CrossMark

ORIGINAL ARTICLE – CANCER RESEARCH

Identification of long noncoding RNAs as potential novel diagnosis and prognosis biomarkers in colorectal cancer

Rui Wang[1] · Lutao Du[1] · Xiaoyun Yang[2] · Xiumei Jiang[1] · Weili Duan[1] · Suzhen Yan[1] ·
Yujiao Xie[1] · Yuntao Zhu[3] · Qingliang Wang[4] · Lili Wang[1] · Yongmei Yang[1] ·
Chuanxin Wang[1]

Abstract

Background Colorectal cancer (CRC) is prevalent worldwide, and improvements in timely and effective diagnosis are imperatively needed. We aimed to identify potential long noncoding RNA (lncRNA) biomarkers for CRC diagnosis along with prognosis prediction.

Methods LncRNA expression profiles were studied by microarray in paired tumor and normal tissues from six patients with CRC. The expression levels of candidate lncRNAs were analyzed in 80 pairs of tissues and two independent cohorts of serum samples. Receiver-operating characteristic (ROC) curves were employed to evaluate the performance of the lncRNAs identified. The correlation between lncRNAs and disease-specific survival rate of CRC patients was assessed to explore prognostic potential.

Results Four lncRNAs (BANCR, NR_026817, NR_029373, and NR_034119) were identified to be significantly dysregulated in both tissue and serum samples with consistent pattern, and a panel was established based on this result. The performance of the 4-lncRNA panel was measured with an area under the ROC curve (AUC) of 0.881. The corresponding AUCs of the panel for patients with TNM stageI, II and III were 0.774, 0.844 and 0.949, respectively, significantly higher than that of CEA. Kaplan–Meier analysis showed that patients with low levels of NR_029373 and NR_034119 had significantly lower disease-specific survival rate ($p = 0.013$ and 0.044, respectively). Multivariate Cox analysis demonstrated that NR_029373 and NR_034119 were both independently associated with disease-specific survival rate ($p = 0.013$ and 0.038, respectively).

Conclusions Our study established a distinctive 4-lncRNA panel with considerable diagnostic value and identified NR_029373 and NR_034119 as potential biomarkers for CRC prognosis prediction.

Rui Wang and Lutao Du contributed equally to this article.

Electronic supplementary material The online version of this article (doi:10.1007/s00432-016-2238-9) contains supplementary material, which is available to authorized users.

✉ Chuanxin Wang
cxwang@sdu.edu.cn

1 Department of Clinical Laboratory, Qilu Hospital of Shandong University, 107 Wenhua Xi Road, Jinan 250012, Shandong Province, China

2 Department of Gastroenterology, Qilu Hospital of Shandong University, 107 Wenhua Xi Road, Jinan 250012, Shandong Province, China

3 Department of Clinical Laboratory, Jinxiang County People's Hospital, Jining 272000, Shandong Province, China

4 Department of Medical Affairs Management, Qilu Hospital of Shandong University, 107 Wenhua Xi Road, Jinan 250012, Shandong Province, China

Background

Colorectal cancer (CRC) is currently the third most common malignancy and the second leading cause of cancer-related death in the world (Siegel et al. 2016). Most of the patients are diagnosed at the advanced stage, where tumor invasion and early systemic dissemination have already occurred, missing the optimum period for curative resection (Figueredo et al. 2008). The dismal outcome of this disease has aroused the attention to the critical importance

 Springer

of early detection. The wide use of fecal occult blood testing (FOBT) and colonoscopy examination in the past decades has moderately increased the detection rate of early-stage tumors. FOBT is a cost-effective and safe test, but it has relatively poor sensitivity and is prone to produce false results as colorectal tumors intermittent bleed (Health Quality Ontario 2009; Duffy et al. 2011). In contrast, colonoscopy examination offers high diagnostic accuracy, but the invasive nature and high cost render it impractical for large-scale screening (Kim et al. 2007; Mandel. 2008). Stool DNA test targeted at molecular biomarkers has been commercially available for twelve years, but the high cost and poor sensitivity limited its clinical application (Bailey et al. 2016; Itzkowitz et al. 2007). CEA and other CRC traditional biomarkers, like CA242, have been commonly used in clinical practice. However, their sensitivity is also far from satisfying (Levin et al. 2008; Zhong et al. 2015). Thus, there is an urgent need to develop a new noninvasive, sensitive and cost-effective method to complement and improve the current CRC screening strategies.

Long noncoding RNAs (lncRNAs) are a class of noncoding RNAs that are longer than 200 nucleotides and do not translate into proteins (van Werven et al. 2012; Ponting et al. 2009). These lncRNAs have recently attracted increasing research interest due to their important role in the regulation of multiple biological processes including initial tumor development and its progression. Aberrant lncRNA expression has been detected in tissues sections from breast cancer, gastric cancer, esophageal squamous cell cancer (ESSC), non-small cell lung cancer (NSCLC) and colorectal cancer (Yang et al. 2016; Nie et al. 2016; Xu et al. 2016; Feng et al. 2015; Sun et al. 2016a). Recently, lncRNAs were found to be present in bloodstream in a stable state and may reflect the physiological and pathological alterations of patients with cancer, which excited great interest among researchers in investigating the possibility of using circulating lncRNAs as surrogate minimally invasive biomarkers. Despite the growing body of the literature characterizing the possibility of circulating lncRNAs as biomarkers, there are still no reports regarding the potential role of circulating lncRNAs in diagnosis of CRC patients (Wang et al. 2015; Tong et al. 2015; Zhou et al. 2015).

Here, we systematically investigated the expression of specific lncRNAs using a three-phase study. In the initial screening phase, microarray analysis was employed to mine potential candidate lncRNAs integrated with previous studies (Ge et al. 2013; Li et al. 2015; Ni et al. 2015; Qi et al. 2013; Sun et al. 2016b; Tong et al. 2014; Wang et al. 2016b; Yao et al. 2016; Yin et al. 2014, 2015) In the training phase, we performed clinical validation of lncRNA expression status using both tissue and serum samples to identify

lncRNAs dysregulated with consistent pattern in these clinical materials and constructed a diagnostic panel based on the result. In the validation phase, lncRNAs identified were further verified and the diagnostic performance of the panel was further validated. In addition, we also assessed the correlation between the expression level of lncRNAs identified and the disease-specific survival rate of CRC patients, to explore their potential for prognostic prediction.

Methods

Ethics statement

This study has been conducted under the supervision of the Clinical Research Ethics Committee of Qilu Hospital, Shandong University. The written informed consent was obtained from each participant prior to tissue and blood sample collection, and all the experiments were performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

Study design

To identify potential serum lncRNA biomarkers for colorectal cancer, step-by-step discovery procedure was designed including three phases—initial screening phase, training phase and validation phase. The lncRNA candidates in each phase were determined based on the profiling results of prior phase of study. In the initial screening phase, primary CRC tissues along with corresponding adjacent non-tumor (NM) tissues from six different CRC patients were subjected to microarray analysis, to identify lncRNAs significantly differentially expressed. Candidate lncRNAs were then selected according to the result of the analysis and previous studies (Ge et al. 2013; Li et al. 2015; Ni et al. 2015; Qi et al. 2013; Sun et al. 2016b; Tong et al. 2014; Wang et al. 2016b; Yao et al. 2016; Yin et al. 2014, 2015). In the training phase, lncRNA candidates were firstly verified by RT-qPCR in 80 pairs of tissue specimens, involving primary CRC tissues and matched NM tissues. Subsequently, the verified lncRNA was further examined in an independent cohort of serum samples obtained from 120 CRC patients and 120 controls. The data from this were used to construct the diagnostic panel, based on the logistic regression model for differential diagnosis between the CRC and the control group. In the validation phase, the diagnostic lncRNA panel constructed in the training phase was applied to another independent cohort of serum samples from 240 patients (120 CRC patients and 120 controls) to validate its diagnostic

performance. In addition, the correlation between the expression of identified lncRNAs and disease-specific rate of CRC patients was assessed to explore their potential as predictors for CRC prognosis.

Patients and control subjects

All participants involved were recruited from Qilu Hospital of Shandong University, between 2007 and 2015. Demographic and clinical features of CRC patients and controls have been described in Table S1. Tissue specimens were obtained from patients who underwent primary tumor resection at the Department of General Surgery. Serum samples were collected before any anticancer treatments such as surgery, chemotherapy and radiotherapy were given. All CRC patients were clearly diagnosed based on histopathology or biopsy analysis. Tumor stage was defined according to the tumor-node-metastasis (TNM) staging system of Union for International Cancer Control (UICC).

CRC patients in the validation phase have been followed up since diagnosis confirmed, at intervals of 3 months in the first 2 years and at 6 months interval thereafter, up to the fifth year. Disease-specific survival was defined as the interval from date of diagnosis to CRC-related death. The date of latest record retrieved was June 20, 2015. Owing to incomplete follow-ups, 13 of all the cases were excluded from the cohort, and the median follow-up time was 62 months (range, 11–74 months).

Sample collection and preparation

Fresh tumor tissue and paired adjacent normal tissue sections were immediately cut from the resected colorectal tissue and kept at −80 °C until RNA extraction. Venous blood was collected and centrifuged at 4000 rpm for 10 min, within 2 h. The supernatant fluids were then collected and further centrifuged at 12,000 rpm for 15 min to completely remove the cell debris. The whole process was strictly controlled to avoid hemolysis, and the supernatant serum was stored at −80 °C, until analysis. The CEA level of each sample was determined using Roche cobas e601 (Roche, Switzerland).

LncRNA microarray analysis

Total RNA extracted from six pairs of fresh primary colorectal cancer tissues and their adjacent normal tissues were subjected to human genome-wide lncRNA microarray (Arraystar Human LncRNA Microarray V2.0; Agilent Technology, Santa Clara, CA) analysis. The whole process and subsequent data analysis were performed by Kangchen Bio-tech, Shanghai P.R. China.

RNA isolation and RT-qPCR analysis

Total RNA from frozen tissues and serum samples were extracted using TRIzol and TRIzol LS reagents (Invitrogen, Carlsbad, CA), respectively, according to the manufacturer's protocol. The quantity and quality of RNA were measured by NanoDrop spectrophotometer. The reverse transcription (RT) reactions were performed using a Prime Script™ RT Reagent Kit (Takara, Dalian, Liaoning).After mixing with 1 μg of template RNA, 4 μL of 5× Prime Script Buffer Mix, 1 μL of Prime Script RT Enzyme MixI, 1 μL of Oligo dT Primer and RNase-free dH$_2$O in a final volume of 20 μL, the reaction volumes were incubated at 37 °C for 30 min, followed by 85 °C for 5 s and 4 °C for 60 min. For real-time PCR, 2 μL of diluted generated cDNAs was mixed with 12.5 μL of SYBR Premix Ex TaqTM, 0.5 μL of DyeII, 1 μL forward and reverse primers (10 μM) and 9 μL of nuclease-free water in a final volume of 25 μL, according to the manufacturer's instructions (Takara, Dalian, Liaoning). The reactions were incubated at 95 °C for 30 s, followed by 45 cycles of 95 °C for 5 s and 60 °C for 34 s. Melting curve analysis was performed to evaluate the specificity of the RT-qPCR products. All reactions were run on CFX96™ real-time system (Bio-Rad, CA, American). Each RT-qPCR experiment was repeated three times. Relative expression of genes was calculated using the comparative cycle threshold (Ct) ($2^{-\Delta\Delta Ct}$) method with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as the endogenous control. The average of Ct values of all normal samples was used as a control sample in this method.

Statistical analysis

LncRNA levels among different groups of subjects were normalized by endogenous control, GAPDH. The distribution of the lncRNA relative expression (based on $2^{-\Delta\Delta Ct}$ method) of each group was determined by Kolmogorov–Smirnov test. Data were presented as median (interquartile range). Nonparametric Mann–Whitney *U* test was used to compare the differences in expression of lncRNAs between the CRC and the control group. Receiver-operating characteristic (ROC) curves and the area under the ROC curve (AUC) were employed to evaluate the performance of the selected lncRNA panel to discriminate CRC patients from controls. Kaplan–Meier method was utilized to estimate survival curves, and the log-rank test was used to make comparisons. The Cox proportional hazards regression model was used to identify the independent prognostic factors. ROC analysis was processed by MedCalc 15.2.2 (MedCalc, Mariakerke, Belgium) software. MATLAB software (MATLAB, R2013a) was used for logistic regression analysis to establish lncRNA panel. Other analysis was

performed using SPSS version 19.0 software (SPSS, Chicago, IL). A *p* value of <0.05 was considered as statistically significant.

Results

Identification of candidate lncRNAs by microarray analysis

The microarray analysis with six pairs of primary colorectal and adjacent normal tissues was conducted with a microarray targeting 33,000 lncRNAs. In total, 5873 lncRNAs were identified with significant differential expression (fold change ≥2.0 and *p* < 0.05). To identify the lncRNAs that were potential biomarkers, we concentrated on the lncRNAs with fold change >10, including 301 lncRNAs upregulated in the tumor tissues and 122 downregulated. Starting from those lncRNAs with the greatest fold changes, we filtered appropriate candidate lncRNAs in descending order. Candidates should be plausible for prime designing, and only those have steady expressions in tissue and serum samples were selected. Finally, we chose five candidate lncRNAs from the upregulated group and five from the downregulated group as well (Table 1). Another eight

lncRNAs were also tested by RT-qPCR because they had been shown dysregulated in malignant cancer tissues (Ge et al. 2013; Li et al. 2015; Ni et al. 2015; Qi et al. 2013; Sun et al. 2016b; Tong et al. 2014; Wang et al. 2016b; Yao et al. 2016; Yin et al. 2014, 2015). Eighteen candidate lncRNAs were selected in total.

Analysis of selected lncRNA expression by RT-qPCR

The expression of all 18 candidate lncRNAs was evaluated by RT-qPCR, using 80 pairs of CRC and control tissue samples. Among these, eight lncRNAs (MALAT-1, NR_029373, NR_046321, NR_034119, GAS5, BANCR, NR_026817 and PCAT-1) were found significantly dysregulated in CRC tumor tissues compared with controls (Table 2). Subsequently, these eight lncRNAs were further analyzed in an independent cohort of 240 serum samples including 120 CRC patients and 120 controls. One lncRNA (BANCR) that was upregulated and three other lncRNAs (NR_026817, NR_029373 and NR_034119) whose expression was downregulated in CRC patients compared to control were finally identified (Fig. 1a–d; Table S2). The corresponding AUCs of the four lncRNAs (BANCR, NR_026817, NR_029373 and NR_034119) were 0.638, 0.708, 0.812 and 0.724, respectively (Fig. 2a–d).

Table 1 Candidate lncRNAs selected on a basis of the microarray analysis

Seqname	Chromosome	RNA length	Regulation	Fold change	*p* value
NR_046321	chr14	1310	Down	40.0482399	4.14303E−05
NR_029373	chr4	1695	Down	30.00485043	0.040111472
NR_034119	chr13	935	Down	26.82814191	0.002132091
NR_038940	chr10	2023	Down	23.91285034	0.010075854
NR_026817	chr1	4241	Down	14.29771802	0.032489718
NR_015423	chr9	536	Up	117.2317631	3.69889E−05
TCONS_00018898	chr10	369	Up	81.88003799	0.000366929
TCONS_00008650	chr4	3118	Up	46.5057635	5.32987E−06
NR_046438	chr2	4115	Up	37.79538509	0.000178468
ENST00000430320	chr1	670	Up	37.1051264	0.00126013

Table 2 Expression of 18 candidate lncRNAs in CRC and corresponding adjacent non-tumor tissues [median (interquartile range)]

LncRNA	Controls	CRCs	*p* value	LncRNA	Controls	CRCs	*p* value
TCONS_00018898	1.21 (0.40–2.87)	1.17 (0.63–2.16)	0.78	GAS5	1.07 (0.58–1.82)	0.88 (0.41–1.33)	<0.01
TCONS_00008650	0.87 (0.36–2.73)	0.66 (0.25–1.65)	0.07	BANCR	1.24 (0.39–2.57)	2.56 (1.28–7.06)	<0.01
MALAT1	0.86 (0.48–1.80)	1.45 (0.96–2.22)	0.03	NR_026817	1.36 (0.33–2.97)	0.61 (0.31–1.70)	<0.01
NR_029373	1.26 (0.42–3.06)	0.59 (0.37–1.94)	<0.01	NR_038940	1.14 (0.45–2.50)	0.97 (0.55–1.53)	0.07
Loc285194	1.08 (0.33–3.72)	0.77 (0.47–1.96)	0.07	PCAT-1	1.15 (0.47–2.84)	2.42 (1.14–3.74)	0.03
NR_046321	1.23 (0.27–4.46)	0.63 (0.28–1.71)	<0.01	MEG3	1.02 (0.39–2.26)	1.28 (0.90–2.56)	0.10
NR_034119	1.15 (0.43–2.81)	0.80 (0.21–1.64)	<0.01	NR_046438	0.93 (0.47–2.12)	1.21 (0.78–2.25)	0.09
ENST00000430320	1.58 (0.30–3.43)	1.06 (0.32–2.82)	0.31	UCA1	1.06 (0.50–2.14)	1.03 (0.49–1.67)	0.31
ANRIL	1.19 (0.45–2.08)	1.09 (0.57–2.11)	0.16	NR_015423	0.88 (0.45–2.47)	0.96 (0.31–2.78)	0.89

Fig. 1 Concentrations of four identified serum lncRNAs in patients with CRC ($n = 120$) and control individuals ($n = 120$) using RT-qPCR assay in training set (**a–d**), *$p < 0.001$
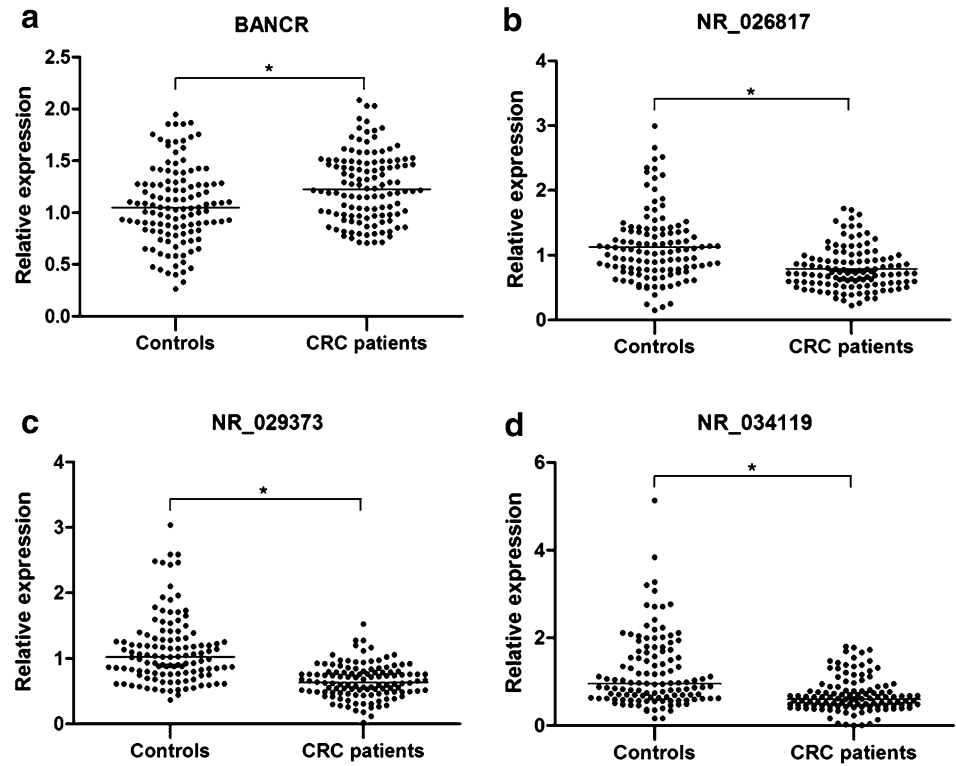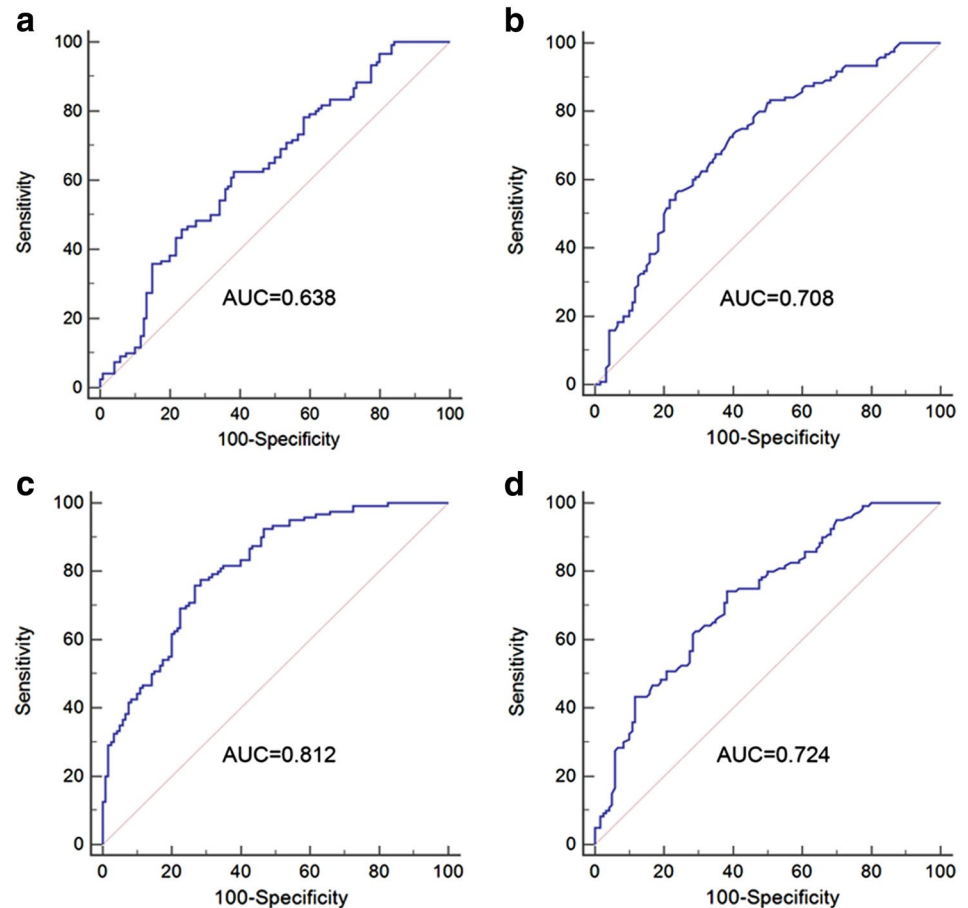

Fig. 2 ROC curve analysis for the detection of CRC using BANCR (**a**), NR_026817 (**b**), NR_029373 (**c**) and NR_034119 (**d**) in patients with CRC ($n = 120$) and control individuals ($n = 120$) in training set

Identification of predictive lncRNA panel

Retrospectively, a stepwise logistic model was constructed for CRC diagnosis using the 240 serum samples enrolled in the training phase. The predicted probability of being diagnosed with CRC from the model based on the 4-lncRNA panel was calculated using the equation as follows: Logit(P) $= -1.0468 - 0.5716 \times$ BANCR $+ 0.6212 \times$ NR_026817 $+ 0.8296 \times$ NR_029373 $+ 0.3967 \times$ NR_034119. ROC analysis was used to evaluate the diagnostic performance of the established lncRNA panel. The AUC for the 4-lncRNA panel was 0.891 (95 % confidence interval [CI] = 0.844–0.927, sensitivity = 81.67 % and specificity = 80.00 %, Fig. 3a).

Validation of the lncRNA panel

The four lncRNAs identified in the training phase were further measured using another validation data set (another independent cohort including 120 CRC patients and 120 controls). No significant differences were observed in the distribution of age, gender and tumor characteristic for the CRC and control group samples between the training and the validation sets (Table S1). The alteration in the lncRNA expression pattern of the samples in the validation phase cohort was consistent with those in the training phase serum samples (Table S2; Fig. S1). The corresponding AUCs of the four lncRNAs (BANCR, NR_026817,
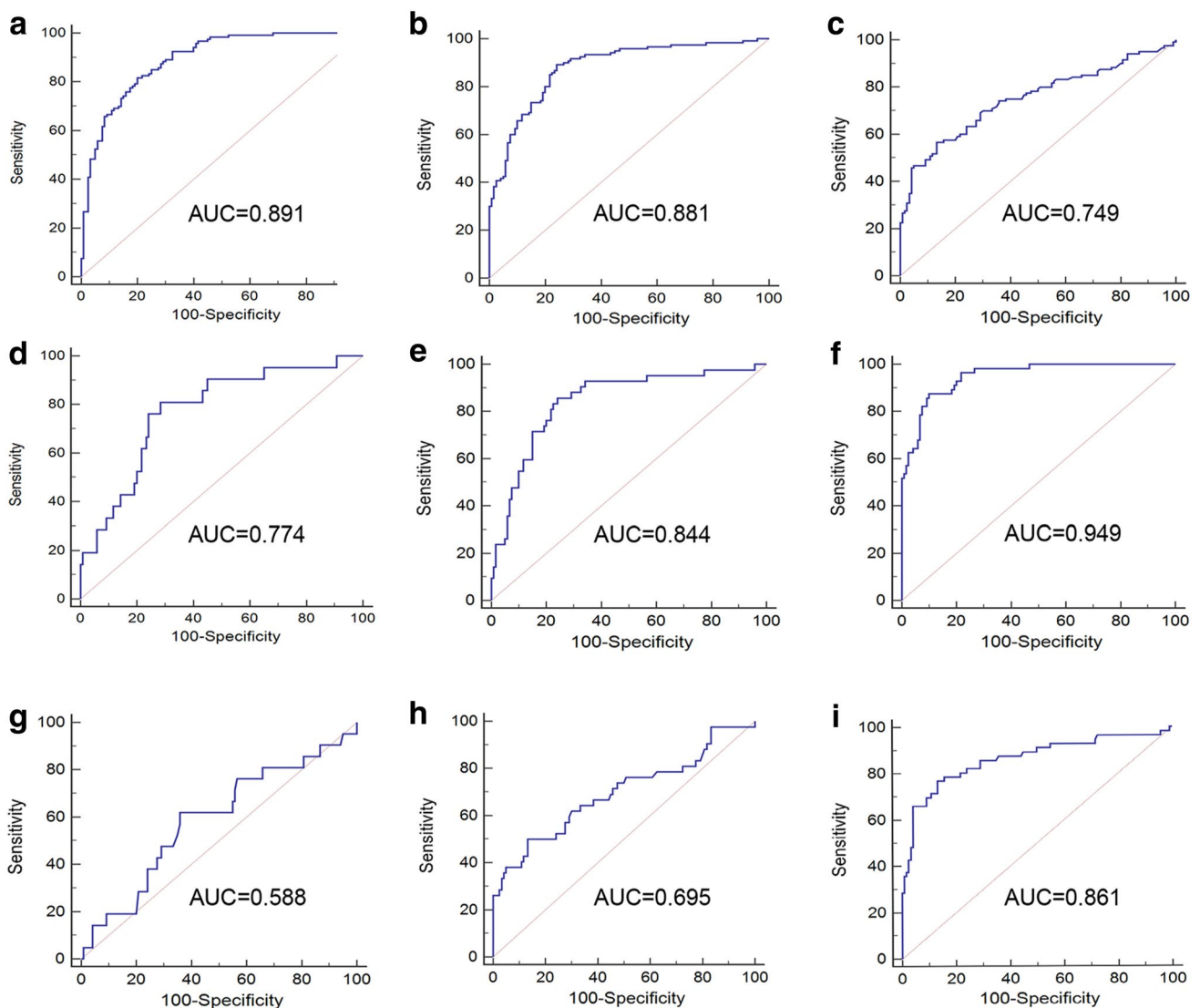

Fig. 3 (**a**, **b**) ROC curves for the detection of CRC using 4-lncRNA panel in training set (**a**) and validation set (**b**); (**c**) ROC curve analysis using CEA for the detection of CRC in validation set; (**d–f**) ROC curves using the 4-lncRNA panel for the detection of patients with TNM stage I (**d**), II (**e**) and III (**f**) in validation set; (**g–i**) ROC curve analysis using CEA for the detection of CRC stage I (**g**), II (**h**) and III (**i**) in validation set

NR_029373 and NR_034119) were 0.659, 0.752, 0.806 and 0.706, respectively (Fig. S2).

The diagnostic performance of the identified 4-lncRNA panel was further assessed. The predicted probability was calculated based on the parameters obtained from the training set and used to construct the ROC curve. The AUC of the 4-lncRNA panel was 0.881 (95 % confidence interval [CI] = 0.833–0.919, sensitivity = 89.17 % and specificity = 75.83 %, Fig. 3b), which was significantly better than that of CEA (AUC: 0.749, 95 % confidence interval [CI] = 0.689–0.803, sensitivity = 56.67 % and specificity = 86.67 %, $p < 0.001$, Fig. 3c).

Furthermore, we then compared the diagnostic performance of this 4-lncRNA panel with CEA, in discriminating CRC patients from control individuals at different TNM stages. The AUCs of the 4-lncRNA panel for patients with TNM stageI, II and III were 0.774, 0.844 and 0.949, respectively (Fig. 3d–f), and were all higher than those of CEA, which were 0.588, 0695 and 0.861, respectively (Fig. 3g–i).

Correlation between the four lncRNAs and clinicopathological characteristics

The data summarized in Table S3 show the relationship between the four identified lncRNAs and the clinicopathological characteristics of the patients with CRC in the validation set. Higher levels of serum lncRNA BANCR and lower levels of NR_026817, NR_029373, NR_034119 significantly correlated with advanced TNM stage ($p < 0.05$) and lymph node metastasis ($p < 0.05$). Higher levels of BANCR correlated with tumor local invasion ($p < 0.05$). However, no significant associations were found between the four lncRNAs with age, gender, tumor location, size or differentiation (all at $p \geq 0.05$).

Correlation of lncRNAs expression between tissues and serum

We analyzed the correlation between the expression levels of BANCR, NR_026817, NR_029373, NR_034119 in CRC tissues and serum. As shown in Fig. S3, a significant correlation was observed for all four lncRNAs (BANCR: $r = 0.504$, $p = 0.005$; NR_026817: $r = 0.611$, $p < 0.001$; NR_029373: $r = 0.573$, $p < 0.001$; NR_034119: $r = 0.589$, $p < 0.001$). It suggested that there is a consistency in the expression of these four lncRNAs at both tissue and serum levels.

Analysis of the identified lncRNAs stability in serum

The serum from five CRC patients was treated under different harsh conditions to assess the stability of these four identified lncRNAs, BANCR, NR_026817, NR_029373 and NR_034119. The serum was collected and subsequently exposed to harsh conditions, such as storage at room temperature for 4, 8 and 24 h or incubated at −80 °C and repetitive freeze–thaw cycles for 0, 2, 4 and 8 times. No obvious alterations in the expression level of these four lncRNAs were observed, as shown in Fig. S4, suggesting their stability.

Correlation between the four lncRNAs expression and disease-specific survival rate

Survival analysis has been finally carried on 107 patients since 13 patients were lost to follow up. Kaplan–Meier survival analysis showed that patients with low expression of NR_029373 and NR_034119 had significantly lower disease-specific survival rate, compared with those with high expression ($p = 0.013$ and 0.044, respectively, Fig. 4). A statistically significant correlation was also observed between disease-specific survival rate and NR_029373 expression ($p = 0.016$), NR_034119 expression ($p = 0.049$), size ($p = 0.007$) and TNM stage ($p < 0.001$) using univariate Cox proportional hazards regression model analysis. Parameters significantly related to disease-specific survival based on univariate analysis were further analyzed by multivariate analysis to identify the independent prognostic factors. The NR_029373 expression ($p = 0.013$), NR_034119 expression ($p = 0.038$), size ($p = 0.006$) and TNM stage ($p = 0.022$) were finally identified as statistically significant prognostic factors (Table 3).

Discussion

In our study, microarray analysis was firstly employed to provide basic information of lncRNAs significantly dysregulated in CRC tissues. Candidate lncRNAs were selected, compiled the microarray analysis result and previous studies, and then evaluated by RT-qPCR in tissues and serum samples to validate their consistent pattern of dysregulation in these clinical materials (Ge et al. 2013; Li et al. 2015; Ni et al. 2015; Qi et al. 2013; Sun et al. 2016b; Tong et al. 2014, Wang et al. 2016b; Yao et al. 2016; Yin et al. 2014, 2015). Four lncRNAs (BANCR, NR_026817, NR_029373 and NR_034119), which showed considerable discriminating potential to identify CRC patients from control with high AUC values, were finally identified. Using the multivariate logistic regression model, we established a panel of four lncRNAs that can diagnose CRC patients with higher accuracy in comparison with traditional diagnostic biomarker like CEA. In addition, we also identified NR_029373 and NR_034119 as independent factors for CRC patient prognosis.
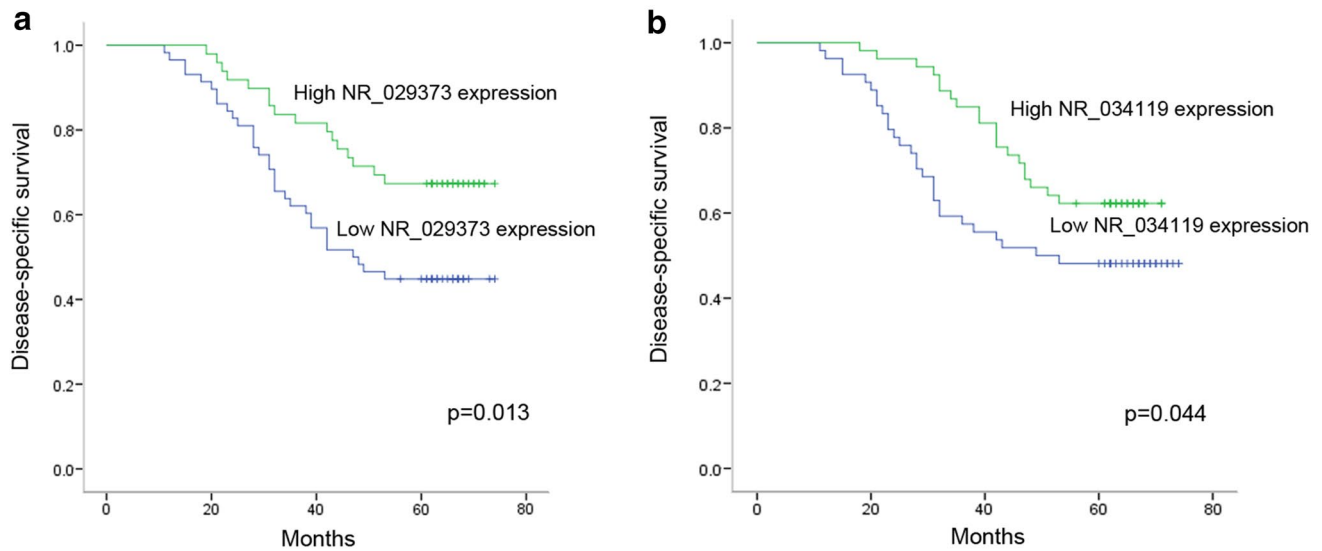
Fig. 4 Kaplan–Meier curves for disease-specific survival rate according to the serum levels of NR_029373 (**a**) and NR_034119 (**b**) in patients with CRC in validation set

Table 3 Univariate and multivariate analysis for predictors of disease-specific survival in the clinical validation cohort

Parameters	Univariate analysis			Multivariate analysis		
	HR	95 % CI	p	HR	95 % CI	p
Gender	1.065	0.602–1.884	0.828			
Age	1.617	0.910–2.873	0.101			
Tumor location	0.968	0.526–1.783	0.917			
Size	2.278	1.247–4.163	0.007	2.402	1.287–4.486	0.006
Differentiation	0.655	0.364–1.177	0.157			
Local invasion	1.418	0.737–2.726	0.295			
Lymph node metastasis	2.670	1.484–4.802	0.001	0.246	0.042–1.445	0.121
TNM stage	2.377	1.511–3.740	<0.001	5.618	1.281–24.647	0.022
BANCR expression	1.306	0.740–2.305	0.357			
NR_026817 expression	0.584	0.329–1.037	0.066			
NR_029373 expression	0.477	0.261–0.870	0.016	0.457	0.246–0.848	0.013
NR_034119 expression	0.561	0.316–0.996	0.049	0.517	0.277–0.964	0.038

HR hazard ratio, *CI* confidence interval

The expression profile of 18 candidate lncRNAs was first explored in CRC tissues and adjacent NM tissues in the training phase of our study (Liu et al. 2015a; Ren et al. 2015). Among these, eight lncRNAs (MALAT-1, NR_029373, NR_046321, NR_034119, GAS5, BANCR, NR_026817 and PCAT-1) were significantly identified to be differentially expressed between CRC tissues and the adjacent NM tissues. The result showed that the expression of GAS5 significantly reduced in CRC tissues, while expression of MALAT-1 and PCAT-1 lncRNAs increased, and this trend was consistent with previous studies (Yin et al. 2014; Ma et al. 2016; Ji et al. 2014; Ge et al. 2013). Guo et al.

have reported that BANCR was overexpressed in CRC, while Shi et al. claimed its reduced expression (Guo et al. 2014; Shi et al. 2015). In the present study, it was observed upregulated in CRC tissues. In addition, exploring the role lncRNAs in tumor carcinogenesis may help to decipher their oncogenic or suppressor function in CRC patients. Yin et al. showed that GAS5 overexpression could inhibit in vivo CRC cell proliferation, which indicated that it may function as a tumor suppressor in CRC progression (Yin et al. 2014). The MALAT-1 lncRNA has been reported to be involved in RC cell proliferation, migration and invasion (Ji et al. 2014). In study by Yang et al., it was demonstrated that MALAT-1

promoted the growth of tumor cells via a target protein called AKAP-9 (Yang et al. 2015). There have been no functional studies about the role of PCAT-1 in development of CRC, but PCAT-1 has been identified to be associated with the tumor cell proliferation in several other cancers including NSCLC and bladder cancer (Zhao et al. 2015; Liu et al. 2015b). The deregulations of the expression of other four lncRNAs (NR_029373, NR_034119, NR_026817 and NR_046321) were firstly reported in our study.

The reliance on surgical resection and invasive procedures limits the application of tissue lncRNAs in cancer diagnosis, which highlights the merit of lncRNAs in circulation (El-Tawdi et al. 2016; Gu et al. 2016; Tong et al. 2015; Wang et al. 2015, 2016a; Wu et al. 2016). We further measured the expression of identified tissue lncRNAs in serum samples. LncRNAs dysregulated in both tissue and serum with the consistent pattern could effectively represent the lncRNA expression alteration of tumor tissues and simultaneously satisfy the demand of noninvasive biomarkers. RT-qPCR was performed in two independent cohorts of serum samples from the training and validation phase of our study, to validate the lncRNAs identified in the tissue specimens. This stringent analysis led to identification of four significantly altered lncRNAs (BANCR, NR_026817, NR_029373 and NR_034119). In addition, ROC analysis further confirmed that expression of these lncRNAs considerably distinguished CRC patients from the controls, with high AUC values. Given the heterogeneous nature of CRC, a 4-lncRNA panel was established to further improve the diagnosis of this disease. The higher diagnostic accuracy in the training and validation phases, with AUC values of 0.891 and 0.881, respectively, indicated that combination of these lncRNAs may comprise a promising method for CRC detection. Furthermore, we directly compared the diagnostic efficiency of 4-lncRNA panel with previously established marker CEA, in the same cohort. The result clearly demonstrated the superiority of 4-lncRNA panel over CEA for CRC diagnosis, with better sensitivity, especially in early-stage tumors. Based on these findings, it seems that serum lncRNA panel may prove to be a much more sensitive method for CRC detection.

The four lncRNAs in the panel were significantly dysregulated in both tissue and serum sample according to our selection procedure. In addition, we examined the correlation between lncRNAs expressions of colorectal cancer tissue and serum. A significant correlation was observed, approving the power of the 4-lncRNA panel to reflect the condition of the solid tumor. Despite the reduced expression of NR_026817, NR_029373 and NR_034119 in CRC patients, all three lncRNAs have a steady expression in both tissue and serum samples. Moreover, we tested the stability of all four identified lncRNAs in harsh conditions. These lncRNAs could stably express even after their storage at room temperature or at −80 °C and repetitive freeze–thaw cycles, suggesting that this 4-lncRNA panel is fit for stably detection even under harsh conditions.

Considering one of the most urgent needs of clinicians, to find adequate predictive biomarker that could discriminate CRC patients with high risk and poor prognosis, we investigated the role of these four identified lncRNAs as prognostic biomarkers. The median was used as the cutoff as we just explore their potentiality for prognosis prediction in this study. Reduced expression of NR_029373 and NR_034119 lncRNAs was correlated with lower disease-specific survival rate. The Cox proportional hazards regression model analysis showed that serum expression level of NR_029373 and NR_034119 was independent factors for disease-specific survival rate of CRC patients, suggesting that they may be employed as biomarkers for CRC prognosis.

Although we have constructed a promising 4-lncRNA panel for CRC detection in serum, it is uncertain if this panel is only specific for CRC. Thus, additional studies will be required to further examine the expression changes of these four lncRNAs in other tumors. Despite demonstrating improved sensitivity and markedly higher AUC value of the 4-lncRNA panel over CEA, in distinguishing early-stage CRC patients from healthy people, the expression of these four lncRNAs has still not been investigated in patients with adenoma. Colorectal adenoma is accepted as the usual benign precursor lesion in the transformation to CRC. Thus, we need to explore the lncRNAs dysregulation in patients suffering from adenoma with various degree of dysplasia and to identify if it has some prognostic benefits to patients who are at the very early risk of having CRC.

In conclusion, we have successfully established a distinctive serum 4-lncRNA panel for CRC detection through stringent step-by-step selection procedures and identified that NR_029373 and NR_034119 can act as independent predictors of CRC-specific survival rate. Further studies, including large clinical samples and diverse ethnic populations, are required to confirm the usefulness of these lncRNAs, as noninvasive markers in CRC patients.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the Clinical Research Ethics Committee of Qilu Hospital, Shandong University and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Bailey JR, Aggarwal A, Imperiale TF (2016) Colorectal cancer screening: stool DNA and other noninvasive modalities. Gut Liver 10:204–211

Duffy MJ, van Rossum LG, van Turenhout ST, Malminiemi O, Sturgeon C, Lamerz R, Nicolini A, Haglund C, Holubec L, Fraser CG, Halloran SP (2011) Use of faecal markers in screening for colorectal neoplasia: a European group on tumor markers position paper. Int J Cancer 128:3–11

El-Tawdi AH, Matboli M, Shehata HH, Tash F, El-Khazragy N, Ael-S Azazy, Abdel-Rahman O (2016) Evaluation of circulatory RNA-based biomarker panel in hepatocellular carcinoma. Mol Diagn Ther 20:265–277

Feng J, Sun Y, Zhang EB, Lu XY, Jin SD, Guo RH (2015) A novel long noncoding RNA IRAIN regulates cell proliferation in non small cell lung cancer. Int J Clin Exp Pathol 8:12268–12275

Figueredo A, Coombes ME, Mukherjee S (2008) Adjuvant therapy for completely resected stage ‖ colon cancer. Cochrane Database Syst Rev (3):CD005390. doi:10.1002/14651858.CD005390.pub2

Ge X, Chen Y, Liao X, Liu D, Li F, Ruan H, Jia W (2013) Overexpression of long noncoding RNA PCAT-1 is a novel biomarker of poor prognosis in patients with colorectal cancer. Med Oncol 30:588

Gu M, Zheng A, Tu W, Zhao J, Li L, Li M, Han S, Hu X, Zhu J, Pan Y, Xu J, Yu Z (2016) circulating LncRNAs as novel, non-invasive biomarkers for prenatal detection of fetal congenital heart defects. Cell Physiol Biochem 38:1459–1471

Guo Q, Zhao Y, Chen J, Hu J, Wang S, Zhang D, Sun Y (2014) BRAF-activated long non-coding RNA contributes to colorectal cancer migration by inducing epithelial-mesenchymal transition. Oncol Lett 8:869–875

Health Quality Ontario (2009) Fecal occult blood test for colorectal cancer screening: an evidence-based analysis. Ont Health Technol Assess Ser 9:1–40

Itzkowitz SH, Jandorf L, Brand R, Rabeneck L, Schroy PC 3rd, Sontag S, Johnson D, Skoletsky J, Durkee K, Markowitz S, Shuber A (2007) Improved fecal DNA test for colorectal cancer screening. Clin Gastroenterol Hepatol 5:111–117

Ji Q, Zhang L, Liu X, Zhou L, Wang W, Han Z, Sui H, Tang Y, Wang Y, Liu N, Ren J, Hou F, Li Q (2014) Long non-coding RNA MALAT1 promotes tumour growth and metastasis in colorectal cancer through binding to SFPQ and releasing oncogene PTBP2 from SFPQ/PTBP2 complex. Br J Cancer 111:736–748

Kim DH, Pickhardt PJ, Taylor AJ, Leung WK, Winter TC, Hinshaw JL, Gopal DV, Reichelderfer M, Hsu RH, Pfau PR (2007) CT colonography versus colonoscopy for the detection of advanced neoplasia. N Engl J Med 357:1403–1412

Levin B, Lieberman DA, McFarland B, Andrews KS, Brooks D, Bond J, Dash C, Giardiello FM, Glick S, Johnson D, Johnson CD, Levin TR, Pickhardt PJ, Rex DK, Smith RA, Thorson A, Winawer SJ, American Cancer Society Colorectal Cancer Advisory Group, US Multi-Society TaskForce, American College of Radiology Colon Cancer Committee (2008) Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps 2008 a joint guideline from the American Cancer Society the US Multi-Society Task Force on Colorectal Cancer and the American College of Radiology. Gastroenterology 134:1570–1595

Li L, Zhang L, Zhang Y, Zhou F (2015) Increased expression of LncRNA BANCR is associated with clinical progression and poor prognosis in gastric cancer. Biomed Pharmacother 72:109–112

Liu Y, Zhang M, Liang L, Li J, Chen YX (2015a) Over-expression of lncRNA DANCR is associated with advanced tumor progression and poor prognosis in patients with colorectal cancer. Int J Clin Exp Pathol 8:11480–11484

Liu L, Liu Y, Zhuang C (2015b) Inducing cell growth arrest and apoptosis by silencing long non-coding RNA PCAT-1 in human bladder cancer. Tumour Biol 36:7685–7689

Ma C, Shi X, Zhu Q, Li Q, Liu Y, Yao Y, Song Y (2016) The growth arrest-specific transcript 5 (GAS5): a pivotal tumor suppressor long noncoding RNA in human cancers. Tumour Biol 37:1437–1444

Mandel JS (2008) Screening for colorectal cancer. Gastroenterol Clin North Am 37:97–115

Ni B, Yu X, Guo X, Fan X, Yang Z, Wu P et al (2015) Increased urothelial cancer associated 1 is associated with tumor proliferation and metastasis and predicts poor prognosis in colorectal cancer. Int J Oncol 47:1329–1338

Nie FQ, Ma S, Xie M, Liu YW, De W, Liu XH (2016) Decreased long noncoding RNA MIR31HG is correlated with poor prognosis and contributes to cell proliferation in gastric cancer. Tumour Biol 37:7693–7701

Ponting CP, Oliver PL, Reik W (2009) Evolution and functions of long noncoding RNAs. Cell 136:629–641

Qi P, Xu MD, Ni SJ, Huang D, Wei P, Tan C et al (2013) Low expression of LOC285194 is associated with poor prognosis in colorectal cancer. J Transl Med 11:122

Ren YK, Xiao Y, Wan XB, Zhao YZ, Li J, Li Y, Han GS, Chen XB, Zou QY, Wang GC, Lu CM, Xu YC, Wang YC (2015) Association of long non-coding RNA HOTTIP with progression and prognosis in colorectal cancer. Int J Clin Exp Pathol 8:11458–11463

Shi Y, Liu Y, Wang J, Jie D, Yun T, Li W, Yan L, Wang K, Feng J (2015) Downregulated long noncoding RNA BANCR promotes the proliferation of colorectal cancer cells via downreguation of p21 expression. PLoS ONE 10:e0122679

Siegel RL, Miller KD, Jemal A (2016) Cancer statistics, 2016. CA Cancer J Clin 66:7–30

Sun J, Ding C, Yang Z, Liu T, Zhang X, Zhao C, Wang J (2016a) The long non-coding RNA TUG1 indicates a poor prognosis for colorectal cancer and promotes metastasis by affecting epithelial-mesenchymal transition. J Transl Med 14:42

Sun Y, Zheng ZP, Li H, Zhang HQ, Ma FQ (2016b) ANRIL is associated with the survival rate of patients with colorectal cancer, and affects cell migration and invasion in vitro. Mol Med Rep. doi:10.3892/mmr.2016.5409

Tong YS, Zhou XL, Wang XW, Wu QQ, Yang TX, Lv J et al (2014) Association of decreased expression of long non-coding RNA LOC285194 with chemoradiotherapy resistance and poor prognosis in esophageal squamous cell carcinoma. J Transl Med 12:233

Tong YS, Wang XW, Zhou XL, Liu ZH, Yang TX, Shi WH, Xie HW, Lv J, Wu QQ, Cao XF (2015) Identification of the long non-coding RNA POU3F3 in plasma as a novel biomarker for diagnosis of esophageal squamous cell carcinoma. Mol Cancer 14:3

van Werven FJ, Neuert G, Hendrick N, Lardenois A, Buratowski S, van Oudenaarden A, Primig M, Amon A (2012) Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. Cell 150:1170–1181

Wang YX, Li ZH, Zheng SY, Zhou Y, Zhao L, Ye H, Zhao X, Gao W, Fu Z, Zhou Q, Liu Y, Chen R (2015) Expression profile of long non-coding RNAs in pancreatic cancer and their clinical significance as biomarkers. Oncotarget 6:35684–35698

Wang WT, Sun YM, Huang W, He B, Zhao YN, Chen YQ (2016a) Genome-wide long non-coding RNA analysis identified circulating LncRNAs as novel non-invasive diagnostic biomarkers for gynecological disease. Sci Rep 6:23343

Wang Y, Xue D, Li Y, Pan X, Zhang X, Kuang B et al (2016b) The long noncoding RNA MALAT-1 is a novel biomarker in various cancers: a meta-analysis based on the GEO database and literature. J Cancer 7:991–1001

Wu Y, Wang YQ, Weng WW, Zhang QY, Yang XQ, Gan HL, Yang YS, Zhang PP, Sun MH, Xu MD, Wang CF (2016) A serum-circulating long noncoding RNA signature can discriminate between patients with clear cell renal cell carcinoma and healthy controls. Oncogenesis 5:e19

Xu Y, Qiu M, Chen Y, Wang J, Xia W, Mao Q, Yang L, Li M, Jiang F, Xu L, Yin R (2016) Long noncoding RNA, tissue differentiation-inducing nonprotein coding RNA is upregulated and promotes development of esophageal squamous cell carcinoma. Dis Esophagus. doi:10.1111/dote.12436

Yang MH, Hu ZY, Xu C, Xie LY, Wang XY, Chen SY, Li ZG (2015) MALAT1 promotes colorectal cancer cell proliferation/migration/invasion via PRKA kinase anchor protein 9. Biochim Biophys Acta 1852:166–174

Yang F, Liu YH, Dong SY, Ma RM, Bhandari A, Zhang XH, Wang OC (2016) A novel long non-coding RNA FGF14-AS2 is correlated with progression and prognosis in breast cancer. Biochem Biophys Res Commun 470:479–483

Yao W, Bai Y, Li Y, Guo L, Zeng P, Wang Y et al (2016) Upregulation of MALAT-1 and its association with survival rate and the effect on cell cycle and migration in patients with esophageal squamous cell carcinoma. Tumour Biol 37:4305–4312

Yin D, He X, Zhang E, Kong R, De W, Zhang Z (2014) Long non-coding RNA GAS5 affects cell proliferation and predicts a poor prognosis in patients with colorectal cancer. Med Oncol 31:253

Yin DD, Liu ZJ, Zhang E, Kong R, Zhang ZH, Guo RH (2015) Decreased expression of long noncoding RNA MEG3 affects cell proliferation and predicts a poor prognosis in patients with colorectal cancer. Tumour Biol 36:4851–4859

Zhao B, Hou X, Zhan H (2015) Long non-coding RNA PCAT-1 overexpression promotes proliferation and metastasis in non-small cell lung cancer cells. Int J Clin Exp Med 8:18482–18487

Zhong W, Yu Z, Zhan J, Yu T, Lin Y, Xia ZS, Yuan YH, Chen QK (2015) Association of serum levels of CEA, CA199, CA125, CYFRA21-1 and CA72-4 and disease characteristics in colorectal cancer. Pathol Oncol Res 21:83–95

Zhou X, Yin C, Dang Y, Ye F, Zhang G (2015) Identification of the long non-coding RNA H19 in plasma as a novel biomarker for diagnosis of gastric cancer. Sci Rep 5:11516