subsequent investigations.

The survival curves of the module were estimated using the Kaplan-Meier method, and similar methods of scoring and classifying of the samples were used to classify the samples using the expression of all genes in the modules instead of triplets. This was performed for the COAD module in the TCGA training set, TCGA test set, and GSE29623 dataset. For the READ module, this was performed in the TCGA training set and TCGA test set.

The above analyses demonstrated that the overall survival rates were significantly different between the high-risk group and the low-risk group of patients with CRC for COAD (training log-rank  $P = 1.58e-06$ ; test log-rank  $P = 1.42e-07$ ) and READ (training log-rank  $P = 4.15e-03$ , test log-rank  $P = 2.83e-04$ ). The independent data validation of COAD also indicated that the module was stable (GSE29623 log-rank  $P = 2.18e-05$ ). The survival curves for COAD and READ are shown in Fig. 4C and D. The prognostic efficiency of the gene module was significantly higher, compared with that of the triplets (Fig. 4C and D, lower panels, Table IV). The Kaplan-Meier estimates of overall survival rates of two triplets with low P-values in the analysis of internal and external independent datasets with multivariate Cox proportional hazards regression model are shown in Table IV, with PDZ binding kinase (PBK) and microtubule associated protein 7

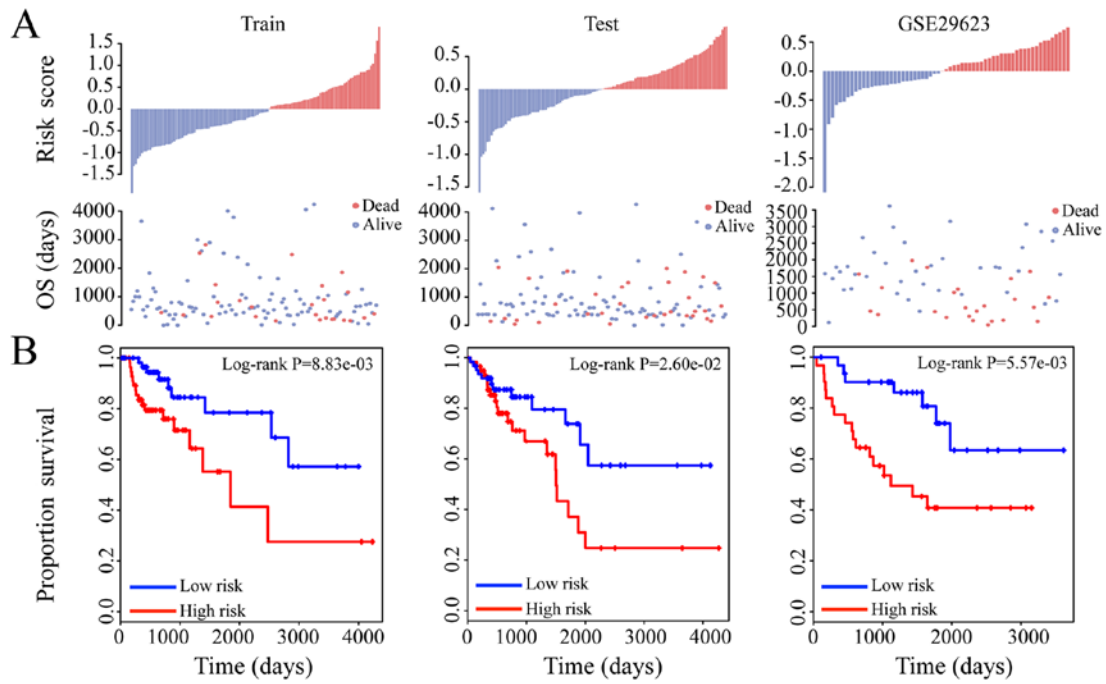


Figure 5. Example of prognostic triplets (MIMAT0000075, PDZ binding kinase and karyopherin subunit  $\alpha 2$ ) in COAD. (A) Distribution of patients' risk score and survival status of the triplet in the training, test and Gene Expression Omnibus datasets. (B) Kaplan-Meier estimates of OS of patients with COAD according to the triple signature. COAD, colon adenocarcinoma; OS, overall survival.

Table IV. Log-rank P-values of two examples of the triple signatures in colorectal cancer.

Cancer	Triplet	Training	Test	GSE29623
COAD	MIMAT0000075, PBK, KPNA2	8.83e-03	2.60e-02	5.57e-03
READ	MIMAT0000063, MAP7, SRPK1	6.14e-03	1.82e-02	-

COAD, colon adenocarcinoma; READ, rectal adenocarcinoma; PBK, PDZ binding kinase; KPNA2, karyopherin subunit  $\alpha 2$ ; MAP7, microtubule associated protein 7; SRPK1, SRSF protein kinase 1.

(MAP7) in the triplets associated with CRC (22,23). The distribution of patient risk score and survival analysis of the triplet [MIMAT0000075, PBK, karyopherin subunit  $\alpha 2$  (KPNA2)] in COAD are also shown in Fig. 5A and B. The corresponding P-values of the log-rank tests in the module signatures were superior to those of triplet signatures. In addition, the mortality rate of the high-risk group was higher than that of the low-risk group, and this trend was more evident in the module analysis from the distribution of the survival status (Fig. 4C, upper panel, and Fig. 5A).

*Predictive capacity of autophagy-related prognostic module for assessing clinical outcome of CRC.* To confirm the stability of the modules in autophagy-mediated ceRNA-ceRNA interaction network, the survival rates between the high-risk and low-risk groups were compared by considering stages, age, and chemotherapy, respectively. Stages I and II were merged into one group, and stage III and IV were merged into one group for READ due to the lack of READ samples. The survival curves were estimated using the method described above.

There were significant differences in survival rates between the high-risk and low-risk groups. The genetic model

Table V. Log-rank P-values of stratification analyses of all patients with age, chemotherapy and stage information using the colorectal cancer module signatures.

Parameter	COAD	READ
Age (years)		
≤65	2.16e-04	6.14e-02
>65	2.95e-05	1.54e-04
Chemotherapy		
Yes	2.81e-06	8.68e-03
No	1.76e-04	2.67e-03
Stage		
I	2.78e-02	-
II	2.51e-05	-
I&II	-	9.10e-03
III	1.61e-05	-
IV	8.83e-08	-
III&IV	-	1.02e-03

COAD, colon adenocarcinoma; READ, rectal adenocarcinoma.

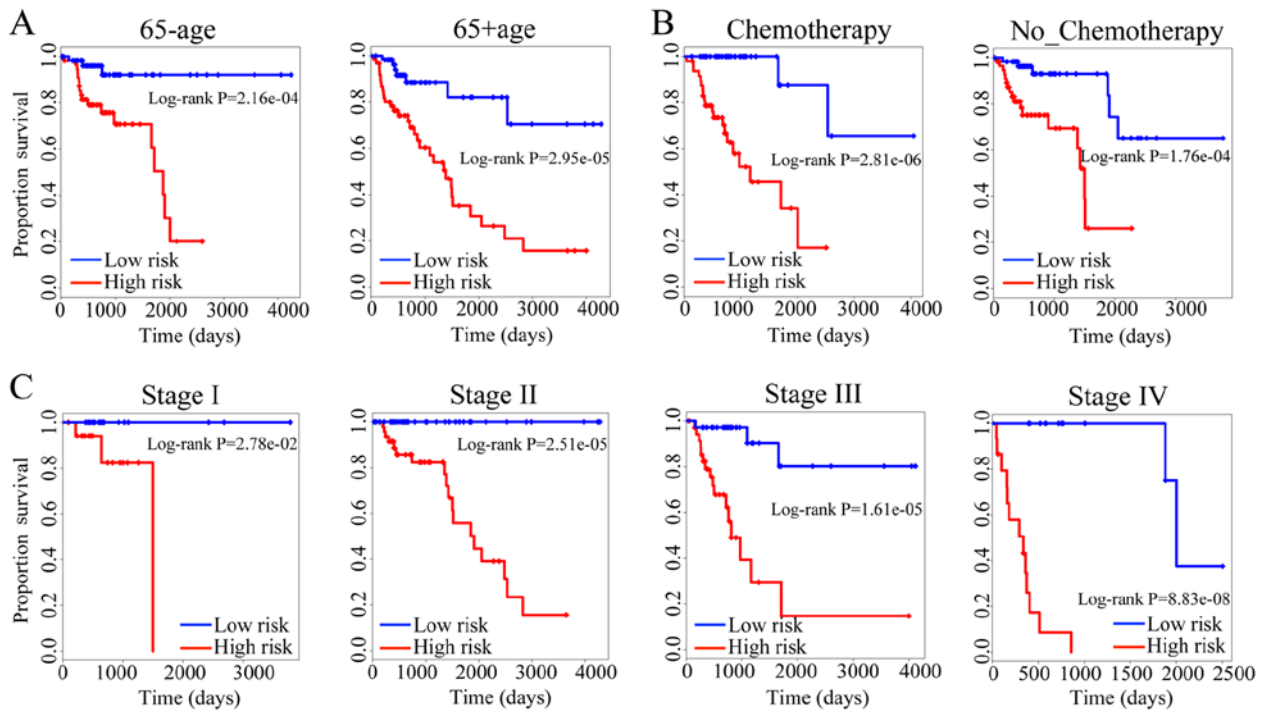


Figure 6. Stratification analyses of all patients with available clinical information. (A) Kaplan-Meier survival curves for all patients with age information. (B) Kaplan-Meier survival curves for all patients with chemotherapy information. (C) Kaplan-Meier survival curves for all patients with tumor stage information. P-values were calculated using the log-rank test.

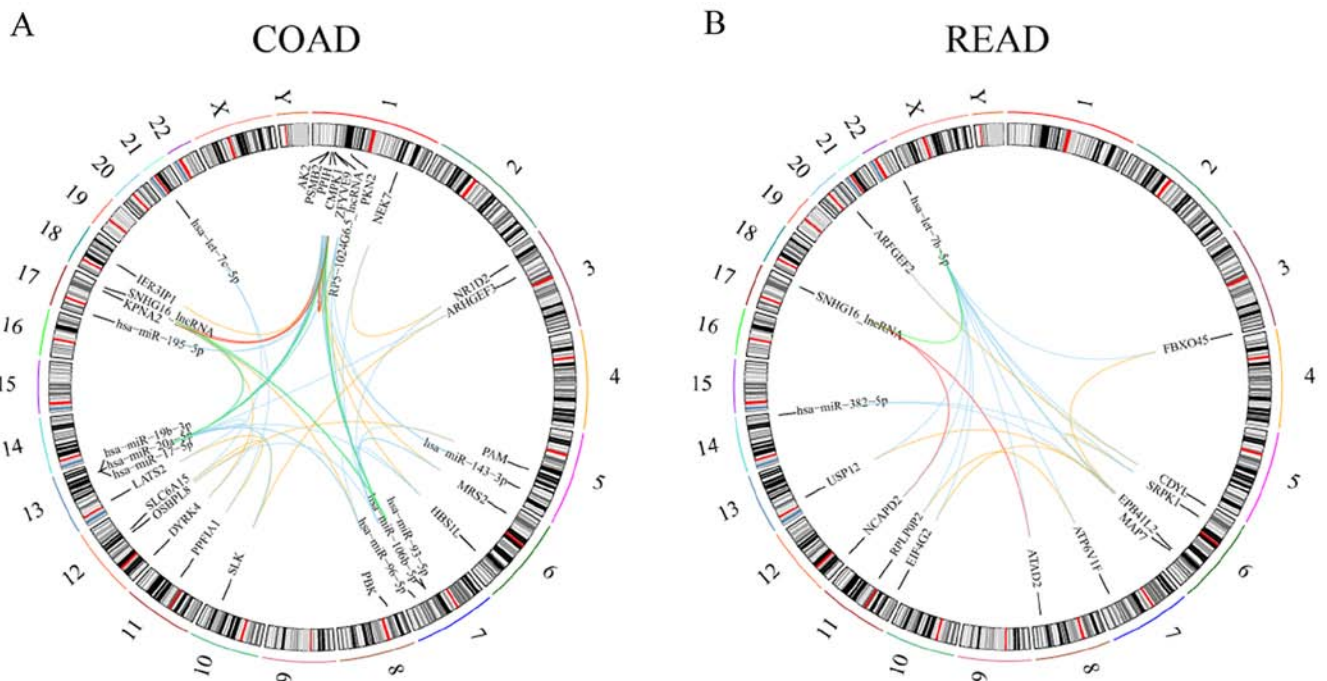


Figure 7. Layouts of chromosomal location of ceRNAs and miRNA regulators in modules. (A) Chromosome locations of genes in COAD module. (B) Chromosome locations of genes in READ module. The red line represents lncRNA-mRNA, the yellow line indicates mRNA-mRNA, and the green line and the blue line indicate miRNA-mRNA, miRNA-lncRNA, respectively. COAD, colon adenocarcinoma; READ, rectal adenocarcinoma; miRNA, microRNA; lncRNA, long non-coding RNA.

was significantly correlated with the overall survival rates of patients with CRC, independent of clinical factors, and the prognostic efficacy was robust. The results also showed that the P-values were lower in patients without chemotherapy, of older age and with advanced-stage disease (Table V). The

results of the stratification analyses of all patients with age, chemotherapy and stage information using the module signature in COAD are also shown in Fig. 6A-C.

The locations of the genes on the human chromosomes and the interactions between genes in prognostic modules



were visualized using the Rcirco 1.2.0 (16). It was found that has-miR-17-5p, has-miR-20a-5p, has-miR-195-5p, large tumor suppressor kinase 2 (LATS2) and karyopherin subunit  $\alpha 2$  (KPNA2) in the prognostic modules were CRC biomarkers, well confirmed by experiments (24-32). Previous studies have found that lncRNA small nucleolar RNA host gene 16 (SNHG16) is an important regulatory gene in CRC, and this was also present in the modules in the present study. It was not possible to further validate this in the GEO due to the absence of the lncRNA expression profile, however, its prognostic efficacy was verified in the TCGA datasets. The results revealed that SNHG16 may be a potential prognostic biomarker. Furthermore, these six RNAs are located on chromosome 13 or 17, and all three miRNAs belong to the precursor miR-17 family. The other RNAs in the modules, particularly has-miR-19b-3p on chromosome 13, and has-miR-93-5p and has-miR-106b-5p, belong to the precursor miR-17 family and are likely to be novel potential CRC biomarkers. These findings require further investigation and experimental confirmation.

## Discussion

CRC is the second most frequently diagnosed type of cancer in Europe, representing 13.2 and 12.7% of all cancer cases in men and women, respectively (33), and it is the fourth common cancer in China (34). Studies have shown that autophagy is closely associated with the biological processes of cancer, including CRC (35-39). Therefore, the development of autophagy-related biomarkers is necessary. In addition, miRNA-mediated ceRNA-ceRNA interaction networks provide a rapid and efficient method of investigating the underlying mechanisms of action of lncRNAs in cancer (40). Differing from previous network analyses of genes, the present study identified network-based survival-associated modules in CRC and further validated their prognostic abilities in independent internal and external datasets based on autophagy-mediated ceRNA-ceRNA interaction networks. In order to verify the ability of the prognostic modules in guiding the treatment of CRC, the overall survival analysis of the modules were performed in terms of age, stage, and presence of chemotherapy.

Considering the crucial role of cellular autophagy in the mechanism of cancer, the present study constructed autophagy-mediated ceRNA-ceRNA interaction networks of CRC. The prognostic triplets were identified, and the multivariate Cox proportional hazards model was then used to examine the association between the expression level of prognostic triplets and survival rate. The patients with CRC were divided into high- and low-risk groups, and the clinically significant prognostic triplets were identified, which were considered as the candidate risk triplets. These triplets have the tendency of collective cooperation in the network, and the largest component of the preserved prognostic triplets was considered as a candidate module; 24-gene and 15-gene modules were identified for COAD and READ, respectively. The predictive ability of the modules was then further confirmed in internal and external independent datasets using the multivariate Cox proportional hazards regression model. Gene modules facilitate elucidation of the underlying cellular infrastructure and

increase understanding of the pathogenesis of CRC at a more comprehensive level of transcriptome. Finally, overall survival analysis of the high- and low-risk groups was performed in terms of age, stage, and chemotherapy, respectively, and it was found that the prognostic power of the two modules was robust.

Following the construction of the autophagy-mediated ceRNA-ceRNA interaction networks, prognostic triplets with significant differences in survival rate were identified, and the association among the triplets was considered, which found that the triplets tended to show collective cooperation. Furthermore, the prognostic modules were identified. Due to the lack of existing lncRNA expression data, the lncRNA expression matrix was obtained by recalculating the RPKM values in TCGA. Although the independent dataset of mRNA and miRNA for validation was found, the lncRNA expression profile was absent. Therefore, validation of the overall survival analysis in the independent dataset was only available for the triplets without lncRNAs. However, the final results showed that the prognostic ability of the selected modules was stable.

The lncRNA SNHG16, existing in the two prognostic-related modules, has been shown to be carcinogenic and upregulated in CRC through dysregulation of the Wnt signaling pathway and affects genes involved in lipid metabolism (24). In addition, hsa-miR-17-5p (25), hsa-miR-20a-5p (26), hsa-miR-195-5p (27), KPNA2 (30-32), and LATS2 (28,29) have been confirmed to be important in CRC. For example, hsa-miR-17-5p promotes chemotherapeutic drug resistance and tumor metastasis of CRC by repressing the expression of phosphatase and tensin homolog (25). Multiple miRNAs in CRC promote tumor growth and metastasis by targeting LATS2 (28,29). The results of Rcirco plots showed that the miRNAs and ceRNA tend to be involved in trans, as shown in Fig. 7A and B. In addition, it was found that certain genes in the prognostic modules may serve as novel potential biomarkers for the diagnosis and treatment of CRC, although the conclusion requires further investigation and experimental confirmation.

In conclusion, the present study integrated the confirmed human autophagy-related genes and the experimental interactions between miRNAs and lncRNAs/mRNAs, combined TCGA CRC gene expression data to construct autophagy-mediated ceRNA-ceRNA interaction networks. Two prognostic modules were identified and subsequent survival analysis showed that the modules were significantly associated with the overall survival rates of patients with CRC, and the prognostic abilities were further confirmed in internal and external independent datasets. The stable prognosis-related modules in patients with CRC are important in clinical guidelines. The prognostic modules from the autophagy-mediated ceRNA-ceRNA interaction networks confirmed that autophagy is crucial in CRC, and that ceRNA is also important in understanding the clinical evaluation of CRC. The findings of the present study may improve current understanding of tumor development and provide more accurate information for the development of novel targeted therapies.

## Acknowledgements

Not applicable.

## Funding

This study was supported by the National Natural Science Foundation of China (grant nos. 31501075 and 31301094), the Natural Science Foundation of Heilongjiang Province (grant no. B201302), the Education Department Foundation of Heilongjiang Province (grant no. 12531227), the Health Department Foundation of Heilongjiang Province (grant no. 2012-798) and the Scientific Research Project of Heilongjiang Provincial Education Department (grant no. 12541565).

## Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

## Authors' contributions

KL and LW conceived and designed the experiments. KQ, HH, JJ, DX, SG, YC and HW analyzed the data. KL and LW wrote the manuscript. All authors have read and approved the final manuscript.

## 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

- Qiang L, Sample A, Shea CR, Soltani K, Macleod KF and He YY: Autophagy gene ATG7 regulates ultraviolet radiation-induced inflammation and skin tumorigenesis. *Autophagy* 13: 2086-2103, 2017.
- Aita VM, Liang XH, Murty VV, Pincus DL, Yu W, Cayanis E, Kalachikov S, Gilliam TC and Levine B: Cloning and genomic organization of beclin 1, a candidate tumor suppressor gene on chromosome 17q21. *Genomics* 59: 59-65, 1999.
- Wu TY, Cho TY, Lu CK, Liou JP and Chen MC: Identification of 7-(4'-Cyanophenyl)indoline-1-benzenesulfonamide as a mitotic inhibitor to induce apoptotic cell death and inhibit autophagy in human colorectal cancer cells. *Sci Rep* 7: 12406, 2017.
- Katheder NS and Rusten TE: Microenvironment and tumors-a nurturing relationship. *Autophagy* 13: 1241-1243, 2017.
- Galluzzi L, Pietrocola F, Levine B and Kroemer G: Metabolic control of autophagy. *Cell* 159: 1263-1276, 2014.
- Mizushima N, Yoshimori T and Ohsumi Y: The role of Atg proteins in autophagosome formation. *Annu Rev Cell Dev Biol* 27: 107-132, 2011.
- Cancer Genome Atlas Research Network: Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455: 1061-1068, 2008.
- Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, *et al*: GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res* 22: 1760-1774, 2012.
- Wang P, Ning S, Zhang Y, Li R, Ye J, Zhao Z, Zhi H, Wang T, Guo Z and Li X: Identification of lncRNA-associated competing triplets reveals global patterns and prognostic markers for cancer. *Nucleic Acids Res* 43: 3478-3489, 2015.
- Chen DT, Hernandez JM, Shibata D, McCarthy SM, Humphries LA, Clark W, Elahi A, Gruidl M, Coppola D and Yeatman T: Complementary strand microRNAs mediate acquisition of metastatic potential in colonic adenocarcinoma. *J Gastrointest Surg* 16: 905-912, discussion 912-913, 2012.
- Li JH, Liu S, Zhou H, Qu LH and Yang JH: starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* 42D: D92-D97, 2014.
- Arntzen MO, Boddie P, Frick R, Koehler CJ and Thiede B: Consolidation of proteomics data in the Cancer Proteomics database. *Proteomics* 15: 3765-3771, 2015.
- Wu D, Huang Y, Kang J, Li K, Bi X, Zhang T, Jin N, Hu Y, Tan P, Zhang L, *et al*: ncRDeathDB: A comprehensive bioinformatics resource for deciphering network organization of the ncRNA-mediated cell death system. *Autophagy* 11: 1917-1926, 2015.
- Xu J and Li YH: miRDeathDB: A database bridging microRNAs and the programmed cell death. *Cell Death Differ* 19: 1571, 2012.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T: Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498-2504, 2003.
- Zhang H, Meltzer P and Davis S: RCircos: An R package for Circos 2D track plots. *BMC Bioinformatics* 14: 244, 2013.
- Sun J, Cheng L, Shi H, Zhang Z, Zhao H, Wang Z and Zhou M: A potential panel of six-long non-coding RNA signature to improve survival prediction of diffuse large-B-cell lymphoma. *Sci Rep* 6: 27842, 2016.
- Li C, Zhu B, Chen J and Huang X: Novel prognostic genes of diffuse large B-cell lymphoma revealed by survival analysis of gene expression data. *Oncotargets Ther* 8: 3407-3413, 2015.
- Li L: Survival prediction of diffuse large-B-cell lymphoma based on both clinical and gene expression information. *Bioinformatics* 22: 466-471, 2006.
- Zhao S, Bai N, Cui J, Xiang R and Li N: Prediction of survival of diffuse large B-cell lymphoma patients via the expression of three inflammatory genes. *Cancer Med* 5: 1950-1961, 2016.
- Alizadeh AA, Gentles AJ, Alencar AJ, Liu CL, Kohrt HE, Houot R, Goldstein MJ, Zhao S, Natkunam Y, Advani RH, *et al*: Prediction of survival in diffuse large B-cell lymphoma based on the expression of 2 genes reflecting tumor and microenvironment. *Blood* 118: 1350-1358, 2011.
- Yang J, Yuan D, Xing T, Su H, Zhang S, Wen J, Bai Q and Dang D: Ginsenoside Rh2 inhibiting HCT116 colon cancer cell proliferation through blocking PDZ-binding kinase/T-LAK cell-originated protein kinase. *J Ginseng Res* 40: 400-408, 2016.
- Blum C, Graham A, Yousefzadeh M, Shrout J, Benjamin K, Krishna M, Hoda R, Hoda R, Cole DJ, Garrett-Mayer E, *et al*: The expression ratio of Map7/B2M is prognostic for survival in patients with stage II colon cancer. *Int J Oncol* 33: 579-584, 2008.
- Christensen LL, True K, Hamilton MP, Nielsen MM, Damas ND, Damgaard CK, Ongen H, Dermitzakis E, Bramsen JB, Pedersen JS, *et al*: SNHG16 is regulated by the Wnt pathway in colorectal cancer and affects genes involved in lipid metabolism. *Mol Oncol* 10: 1266-1282, 2016.
- Fang L, Li H, Wang L, Hu J, Jin T, Wang J and Yang BB: MicroRNA-17-5p promotes chemotherapeutic drug resistance and tumour metastasis of colorectal cancer by repressing PTEN expression. *Oncotarget* 5: 2974-2987, 2014.
- Cheng D, Zhao S, Tang H, Zhang D, Sun H, Yu F, Jiang W, Yue B, Wang J, Zhang M, *et al*: MicroRNA-20a-5p promotes colorectal cancer invasion and metastasis by downregulating Smad4. *Oncotarget* 7: 45199-45213, 2016.
- Sun M, Song H, Wang S, Zhang C, Zheng L, Chen F, Shi D, Chen Y, Yang C, Xiang Z, *et al*: Integrated analysis identifies microRNA-195 as a suppressor of Hippo-YAP pathway in colorectal cancer. *J Hematol Oncol* 10: 79, 2017.
- Zheng YB, Xiao K, Xiao GC, Tong SL, Ding Y, Wang QS, Li SB and Hao ZN: MicroRNA-103 promotes tumor growth and metastasis in colorectal cancer by directly targeting LATS2. *Oncol Lett* 12: 2194-2200, 2016.
- He Y, Wang J, Wang J, Yung VY, Hsu E, Li A, Kang Q, Ma J, Han Q, Jin P, *et al*: MicroRNA-135b regulates apoptosis and chemoresistance in colorectal cancer by targeting large tumor suppressor kinase 2. *Am J Cancer Res* 5: 1382-1395, 2015.
- Zhang Y, Zhang M, Yu F, Lu S, Sun H, Tang H and Peng Z: Karyopherin alpha 2 is a novel prognostic marker and a potential therapeutic target for colon cancer. *J Exp Clin Cancer Res* 34: 145, 2015.

31. Yu L, Wang G, Zhang Q, Gao L, Huang R, Chen Y, Tang Q, Liu J, Liu C, Wang H, *et al*: Karyopherin alpha 2 expression is a novel diagnostic and prognostic factor for colorectal cancer. *Oncol Lett* 13: 1194-1200, 2017.
32. Takada T, Tsutsumi S, Takahashi R, Ohson K, Tatsuki H, Suto T, Kato T, Fujii T, Yokobori T and Kuwano H: KPNA2 over-expression is a potential marker of prognosis and therapeutic sensitivity in colorectal cancer patients. *J Surg Oncol* 113: 213-217, 2016.
33. Stanel SC, Sjöberg J, Salmonson T, Foggi P, Caleno M, Melchiorri D, Gravanis I, Tzogani K and Pignatti F: European Medicines Agency approval summary: Zaltrap for the treatment of patients with oxaliplatin-resistant metastatic colorectal cancer. *ESMO Open* 2: e000190, 2017.
34. Gu M-J, Huang Q-C, Bao C-Z, Li YJ, Li XQ, Ye D, Ye ZH, Chen K and Wang JB: Attributable causes of colorectal cancer in China. *BMC Cancer* 18: 38, 2018.
35. Zhang W, Yuan W, Song J, Wang S and Gu X: LncRNA CPS1-IT1 suppresses EMT and metastasis of colorectal cancer by inhibiting hypoxia-induced autophagy through inactivation of HIF-1 $\alpha$ . *Biochimie* 144: 21-27, 2018.
36. Liu L, Zhao WM, Yang XH, Sun ZQ, Jin HZ, Lei C, Jin B and Wang HJ: Effect of inhibiting Beclin-1 expression on autophagy, proliferation and apoptosis in colorectal cancer. *Oncol Lett* 14: 4319-4324, 2017.
37. Wang S, Wang K, Zhang C, Zhang W, Xu Q, Wang Y, Zhang Y, Li Y, Zhang Y, Zhu H, *et al*: Overaccumulation of p53-mediated autophagy protects against betulinic acid-induced apoptotic cell death in colorectal cancer cells. *Cell Death Dis* 8: e3087, 2017.
38. Qian HR, Shi ZQ, Zhu HP, Gu LH, Wang XF and Yang Y: Interplay between apoptosis and autophagy in colorectal cancer. *Oncotarget* 8: 62759-62768, 2017.
39. Zhao W, Shi F, Guo Z, Zhao J, Song X and Yang H: Metabolite of ellagitannins, urolithin A induces autophagy and inhibits metastasis in human sw620 colorectal cancer cells. *Mol Carcinog* 57: 193-200, 2018.
40. Conte F, Fisco G, Chiara M, Colombo T, Farina L and Paci P: Role of the long non-coding RNA PVT1 in the dysregulation of the ceRNA-ceRNA network in human breast cancer. *PLoS One* 12: e0171661, 2017.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.