Digestive Diseases and Sciences
https://doi.org/10.1007/s10620-020-06064-1

REVIEW

Identification of MFI2-AS1, a Novel Pivotal lncRNA for Prognosis of Stage III/IV Colorectal Cancer

Ruihan Luo[1] · Jing Song[1] · Wanfeng Zhang[1] · Longke Ran[1]

Received: 17 July 2019 / Accepted: 9 January 2020

Abstract

Background Long noncoding RNAs (lncRNAs) have been shown to play pivotal role in pathogenesis and prognosis of cancers. Identification of novel clinical biomarkers in advanced stage colorectal cancer (CRC) is warranted.

Aims To identify potential lncRNAs associated with progression of stage III/IV CRC and illuminate regulatory mechanisms.

Methods Differentially expressed lncRNAs, mRNAs and miRNAs (DElncRNAs, DEmRNAs, and DEmiRNAs) were extracted between stage III/IV CRC and normal tissues. We used DEGs to construct a ceRNA network and analyzed correlations between key lncRNAs and overall survivals (OS) of stage III/IV CRC patients. Weighted gene co-expression network analysis (WGCNA) was applied to a pivotal lncRNA. We conducted functional enrichment analysis on target genes and constructed lncRNA–TF–mRNA network by overlapping mRNAs co-expressed with the key lncRNA and target genes of transcriptional factors (TFs).

Results A total of 26 DElncRNAs, 398 DEmiRNAs, 2155 DEmRNAs were identified. A ceRNA network was constructed with 16 lncRNAs, 20 miRNAs, and 59 mRNAs, in which MFI2-AS1 exhibited promising diagnostic efficiency. (AUC was 0.938.) MFI2-AS1 was negatively correlated to OS of stage III/IV CRC patients (P value < 0.05). KEGG analysis showed potential mRNA targets of MFI2-AS1 mainly involved in cell cycle and cytokine–cytokine receptor interaction. We identified 17 potential TFs of MFI2-AS1 and built a lncRNA–TF–mRNA network.

Conclusion Our study provides novel insights into lncRNAs associated regulatory networks and reveals a promising lncRNA biomarker, MFI2-AS1, as an independent prognostic indicator and potential therapeutic target for CRC.

Keywords Colorectal cancer · LncRNAs · CeRNA network · Co-expression analysis · Prognosis

Introduction

Colorectal cancer (CRC) is the third most common malignancies with the fourth highest mortality rate in the world for several decades [1]. For the majority of early stage (stage I and II) CRC patients, partial or total colectomy alone is considered as the primary treatment option; while late-stage (stage III and IV) CRCs are often treated with neoadjuvant chemotherapy plus radiation and chemotherapy [2]. Available therapeutical approaches and effective prognostic biomarkers are very limited for metastatic CRC, frequently contributing to inappropriate clinical decision making. In spite of substantial advancements in medical science, surgical techniques, and chemotherapy of CRC treatment, tumor recurrence, metastasis and chemotherapy resistance are still obstacles to clinical management and improved survivals in advanced stage CRC patients, suggesting existing therapeutics are insufficient for patients with locally advanced or distantly metastatic CRC. Long noncoding RNAs, defined as a type of noncoding RNA molecules of more than 200 nucleotides, can regulate gene expression in different levels (transcriptional, post-transcriptional, translational, post-translational, and epigenetic) [3]. CRC has recently been reported to be one of the cancers showing the most associated deregulation of lncRNAs' functions that are modulating

✉ Longke Ran
ranlongke@cqmu.edu.cn

Ruihan Luo
2766969847@qq.com

Jing Song
707586478@qq.com

Wanfeng Zhang
1045014659@qq.com

[1] Department of Bioinformatics, Chongqing Medical University, Chongqing, China

Published online: 20 January 2020

a bunch of pathophysiological processes, including chromatin remodeling, histone modifications, cell cycle, splicing and apoptosis [3]. For instance, downregulation of LET and UC.388 and the upregulation of HOTAIR, MALAT1, and CCATA2 were found to promote metastasis of CRC [4]. High expression levels of HOTAIR, HULC, and LINC00152 have predicted a poor prognosis of CRC [5, 6]. To be sure, despite over 2 decades of vast investigation in small ncRNAs, in particular miRNAs whose biological implications have been unraveled, still very little is elucidated for exact molecular biological mechanisms of lncRNAs.

Based on the competing endogenous RNA (ceRNA) hypothesis whereby lncRNAs can bind and seclude miRNAs via miRNA-binding sites [7], acting as ceRNAs that interact with mRNAs and functioning as miRNA sponges, to restrict miRNAs from repressing expression level of target genes [8, 9], the dysregulated lncRNA-mediated ceRNAs have been identified to have tumorigenic or cancer-promoting effect on various cancers. For example, LINC00460 regulates the expression of stanniocalcin 2 that is correlated with a poor prognosis in head and neck squamous cell carcinoma via sponging miR-206 [10]. In this context, it is imperative that more comprehensive function analysis of lncRNAs be performed to better clarify the molecular mechanism of CRC. Therefore, with the use of bioinformatics, we firstly contributed to investigate the differentially expressed profiles of lncRNAs, mRNAs, and miRNAs of stage III/IV CRC samples in datasets from TCGA and GEO, respectively, and then constructed a ceRNA network to predict their potential interactions. Next, we validated the significant differentially expressions of candidate lncRNAs from this network in another GEO dataset. Moreover, to illuminate the lncRNA involved in the ceRNA network and also significantly correlated to OS, we sought to perform Gene Ontology and pathway analysis on predicted target genes via WGCNA, and built a lncRNA–TF–mRNA network. Finally, our work revealed a novel lncRNA associated with overall survivals, which could be considered as an independent prognostic biomarker for advanced stage CRC. We suggest that our findings might help clinicians to better understand the function of lncRNAs, explore new physiological and pathological mechanisms of CRC, and provide innovative as well as individualized therapeutics for CRC patients in advanced stages.

Materials and Methods

CRC Datasets

A total of 453 patients from The Cancer Genome Atlas (TCGA) dataset were included in this study, on which information included the TCGA-COAD miRNA expression profiles (454 tumor tissues, 8 adjacent normal samples), RNA expression profiles (477 tumor tissues, 41 adjacent normal samples) and corresponding clinical data. For another, two microarray expression datasets (accession numbers GSE89076 and GSE41657) based on GPL16699 and GPL6480 platforms were downloaded from the Gene Expression Omnibus (GEO) for more reliable analysis; consequently, 75 (36 tumor, 39 adjacent normal) and 37 (25 tumor, 12 adjacent normal) CRC tissue samples were extracted as training and testing set, respectively.

Preprocessing and Analysis of Expression Profiles

According to the eighth revision of AJCC TNM classification, the advanced stage (stage III/IV) tumor samples from 3 CRC datasets were extracted based on each of their pathologic stage information. Ultimately, the number of stage III/IV tumor tissues for datasets above were 19 (GSE89076), 12 (GSE41657), 199 and 194 (TCGA-COAD RNA-seq and miRNA-seq data), respectively.

Then, microarray expression profiles analysis (GSE89076 data were preprocessed through log2-standardization and quantile normalization) was performed with R package limma, and edgeR package was used to analyze RNA and microRNA sequencing data. The significantly differentially expressed genes (DEGs) between stage III/IV tumor and adjacent normal tissue samples as well as between all tumor and adjacent normal samples of datasets above with $|\log2 \text{ fold change}| \geq 1.0$ and P value or false discovery rate (FDR) < 0.05 were retained for further analysis. The differentially expressed lncRNAs (DElncRNAs) and mRNAs (DEmRNAs) were subsequently screened from the DEGs based on the GENCODE (V22) annotation. We named intersections of DElncRNAs and DEmRNAs from four comparative groups of two datasets (GSE89076 and TCGA-COAD), and of DEmiRNAs from two comparative groups (TCGA-COAD only) as intersecting DElncRNAs, DEmRNAs, and DEmiRNAs, respectively, which were also selected for subsequent analysis. The workflow of this selection process is shown in Fig. 1.

GO and Pathway Enrichment Analysis

To construct meaningful annotation of DEGs and better analyze the biological functions of DEGs as well as pathways of the aberrantly expression, the Gene Ontology (GO), including biological process (BP), cellular component (CC) and molecular function (MF) as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were carried out by R package of clusterprofiler with a cut-off of P value < 0.05.

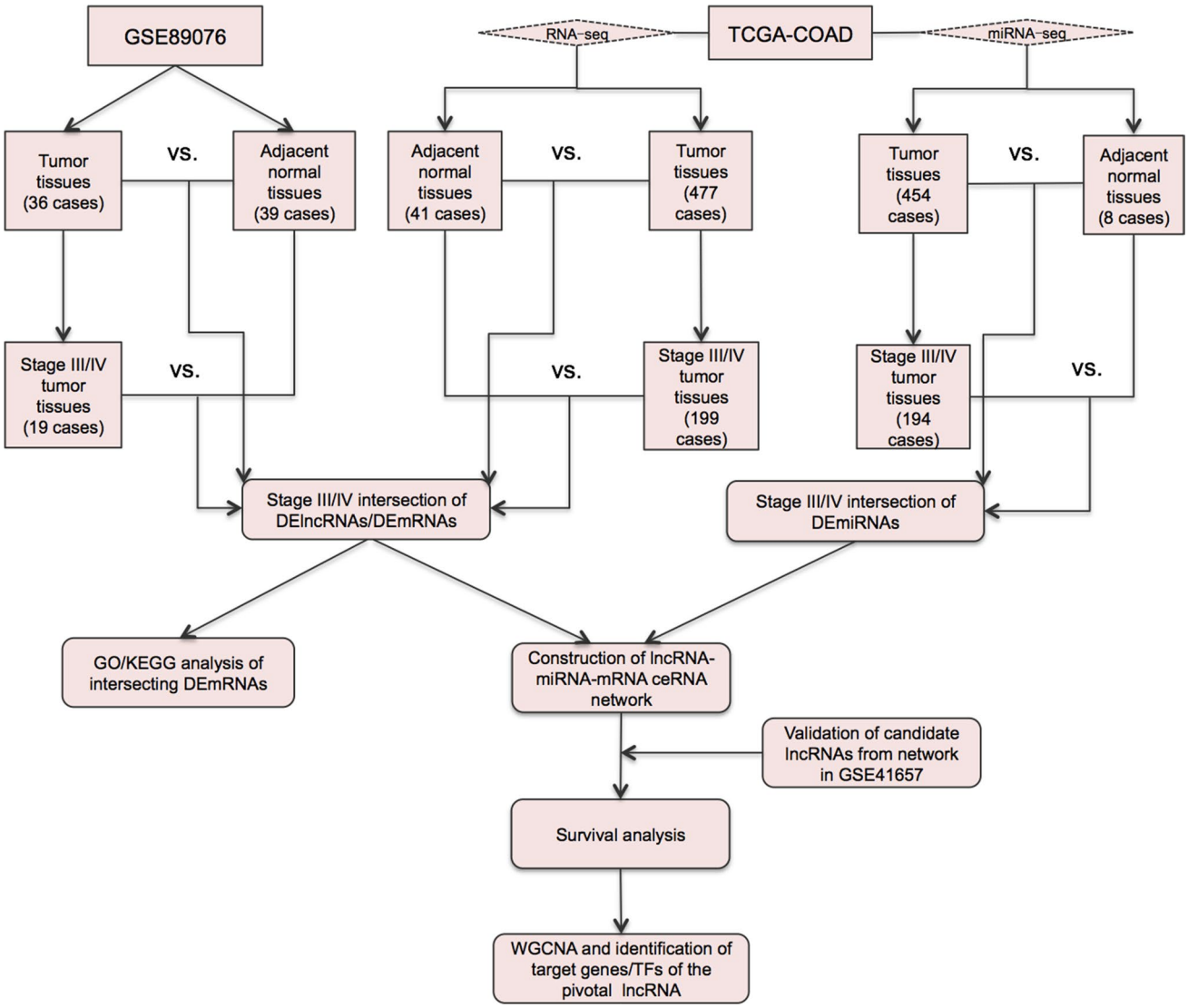

Fig. 1 Flowchart of data generation and analysis

Establishing the ceRNA Network in CRC and Validating Key lncRNAs with GEO Expression Profile

That lncRNAs can affect mRNAs through miRNAs is a well-known theory according to previous studies, on which the lncRNA-associated ceRNA network could be constructed. We used intersecting DElncRNAs, DEmiRNAs, and DEmRNAs mentioned above for establishing this kind of network in late-stage CRC. The specific steps are as follows: Firstly, highly reliable online miRNA reference database of miRcode (http://www.mircode.org/) was applied to search for lncRNA-miRNA interactions based on intersecting DEmiRNAs. Then, miRNA-targeted mRNAs were retrieved from Targetscan v7.1 (http://www.targetscan.org/vert_71/), miRDB (http://www.mirdb.org/miRDB/) and DIANA-Tarbase v7.0 (http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=site/index). We retained the intersections of overlapping predictive results and the aforementioned DElncRNAs and DEmRNAs, respectively. Additionally, Pearson correlation analysis between intersecting DEmiRNAs and corresponding mRNAs enabled us to remain potential target genes. Cytoscape v3.7.1 was used to visualize the lncRNA-miRNA-mRNA ceRNA network. LncRNAs shown in the network were considered as potential candidate lncRNAs in this study. The candidate lncRNAs were further validated in testing set, GSE41657. Receiver operating characteristic (ROC) curve analysis was performed to identify more reliable lncRNAs.

Survival Analysis of Candidate lncRNAs

Key lncRNAs validated by GSE41657 were extracted to investigate their potential of prognostic prediction for advanced stage CRC patients in combination with demographic and clinical characteristics including sex, history of colon polyps, lymphatic invasion, venous invasion, preoperative CEA level, lymphatic metastasis and distant metastasis. Kaplan–Meier method and Cox's proportional hazards (PH) regression model were performed to evaluate overall survivals by conducting the survival and survminer packages. And the final lncRNA correlated with overall survivals was thus regarded as the pivotal one.

Constructing WGCNA and Predicting Target mRNAs of the Pivotal lncRNA

To describe gene expression profiles from TCGA-COAD and their correlations, the WGCNA R package was applied to assess the potential genes associated with the pivotal lncRNA and their module membership by setting soft-thresholding power of 7 (scale-free $R^2 = 0.86$) and a module size cut-off of 30. A topological overlap measurement (TOM) was generated by the WGCNA algorithm, which also provided a generalized evaluation of the edges between two gene nodes. To present the interactions between the key lncRNA and mRNAs as well as their weighted relevance, a co-expressed network was built based on co-expression module of stage III/IV CRC samples and visualized with the use of Cytoscape 3.7.1. The prediction and biological function enrichment analysis for potential target mRNAs of the pivotal lncRNA was also implemented via the co-expression network.

Identification of Trans-regulation of the lncRNA

To further investigate the trans-regulatory function of lncRNAs by the transcriptional factors (TFs) in CRC tumor tissues, predicted TFs that potentially involved in trans-regulation of the pivotal lncRNA were downloaded from three major databases, including GeneCards, JASPAR, and Promo. All potential TFs were identified from Ensembl (http://asia.ensembl.org/index.html). And the lncRNA–TF–mRNA network was built by overlapping target mRNAs of the lncRNA above with the target genes of predicted TFs. The significant correlations of expressions between lncRNA and TFs in stage III/IV CRC were determined in TCGA dataset.

Statistical Analysis

R version 3.5.0 was utilized for all statistical analyses with statistical significance being set at P value < 0.05. A multiple-comparison correction was performed for Pearson correlation analyses using Bonferroni correction. The ROC curve analysis was conducted via R package ROCR. To identify lncRNAs differentially expressed between stage III/IV CRC and normal tissue samples in validation set, we used the Wilcoxon rank test followed by multiple testing using the Benjamini–Hochberg. *, P value < 0.05, P value < 0.01, ***, P value < 0.001, ****, P value < 0.0001.

Results

DEmRNA, DElncRNA, and DEmiRNA in CRC Patients

A total of 4799 mRNAs and 470 lncRNAs were identified as being aberrantly expressed between all tumor and adjacent normal tissues from TCGA-COAD. To investigate targeted biomarkers for CRC patients in advanced stages, further analysis was conducted between stage III/IV tumor tissues and adjacent normal samples, which identified 4774 DEmRNAs and 457 DElncRNAs, among which 2704 mRNAs and 386 lncRNAs were upregulated, while 2070 mRNAs and 71 lncRNAs were downregulated (Fig. 2a). To enhance the credibility of observations, a similar analysis for GEO dataset (GSE89076) was carried out, identifying 3265 DEmRNAs, 57 DElncRNAs (integrated tumor tissues vs. normal samples), 3323 DEmRNAs and 57 DElncRNAs (stage III/IV CRC vs. normal tissues), respectively. Eventually, more reliable 2155 DEmRNAs and 26 DElncRNAs in intersections of the four comparative groups derived from 2 datasets were retained for further analysis (Fig. 2b, c). To construct an lncRNA-miRNA-mRNA ceRNA network, a total of 427 and 466 DEmiRNAs were found to be differentially expressed between total tumor tissues, stage III/IV tumor tissues and adjacent normal tissues. Similarly, we obtained 398 intersecting DEmiRNAs (Fig. 2d).

Go and KEGG Enrichment Analysis of DEmRNAs

To investigate the function of DEGs, intersecting DEmRNAs were selected for functional enrichment analysis, which might also reflect functional implications of relevant lncRNAs. The results of GO BP terms (Top 15, P value < 0.05), MF terms (Top 10, P value < 0.05), CC terms (Top 15, P value < 0.05) and KEGG enrichment analysis (Top 20, P value < 0.05) were presented in Fig. 3. The GO analytical results showed that compared with adjacent normal tissues, DEmRNAs in stage III/IV CRC with respect to biological processes were organelle fission, chromatid segregation and DNA replication (Fig. 3a). KEGG pathway enrichment analysis demonstrated that intersecting DEmRNAs were mainly enriched in several pathways, including cytokine–cytokine

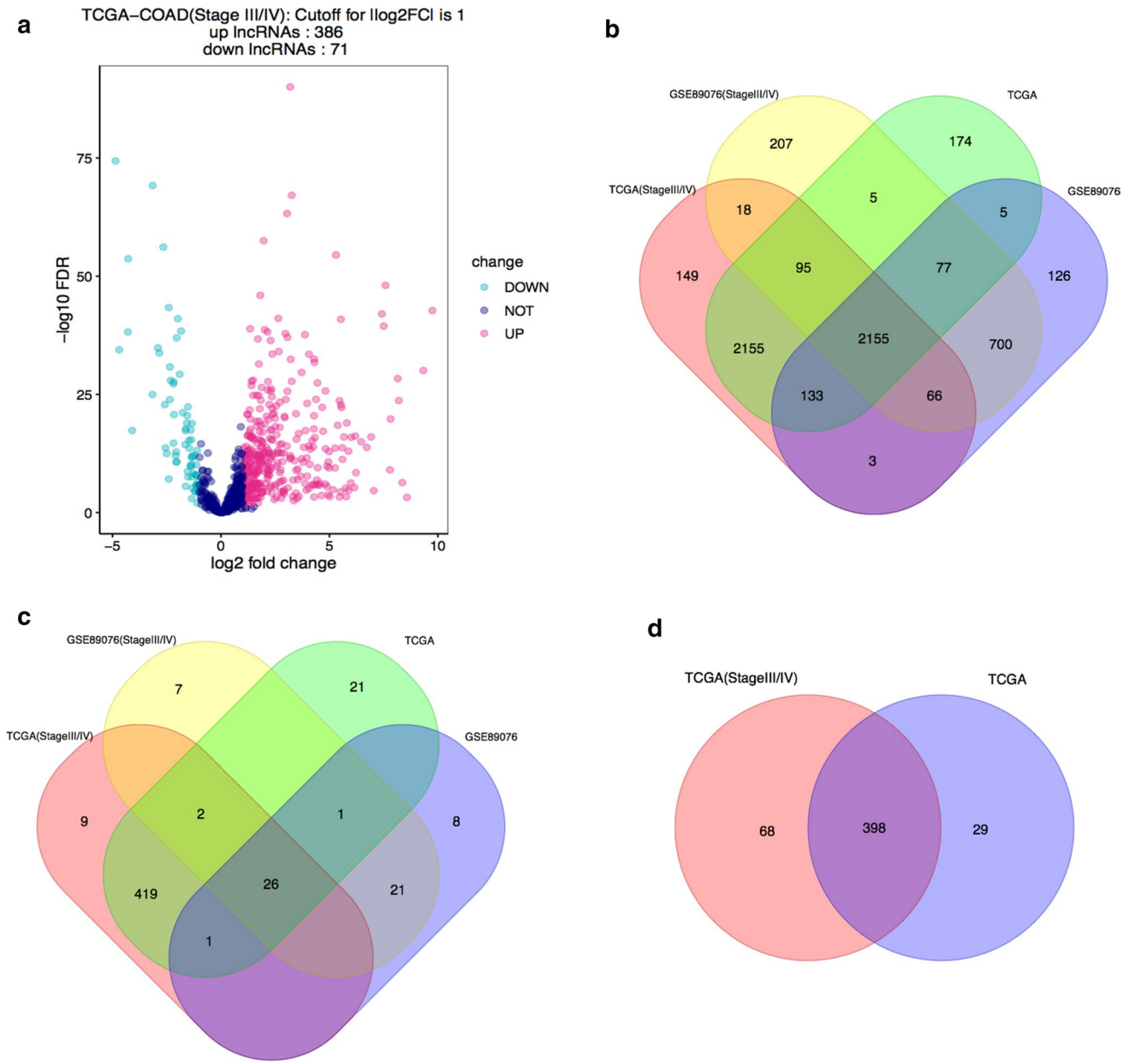

Fig. 2 **a** The volcano plot demonstrates the lncRNA expression profiles in stage III/IV TCGA-COAD data. Venn diagram analysis of differentially expressed **b** mRNAs, **c** lncRNAs and **d** miRNAs between all CRCs, stage III/IV tumors and non-tumor samples of TCGA and GSE89076 datasets, respectively

receptor interaction, cell cycle and ECM-receptor interaction (Fig. 3b).

Construction and Analysis of ceRNA Network

To determine whether 398 intersecting DEmiRNAs target 26 DElncRNAs shown in the venn diagrams above, according to miRcode, 36 miRNAs were screened to predict 17 unique lncRNAs. Targetscan, miRdb, and DIANA-Tarbase measured 455 miRNA-targeted mRNAs that were overlapped with intersecting DEmRNAs. We subsequently analyzed the correlation between 455 predicted mRNAs and 36 screened miRNAs with adj. $P < 0.05$ and $r > 0.4$, and then, 59 more targeted mRNAs interacted with 20 miRNAs were harvested. Ultimately, a total of 16 lncRNAs, 20miRNAs, and 59 mRNAs (specific information is shown in Table 1) were involved in constructing ceRNA network (Fig. 4) using Cytoscape 3.7.1.

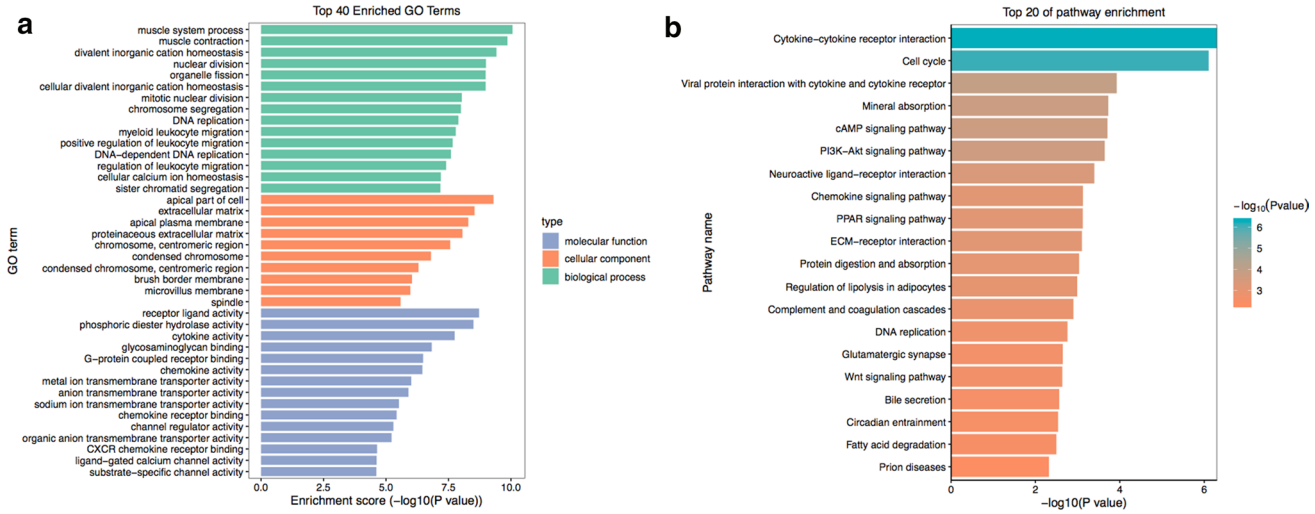

Fig. 3 The GO and KEGG enrichment analysis results of intersecting DEmRNAs. **a** Top 15 BP, top 10 MF, top 15 CC terms. **b** Top 20 KEGG pathways

Table 1 Specific interactions between lncRNAs, miRNAs and mRNAs

miRNA	lncRNA	mRNA
hsa-mir-101	SNHG1, LINC00483	EZH2, ARNTL2, ADAMTSL3, GRIN2A, NOVA1, GFRA1
hsa-mir-103a	UCA1, LINC00114, MFI2-AS1, SNHG7, MLK7-AS1	IARS, EBF1
hsa-mir-125a	MFI2-AS1, CYP1B1-AS1, GNAS-AS1	FANCI, JAM2, PRKCB
hsa-mir-129	UCA1, SNHG1, LINC00483	JPH4, PDRG1, SCN3B
hsa-mir-132	GAS5, MLK7-AS1	KCNK2, NOVA1, FOXN3, EDIL3
hsa-mir-142	MLK7-AS1	C20orf194
hsa-mir-143	UCA1, SFTA1P, SNHG1, LINC00114, MRPL23-AS1, SNHG7, MLK7-AS1, PVT1	KIAA2022
hsa-mir-144	SNHG1, LINC00483, GAS5, MLK7-AS1	ANO1
hsa-mir-145	SNHG1, MRPL23-AS1, MFI2-AS1, CYP1B1-AS1, MLK7-AS1, PVT1	AHCYL2
hsa-mir-155	LINC00240, GAS5	SLITRK3
hsa-mir-17	LINC00483, PVT1	DCLK1, EBF1, DPT, ANGPTL1
hsa-mir-194	SNHG1, PVT1	HECW2, TNFAIP6
hsa-mir-218	LINC00483, CYP1B1-AS1, GNAS-AS1, MLK7-AS1	ITM2C
hsa-mir-22	LINC00114, MRPL23-AS1, SNHG7	CPEB1, OGN, MEIS2
hsa-mir-24	LINC00240, LINC00471, MRPL23-AS1, CYP1B1-AS1, SNHG7, GAS5, GNAS-AS1, MLK7-AS1, PVT1	SCML1
hsa-mir-338	ABHD11-AS1, CYP1B1-AS1, SNHG7, GNAS-AS1	CHL1, RAB30, AKAP12, C10orf54
hsa-mir-33a	LINC00114, MRPL23-AS1, CYP1B1-AS1, GAS5, PVT1	SLC4A4
hsa-mir-7	ABHD11-AS1, SNHG1, LINC00483, PVT1	EDIL3, MIER3, SEMA6D, PHF19, RPGRIP1L, KIAA1549, EIF5A2, CDC14A, SLC41A2, MIER1, GPD1L, SEMA6A, MSX2
hsa-mir-9	SFTA1P, SNHG1, LINC00483, LINC00471, LINC00114, MFI2-AS1, SNHG7, PVT1	STC1, SOS2, SLC26A3, CGN, OSBPL3, MORC4, PDCD4
hsa-mir-96	UCA1, LINC00483, LINC00114, GAS5	CLIP4, PINK1, PTN

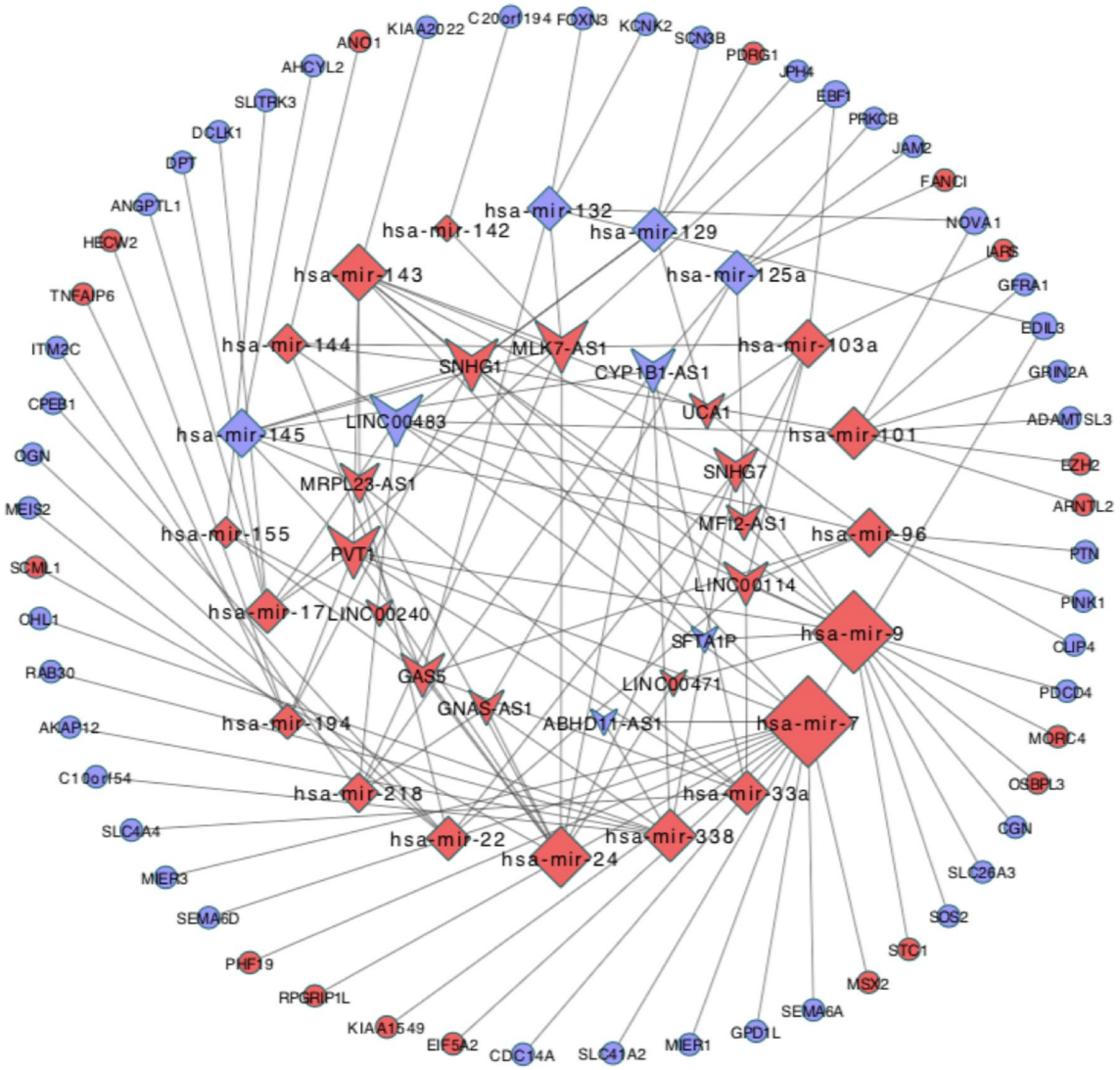

Fig. 4 The ceRNA network in stage III/IV CRC. Red nodes represent upregulated genes, while blue nodes represent downregulated genes. Balls represent mRNAs; diamonds represent miRNAs; Vs represent lncRNAs; gray edges indicate lncRNA–miRNA–mRNA interactions

Validation of Candidate lncRNAs in GSE41657

The expression profiles of candidate lncRNAs shown in the ceRNA network were further validated with GEO: GSE41657 (12 stage III/IV CRC and 12 adjacent normal samples). A total of 11 lncRNAs were detected among 16 lncRNAs, and the differential expressions of 7 lncRNAs were statistically significant between the stage III/IV tumor and normal tissues, with SFTA1P downregulated, PVT1, SNHG7, MFI2-AS1, LINC00114, SNHG1, and UCA1 consistently upregulated (Fig. 5a). Superior sensitivity and specificity on advanced stage CRC diagnosis as estimated by ROC curve analysis was provided by each of significantly DElncRNAs (Fig. 5b), indicating their reliabilities of differentiating stage III/IV CRC patients from the normal.

The Pivotal lncRNA Presented Clinical Prognostic Significance for Stage III/IV CRC Patients

We then investigated the performance of 7 aforementioned lncRNAs in prediction of prognosis for advanced stage CRC of TCGA dataset. Kaplan–Meier curve analysis was performed to study the overall survivals of stage III/IV CRC patients using expressions of 7 lncRNAs, of which results demonstrated that LINC00114 was positively correlated with OS (Fig. 6a), whereas MFI2-AS1 negatively related to OS (log rank P value < 0.05) (Fig. 6b). Next, multivariate Cox's PH regression model analysis was conducted to explore the association between two lncRNAs and OS in combination with clinical characteristics. MFI2-AS1 was found as an independent prognostic indicator of stage III/IV CRCs. High MFI2-AS1 level was significantly associated

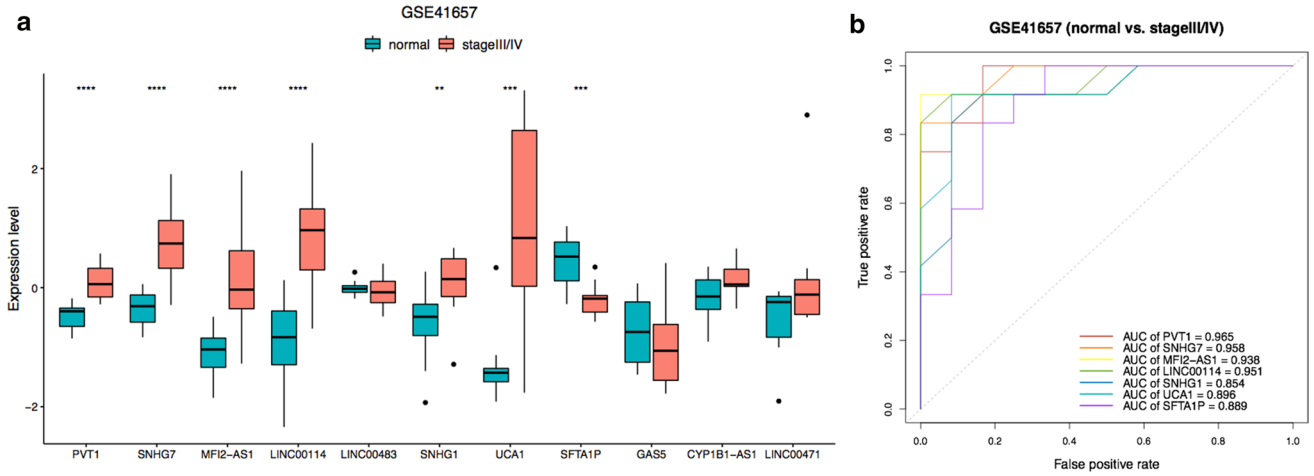

Fig. 5 **a** The Boxplot of validation for expressions of selected lncRNAs in GSE41657. **b** ROC curves of 7 candidate lncRNAs

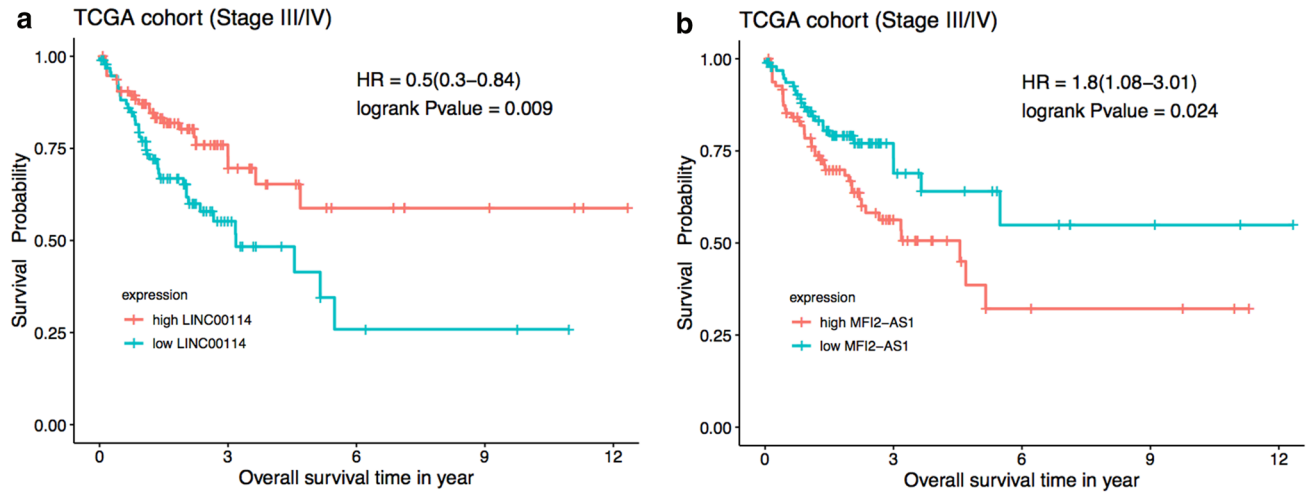

Fig. 6 KaplanMeier curves of two key lncRNAs related to OS in TCGA-COAD

with poor overall survivals of stage III/IV tumors after adjusting for clinical factors, indicating its significance in CRC pathogenesis and prognostic value in advanced stage tumors (Table 1).

WGCNA and Functional Annotation of Target mRNAs

According to previous research, the function of lncRNAs might be reflected on the adjacent coding genes by regulating their expression. In order to further investigate the regulatory mechanisms of this pivotal lncRNA related to OS via their target mRNAs, the co-expression network was constructed using WGCNA, yielding the identification of genes with similar expression patterns clustered into the same module. Clusters and traits of stage III/IV tissue samples from TCGA-COAD patients together with normal samples are shown in Fig. 7a. Twelve modules were ultimately determined in this co-expression network, with MFI2-AS1 included in the turquoise module (Fig. 7b). A total of 2276 genes were categorized in the turquoise module, of which 977 protein-coding genes were differentially expressed in stage III/IV CRCs and potentially regulated by MFI2-AS1. Among target genes of MFI2-AS1, 45 potentially associated mRNAs with weight above a threshold of 0.045 are presented in Fig. 7c.

Since the specific lncRNAs may function or enrich in potential pathways in a manner similar to mRNAs, inference on the function of MFI2-AS1 was done based on its correlated genes (in the turquoise module). Results from GO biological process displayed that MFI2-AS1 was mainly related to nuclear division and DNA replication. In addition, target genes of MFI2-AS1 were involved in

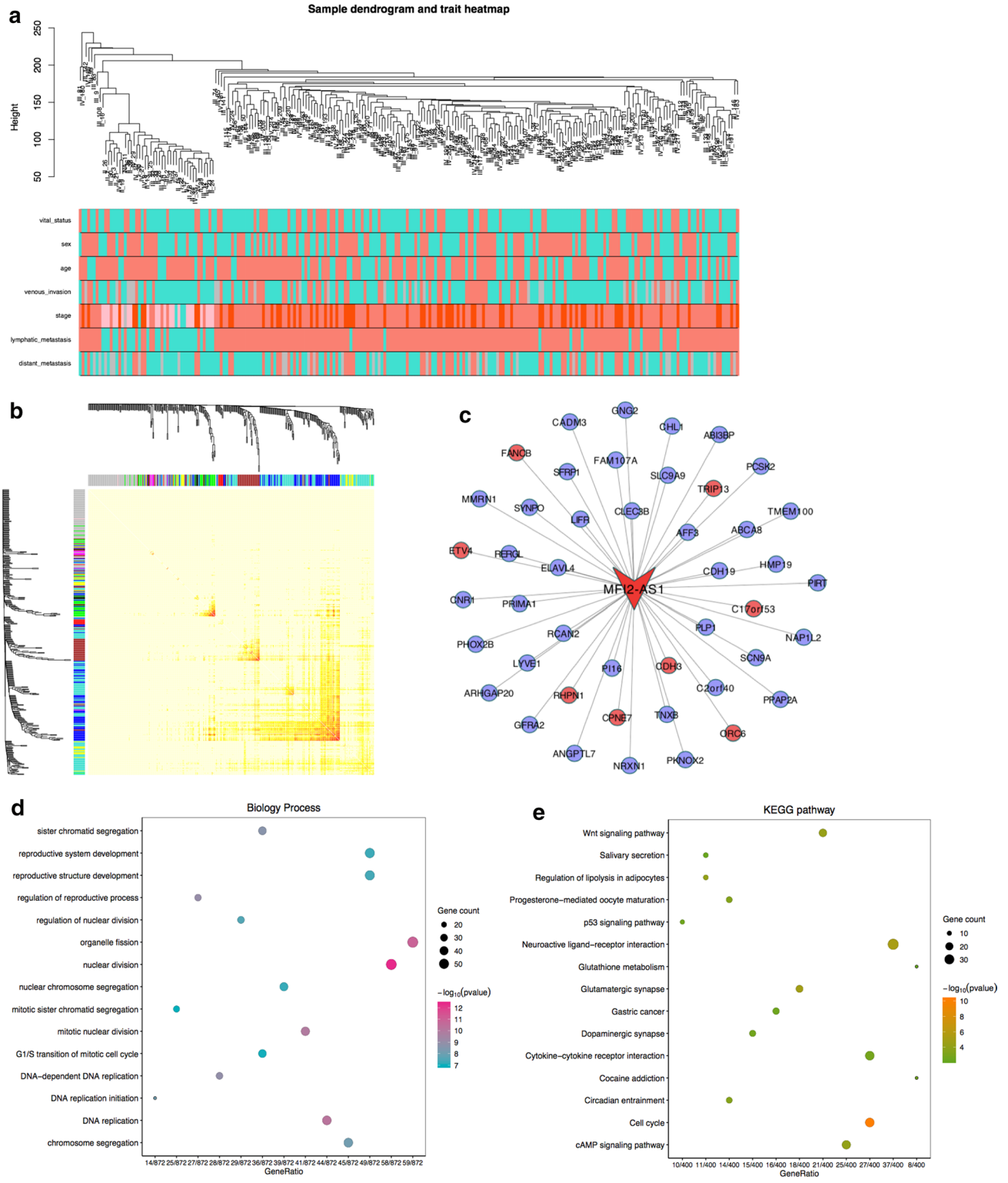

Fig. 7 **a** Clusters and traits of 199 stage III/IV tumor tissues and 41 normal samples. For stage, turquoise, pink, salmon and red represent stage I–IV, respectively; For sex, vital status and age, salmon, and turquoise represent male and female, dead and living, above and less than 60 years old, respectively; for venous invasion, lymphatic, and distant metastasis, salmon and turquoise indicate present and absent, respectively. **b** Construction of the gene correlation network. Each module is assigned a unique color. **c** Forty-five target mRNAs of MFI2-AS1 with weight above a threshold of 0.045. **d** Top 15 biological process GO terms. **e** Top 15 KEGG pathways about target mRNAs of MFI2-AS1

Table 2 Univariate and multivariate analyses of overall survivals in stage III/IV TCGA-COAD

Characteristics		Hazard ratio	95% CI	*P* value	Hazard ratio	95% CI	*P* value
Sex	Male versus female	0.94	0.57–1.54	0.802			
Age	≥ 60 versus < 60	1.52	0.88–2.61	0.133			
Adjuvant chemotherapy	Yes versus no	0.9	0.35–2.33	0.822			
	Unknown versus no	0.77	0.35–1.72	0.523			
History of colon polyps	Present versus absent	0.89	0.47–1.68	0.712			
	Unknown versus absent	1.61	0.85–3.05	0.14			
Lymphatic invasion	Present versus absent	1.26	0.72–2.22	0.415			
	Unknown versus absent	1.57	0.65–3.76	0.313			
Venous invasion	Present versus absent	1.83	1.09–3.1	0.023	1.82	**1.05–3.15**	0.033
	Unknown versus absent	1.1	0.48–2.52	0.82	1.25	**0.54–2.91**	0.602
Lymphatic metastasis	Present versus absent	0.34	0.14–0.87	0.024	0.42	**0.15–1.15**	0.092
Distant metastasis	Present versus absent	2.35	1.4–3.96	0.001	1.77	**1–3.14**	0.05
	Unknown versus absent	1.12	0.46–2.73	0.802	1.11	**0.45–2.71**	0.821
Preoperative CEA level	≥ 5 versus < 5 μg/mL	1.91	0.95–3.85	0.069			
	Unknown versus < 5 μg/mL	1.94	0.98–3.82	0.056			
UCA1	High versus low	1.02	0.62–1.68	0.941			
LINC00114	High versus low	0.5	0.3–0.84	0.009	0.59	**0.35–1**	0.052
SFTA1P	High versus low	1.09	0.67–1.79	0.722			
MFI2-AS1	High versus low	1.8	1.08–3.01	0.024	1.85	**1.09–3.13**	0.022
SNHG7	high versus low	1.33	0.81–2.19	0.258			
PVT1	High versus low	1.17	0.71–1.93	0.533			
SNHG1	High versus low	0.91	0.55–1.49	0.699			

the most paramount pathways, including cell cycle and cytokine–cytokine receptor interaction, indicating that these pathways might play important roles in the pathogenesis, metastasis and progression of CRC (Table 2).

Identification of Regulation of MFI2-AS1 by TFs

A total of 279 transcriptional factors potentially involved in regulating the expression of MFI2-AS1 were predicted by three reliable databases. Ultimately, 17 TFs were remained by Pearson correlation analysis (adj. $P < 0.05$) and were considered as potential modulators of MFI2-AS1. Differential expression of these TFs in stage III/IV CRC tumor tissues compared with normal samples and their correlations with expression of MFI2-AS1 are shown in Table 3. To further enrich the regulation mechanism, based on the co-expression network mentioned above, a total of 68 mRNAs co-expressed with MFI2-AS1 (weight above a threshold of 0.03) as well as overlapped with the target genes of a specific TF were included in the lncRNA–TF–mRNA network (Fig. 8). We found EBF1

Table 3 Expression of TFs and their correlations with MFI2-AS1 in TCGA

TF	*r*	adj. *P*	logFC	FDR
EGR1	−0.2	0.0185	–	–
EBF1	−0.5	< 0.0001	−1.14	< 0.0001
MEF2D	−0.58	< 0.0001	–	–
MAX	−0.3	0.0001	–	–
MNT	−0.2	0.0263	–	–
TFE3	−0.2	0.0023	–	–
ETS1	−0.5	< 0.0001	–	–
FOS	−0.3	0.0004	2.56	< 0.0001
AR	−0.4	< 0.0001	−1.3	< 0.0001
CEBPB	0.28	0.0001	1.31	< 0.0001
BACH1	−0.2	0.0025	–	–
NFE2	0.24	0.0007	2.2	< 0.0001
FOXP3	0.22	0.0019	–	–
ATF4	0.22	0.0017	–	–
ELF1	−0.2	0.0155	–	–
MGA	−0.2	0.0132	–	–
GABPA	−0.36	< 0.0001	–	–

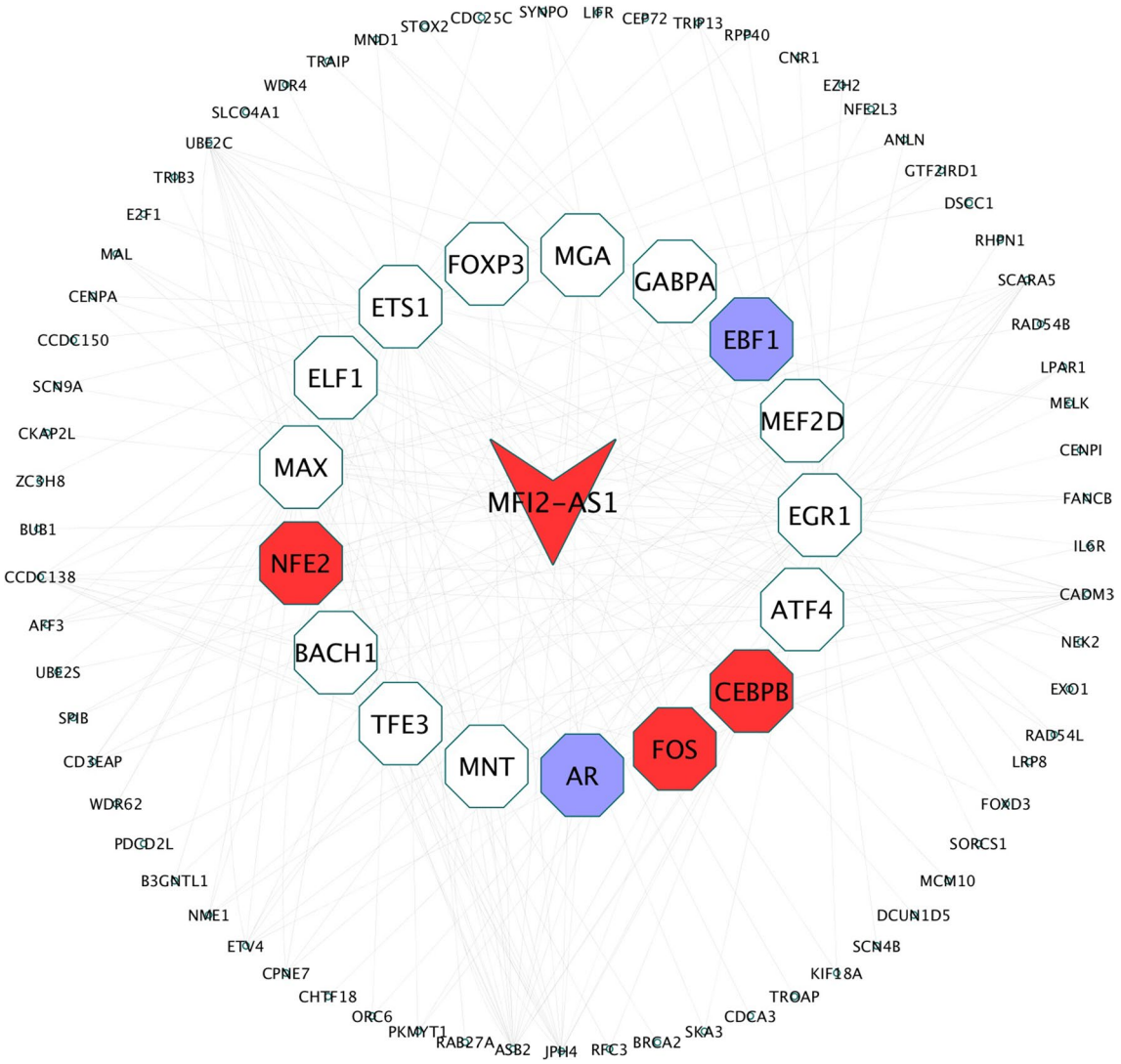

Fig. 8 The TF-lncRNA-coding gene regulation network of MFI2-AS1. Red represents being upregulated, while blue represents being downregulated. Balls represent mRNAs; Hexagans represent miRNAs; Vs represent lncRNAs; Gray edges indicate lncRNA-miRNA-mRNA interactions

and AR were downregulated and negatively correlated to MFI2-AS1, while FOS, CEBPB and NFE2 showed significant upregulation and positively associated with MFI2-AS1.

Discussion

Colorectal cancer has been regarded as a paramount public health problem that seriously endangers thousands and millions of lives, with 5-year survival rate of CRC patients remaining low (30–57%), especially of those in advanced stages [11]. On that account, the diagnosis and prognostic prediction of late-stage CRC is expected to be more accurate for personalized therapy and medication, which makes research and discovery for potent and effective biomarkers of CRC more and more essential and urgent. In this study, we are aiming at identifying the pivotal lncRNA as a reliable prognostic biomarker in advanced pathologic stages (III/IV) of CRC, and constructing robust regulatory networks.

Admittedly, in traditional point of view, the focus and concentration on investigation for cancer drivers should be protein-coding genes that are more likely to exist in reoccurring alterations in cancer genome. Nevertheless, after the discovery of noncoding RNAs, researchers have found that lncRNAs act as indispensible roles and have miscellaneous functions through diverse mechanisms, including interfering neighboring genes when transcribing, mediating chromatin remodeling and histone modifications, binding specific proteins, regulating protein activity and so on. As a new type of nonprotein-coding molecules, lncRNAs have been found in tumorigenesis and prognosis in multiple kinds of cancers. It

has been reported that lncRNAs may function as oncogenes or tumor inhibitors of cancers, such as CRC and lung cancer [12, 13]. Beyond any dispute, the ceRNA hypothesis that lncRNAs may act as an miRNA sponge and protect mRNA targets from miRNA-induced decay these days does suggest a novel regulatory mechanism that can be mediated by lncRNAs, which sheds a new light on our study for lncRNAs.

In the present work, based on TCGA and GEO data, we firstly harvested DEmRNAs, DElncRNAs, and DE miRNAs between integrated tumor tissues, stage III/IV CRC and adjacent normal tissues, respectively. Intersecting DEGs were then selected in four comparative groups. GO and KEGG pathway enrichment analysis was performed on intersecting DEmRNAs, of which results also demonstrated the function of DElncRNAs. Next, online bioinformatics tools were applied to construct the ceRNA network based on intersecting DEmiRNA, DElncRNAs, and DEmRNAs. We further validated the expression of key lncRNAs from aforementioned network in another GEO dataset and evaluated their accuracy for differentiating stage III/IV CRC from normal tissues by ROC curve analysis. Several candidate lncRNAs from the ceRNA network have already been verified previously. For example, UCA1 modulates breast cancer cell growth and apoptosis through UCA1/miR-143 regulation axes [14]. UCA1 and PVT1 were also upregulated and influence apoptosis in CRC [15, 16]. MFI2-AS1 presented upregulation in clear-cell renal cell carcinoma tissue, and the high level of its expression was correlated to dramatically increased risk of relaps and metastatic dissemination [17]. These studies also implied the credibility of our analytical results.

Subsequently, the univariate and multivariate Cox analyses were employed to confirm whether key lncRNAs involved in ceRNA networks could be independent prognostic indicators of CRC, finally yielding a pivotal lncRNA related to overall survivals, MFI2-AS1. We found that MFI2-AS1 significantly upregulated in stage III/IV CRC patients with a poor prognosis, implicating it is a prognostic indicator for the OS of advanced stage CRCs.

Then, a total of 977 target coding genes were identified to be co-expressed with MFI2-AS1 using WGCNA. According to enrichment analysis of potential target genes, these genes were found to function mainly through cell cycle as well as cytokine–cytokine receptor interaction, a signaling pathway that is reported to control stromal blood vessel network formation and mediate cancer cell proliferation [18]. Cytokine secreted by tumor stromal cells may give rise to chemotherapy resistance [19], which indicates this pathway associates with regulating stage III/IV CRC process and is of paramount significance for disease progression.

Ultimately, to optimize the regulatory mechanism of lncRNAs, we further revealed several reliable TFs on regulatory region of MFI2-AS1 and built lncRNA–TF–mRNA network. Among these TFs, CEBPB might function as a transcription activator in CRC through the MAPK pathway [20]. Elevated NFE2-related factor 2 expression promotes NSCLC progression through activating autophage [21]. EBF1 regulated transcription of USP5, a ubiquitous expressed deubiquitinating enzyme, and upregulated of which promoted growth of CRC cells and resistance to chemotherapeutics [22].

ZFAS1 has been reported to be a lncRNA that upregulated in CRC, whose downregulation inhibits proliferation and cell cycle of CRC cells [23]. What' more, ZFAS1 overexpression was induced by SP1 in CRC and ZFAS1 directly interacted with miR-150-5p, increasing expression of VEGFA and contributing to CRC progression and metastasis [24]. Our current study demonstrated that CEBPB ($r = 0.28$, adj. $P = 0.0001$) and EBF1 ($r = -0.5$, adj. $P < 0.0001$) were significantly correlated to MFI2-AS1, which was predicted competitively binding to 4 miRNAs (miR-103a, miR-125a, miR-145, miR-96). These observations suggested that MFI2-AS1 might involve in tumorigenicity of advanced stage CRC through RNA–RNA interactions and crucial regulatory role of TFs. Nonetheless, the specific regulatory mechanisms remain to be further explored.

Still, in this current study exists some limitations: firstly, failing to explore relationships between cancer transcriptomics and premalignant lesion due to excluding colorectal adenoma and stage I/II CRC samples in GEO and TCGA dataset. Second, ignoring the importance of unifying cross-platform standardization of gene profiling and validating with another independent cohort. Thirdly, being challenged by vitro and vivo experiments on the reliability of our analytical results owing to no experimental data on mechanisms of lncRNAs.

In summary, this study revealed a novel pivotal lncRNA, MFI2-AS1 in stage III/IV CRC, predicted its function and constructed regulatory networks in different levels by a series of bioinformatics methods. The lncRNA-miRNA-mRNA ceRNA network as well as TF-lncRNA-mRNA interactions shed substantial light on the regulation mechanism in CRC. AJCC classification helps identify some novel prognostic biomarkers and also enables a better investigation for carcinogenesis of CRC in advanced stages. Our findings provided a potential target that could be applied to novel personalized therapeutics for late-stage CRC patients, and further experimental validation is warranted.

Acknowledgments The authors would like to appreciate the GEO and TCGA databases for availability of data.

Author's contribution Ruihan Luo and Longke Ran conceived of the project; Ruihan Luo and Wangfeng Zhang collected and assembled the data; Ruihan Luo conducted the data analysis, interpretation and manuscript writing; Longke Ran, Jing Song and Ruihan Luo participated in

the discussion; Longke Ran and Jing Song performed paper revision. All of the authors read and approved the manuscript.

Funding Natural Science Foundation of Chongqing in China (Grant no.cstc2018jcyjAX0019 to Longke Ran).

Compliance with Ethical Standards

Conflict of interest The authors declare no conflicts of interest.

References

1. Melina A, Sierra Mónica S, Mathieu L, et al. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*. 2017;66:683–691.
2. Miller Kimberly D, Siegel Rebecca L, Chieh LC, et al. Cancer treatment and survivorship statistics, 2016. *CA Cancer J Clin*. 2016;66:271–289.
3. Gomes AQ, Nolasco S, Soares H. Non-coding RNAs: multi-tasking molecules in the cell. *Int J Mol Sci*. 2013;14:16010–16039.
4. Ye LC, Zhu X, Qiu JJ, Xu J, Wei Y. Involvement of long non-coding RNA in colorectal cancer: from benchtop to bedside (review). *Oncol Lett*. 2015;9:1039–1045.
5. Li J, Wang X, Tang J, et al. HULC and Linc00152 act as novel biomarkers in predicting diagnosis of hepatocellular carcinoma. *Cell Physiol Biochem*. 2015;37:687–696.
6. Xu CZ, Jiang C, Wu Q, Liu L, Yan X, Shi R. A feed-forward regulatory loop between HuR and the long noncoding RNA HOTAIR promotes head and neck squamous cell carcinoma progression and metastasis. *Cell Physiol Biochem*. 2016;40:1039–1051.
7. Hansen TB, Jensen TI, Clausen BH, et al. Natural RNA circles function as efficient microRNA sponges. *Nature*. 2013;495:384–388.
8. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. *Nature*. 2014;505:344–352.
9. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta stone of a hidden RNA language? *Cell*. 2011;146:353–358.
10. Zhao G, Fu Y, Su Z. How long non-coding RNAs and MicroRNAs mediate the endogenous RNA network of head and neck squamous cell carcinoma: a comprehensive analysis. *Cell Physiol Biochem*. 2018;50:332–341.
11. De Jong MC, Pulitano C, Ribero D, et al. Rates and patterns of recurrence following curative intent surgery for colorectal liver metastasis: an international multi-institutional analysis of 1669 patients. *Ann Surg*. 2009;250:440–448.
12. Qiao HP, Gao WS, Huo JX, Yang ZS. Long non-coding RNA GAS5 functions as a tumor suppressor in renal cell carcinoma. *Asian Pac J Cancer Prev*. 2013;14:1077–1082.
13. Enfield KS, Pikor LA, Martinez VD, Lam WL. Mechanistic roles of noncoding RNAs in lung cancer biology and their clinical implications. *Genet Res Int*. 2012;2012:737416.
14. Tuo YL, Li XM, Luo J. Long noncoding RNA UCA1 modulates breast cancer cell growth and apoptosis through decreasing tumor suppressive miR-143. *Eur Rev Med Pharmacol Sci*. 2015;19:3403–3411.
15. Han Y, Yang YN, Yuan HH, et al. UCA1, a long non-coding RNA up-regulated in colorectal cancer influences cell proliferation, apoptosis and cell cycle distribution. *Pathology*. 2014;46:396–401.
16. Takahashi Y, Sawada G, Kurashige J, et al. Amplification of PVT-1 is involved in poor prognosis via apoptosis inhibition in colorectal cancers. *Br J Cancer*. 2014;110:164–171.
17. Flippot R, Mouawad R, Spano JP, et al. Expression of long non-coding RNA MFI2-AS1 is a strong predictor of recurrence in sporadic localized clear-cell renal cell carcinoma. *Sci Rep*. 2017;7:8540.
18. Jones VS, Huang RY, Chen LP, et al. Cytokines in cancer drug resistance: cues to new therapeutic strategies. *Biochim Biophys Acta (BBA) Rev Cancer*. 1865;2:255–265.
19. Tan C, Hu W, He Y, et al. Cytokine-mediated therapeutic resistance in breast cancer. *Cytokine*. 2018;108:151–159.
20. Barbagallo C, Brex D, Caponnetto A, et al. LncRNA UCA1, upregulated in CRC biopsies and downregulated in serum exosomes, controls mRNA expression by RNA–RNA interactions. *Mol Ther Nucleic Acids*. 2018;12:229–241.
21. Wang J, Liu Z, Hu T, et al. Nrf2 promotes progression of non-small cell lung cancer through activating autophagy. *Cell Cycle*. 2017;16:1053–1062.
22. Xin X, Aiwen H, Ximao C, et al. Ubiquitin specific peptidase 5 regulates colorectal cancer cell growth by stabilizing Tu translation elongation factor. *Theranostics*. 2019;9:4208–4220.
23. Nithyananda T, Petra F-V, Sonja H, et al. Long non-coding RNA ZFAS1 interacts with CDK1 and is involved in p53-dependent cell cycle control and apoptosis in colorectal cancer. *Oncotarget*. 2016;7:622–637.
24. Xiaoxiang C, Zeng Kaixuan X, et al. SP1-induced lncRNA-ZFAS1 contributes to colorectal cancer progression via the miR-150-5p/VEGFA axis. *Cell Death Dis*. 2018;9:982.