Int J Clin Exp Pathol 2017;10(5):5227-5239
www.ijcep.com /ISSN:1936-2625/IJCEP0047640

Original Article

Serum long non-coding RNA signatures serve as novel noninvasive biomarkers for diagnosis and prognosis of gastric cancer

Juan Ding[1], Chao Che[2], Yongyuan Liang[1], Yifeng Sun[1], Qian Wang[1]

Departments of [1]Clinical Laboratory, [2]Pediatrics, Qilu Hospital, Shandong University, Jinan, Shandong Province, China

Received December 29, 2016; Accepted March 9, 2017; Epub May 1, 2017; Published May 15, 2017

Abstract: Gastric cancer (GC) is prevalent worldwide, and improvements in timely and effective diagnosis are imperatively needed. Recent advantages of cell free long non-coding RNAs (lncRNAs) open a new realm of possibilities for noninvasive diagnosis and prognosis of GC. The aim of this study was to identify potential lncRNA expression signatures for diagnosis of patients with GC along with prognostic prediction. We performed genome-wide serum lncRNA analysis by Hiseq sequencing followed by evaluations in the training and validation sets with reverse transcription quantitative real-time PCR assays from serum samples of 230 patients with GC and 230 controls. Four lncRNAs (XIST, LOC100506474, UCA1 and LINC00467) were identified to be significantly dysregulated in above serum samples, and a panel was finally developed by multivariate logistic regression model with an area under the receiver operating characteristic curve of 0.888 on validation cohort. The corresponding AUCs of the panel for patients with TNM stage I, II and III were 0.784, 0.851 and 0.931, respectively, significantly higher than that of CA19-9. Kaplan-Meier analysis showed that patients with high levels of XIST and LOC100506474 had worse recurrence-free survival (P=0.008 and 0.019, respectively). The multivariate Cox analysis demonstrated that XIST and LOC100506474 were both independently associated with tumor recurrence of GC (P=0.003 and 0.010, respectively). In conclusion, our study established a serum lncRNA panel with considerable clinical values in predicting and providing prognostic information for GC and identified XIST and LOC100506474 as potential biomarkers that can provide information on the recurrence risk of GC.

Keywords: Gastric cancer, long non-coding RNA, noninvasive biomarkers, diagnosis, prognosis

Introduction

Gastric cancer is the most general malignant cancer in digestive system, which has high morbidity and mortality; and it has been estimated that 360,000 individuals die of GC each year in China [1]. 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. The dismal outcome of this disease has aroused the attention to the critical importance of early detection. Tumor circulation biomarkers for gastric cancer are indispensable for minimally invasive diagnosis in the early stage to reduce gastric cancer mortality rates. However, the currently known tumor antigens in serum, such as CA72-4, CEA, and CA19-9, have relatively poor sensitivity in GC diagnosis [2, 3]. In contrast, gastroscopy examination and gastroscopy-guided biopsy for histological evaluation can offer high diagnostic accuracy, but it is invasive and inconvenient, which limits its use for general cancer screening. Thus, there is an urgent need to develop a new noninvasive, sensitive and cost-effective method to complement and improve the current gastric cancer screening strategies.

Long noncoding RNAs (lncRNAs) are most commonly defined as the RNA transcripts of more than 200 nucleotides (nt) and located in nuclear or cytosolic fractions with no protein-coding capacity [4]. Altered expression of several lncRNAs has recently been attributed to pathogenesis of some malignant neoplasia, including gastric cancer [5, 6]. The discovery and study of lncRNAs is thus of major relevance to human biology and disease, as they represent an extensive, largely unexplored, and functional

Table 1. Characteristics of study participants in training set and validation set

Variables	Training set (n=240)	Validation set (n=220)
Control (number)	120	110
Gender		
Male	76	68
Female	44	42
Age		
≤60 years	51	50
>60 years	69	60
Tumor (number)	120	110
Gender		
Male	76	68
Female	44	42
Age		
≤60 years	51	50
>60 years	69	60
Size		
<6 cm	83	85
≥6 cm	37	25
Differentiation		
Well	17	12
Moderate	90	86
Poor	13	12
Local invasion		
T1-T2	36	31
T3-T4	84	79
Lymph node metastasis		
No	52	48
Yes	68	62
TNM stage		
I	27	21
II	35	40
III	58	49

component of the genome. Aberrant lncRNA expression has been detected in tissues sections and identified as promising biomarkers for diagnosis and prognosis in breast cancer, colorectal cancer, non-small cell lung cancer (NSCLC) and GC [7-10].

Predictive biomarkers are better blood-based, as blood is easily available and provides the chance to monitor cancer progression. 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. One or a cluster of specific marker is urgently needed for increasing the early detection rate and decreasing the mortality rate in GC [11]. Therefore, it is important to identify blood markers that predict the initiation and progression of GC, which may allow for the development of early detection rate of GC.

In this study, we conducted high-throughput Hiseq sequencing followed by reverse transcription quantitative real-time PCR (RT-QPCR) assays to test the hypothesis that specific lncRNAs can be useful in improving the diagnostic and prognostic efficiency of gastric cancer. In addition, we also assessed the correlation between the expression level of lncRNAs identified and the recurrence-free survival rate of GC patients, to explore their potential for prognostic prediction.

Materials and 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 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 GC, 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, serum samples pooled from six patients with GC and six healthy donors were subjected to Hiseq sequencing, to identify lncRNAs significantly differentially expressed. Candidate lncRNAs were then selected according to the result of the analysis and previous studies [12-18]. In the training phase, the candidate lncRNAs were firstly tested with RT-QPCR in an independent cohort of serum samples from 40 GC patients with GC and 40 controls. Subsequently, lncRNAs differentially expressed among

Table 2. The information of GC patients and healthy controls in Hiseq sequencing set [median (interquartile range)]

Variable	GC (n=6)	Healthy controls (n=6)
Age (years)	65 (54-77)	65 (56-71)
Sex		
Male	4 (66.7%)	4 (66.7%)
Female	2 (33.3%)	2 (33.3%)
Size		
<6 cm	3 (50.0%)	
≥6 cm	3 (50.0%)	
Differentiation		
Well	1 (16.7%)	
Moderate	4 (66.7%)	
Poor	1 (16.7%)	
Local invasion		
T1-T2	2 (33.3%)	
T3-T4	4 (66.7%)	
Lymph node metastasis		
Negative	3 (50.0%)	
Positive	3 (50.0%)	
TNM stage		
I	1 (16.7%)	
II	2 (33.3%)	
III	3 (50.0%)	
IV	0 (0)	

the two groups were further tested in 80 GC patients and 80 controls. These 120 GC patients and 120 healthy donors were used to construct the diagnostic lncRNA panel based on the logistic regression model for the differentiation between the GC group 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 220 patients (110 GC patients and 110 controls) to validate its diagnostic performance.

Patients and control subjects

Patients with GC were recruited from Qilu Hospital of Shandong University between 2009 and 2014. Serum samples were collected before any anticancer treatments such as surgery, chemotherapy and radiotherapy were given. All GC 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), and tumor grade was defined according to the WHO 2004 grading scheme. Control participants without history of GC were recruited from a large pool of individuals seeking a routine health checkup at the Healthy Physical Examination Centre of Qilu Hospital, Shandong University. People who showed no evidence of disease were selected as tumor-free controls. The demographic and clinical features of participants in training phase and validation phase are summarized in **Table 1**.

The correlation between the expression of identified lncRNAs and survival rate of GC patients was assessed to explore their potential as predictors for GC prognosis. Recurrence free survival (RFS) was defined as the time from inclusion to recurrence or metastasis progression. GC patients in the validation phase were followed up at intervals of 3 months during the first 2 years and 6 months up to the fifth year, and the date of the latest record retrieved was October 31, 2016. Owing to incomplete follow-ups, 14 of all the cases were excluded from the cohort, and the median follow-up time was 53 months (range, 16-69 months).

Serum preparation

Venous blood samples were collected by vena puncture from each participant before any treatment. Serum was separated via centrifugation at 1,600 g for 10 min at 4°C within 2 h of collection, followed by a second centrifugation at 16,000 g for 10 min at 4°C to eliminate residual cells debris. Supernatant serum was then stored at -80°C till use.

Hiseq sequencing

Serum samples pooled from six patients with GC and six healthy donors were subjected to Hiseq sequencing, to identify lncRNAs significantly differentially expressed. Demographic and clinical features of the GC patients and controls are summarized in **Table 2**. Total serum RNA was extracted by one-step extraction using a Trizol kit (Life Technologies, USA), and the purity and quantity of RNA were determined by UV spectrophotometry. cDNA library construction and sequencing were performed according to previously described methods [19]. Briefly, after extraction of total RNA, ribosomal RNA was separated to isolate as ncRNA as possible. RNA containing poly(A) was then removed. RNA fragments were broken into short fragments randomly. The first chain of cDNA was

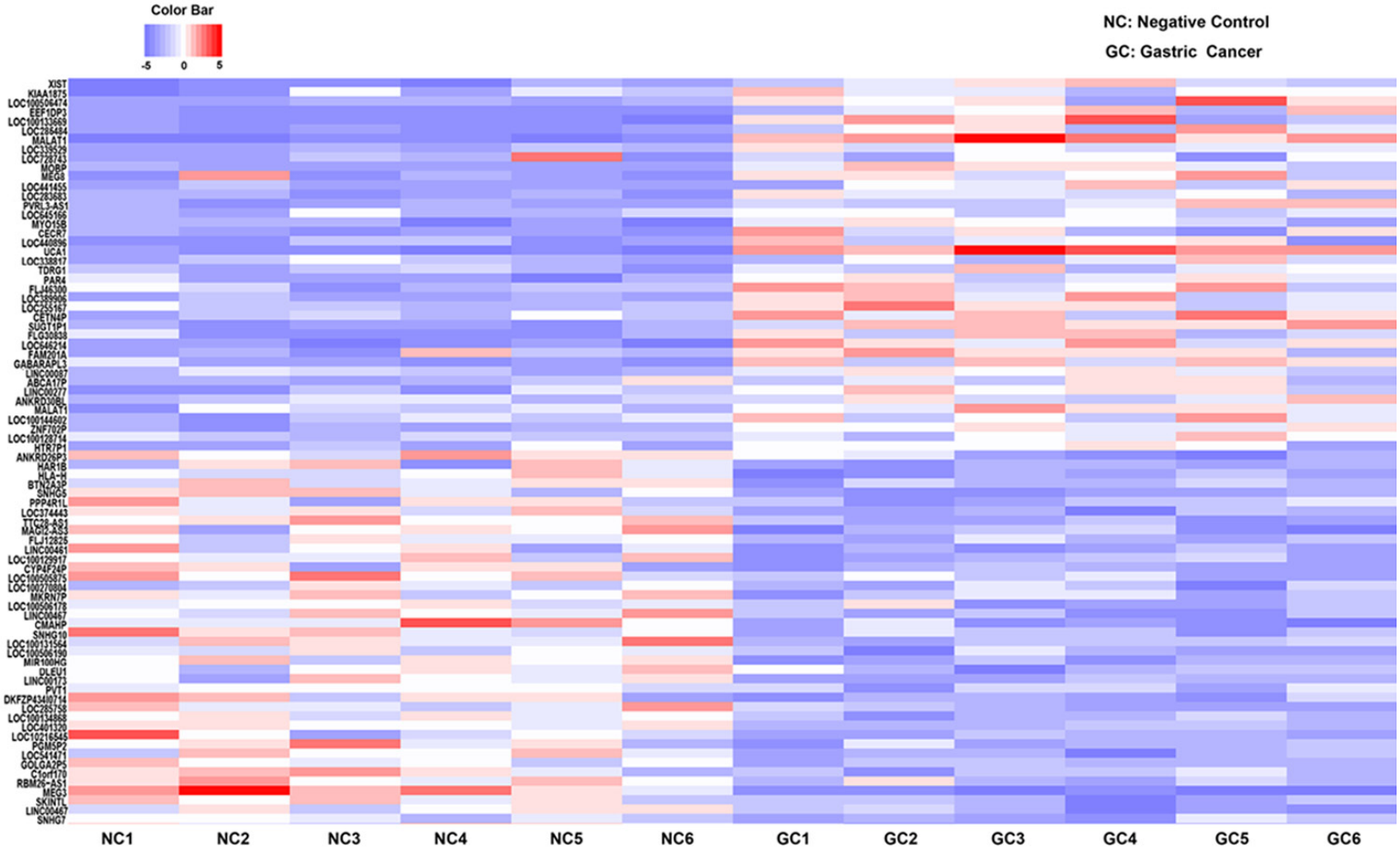

Figure 1. The heat map shows expression of the 80 lncRNAs most up- or down-regulated in serum from GC patients compared with healthy controls. The top 40 lncRNAs up- and down-regulated in GC are shown in the top and bottom halves, respectively. The heat map was generated with an R package using normalization across rows.

Table 3. Candidate lncRNAs selected on a basis of the Hiseq analysis

Seqname	Location	Regulation (GC vs NC)	Fold change	*P* value
XIST	ChrXq13.2	Up	49.7132126	0.00012778
LOC100506474	Chr2p24.3	Up	32.1463807	0.00100932
UCA1	Chr19p13.12	Up	21.6289351	0.00927686
LINC00467	Chr1q32.3	Down	43.8351239	0.00039420
MEG3	Chr14q32.2	Down	36.9371451	0.00084693
LOC100129917	Chr4p16.3	Down	16.1247927	0.01928476

GC: Gastric cancer; NC: Negative Control.

Table 4. Expression of 10 candidate lncRNAs in GC patients and healthy controls [median (interquartile range)]

LncRNA	Healthy controls	GC patients	*P* value
XIST	1.37 (0.41-2.99)	2.43 (1.18-3.72)	<0.01
LOC100506474	0.74 (0.45-1.93)	1.51 (0.32-2.11)	<0.01
UCA1	0.66 (0.35-1.82)	1.42 (0.25-1.93)	<0.01
HOTAIR	1.01 (0.30-2.53)	1.44 (0.56-3.08)	0.23
H19	0.72 (0.16-2.23)	1.39 (0.63-2.67)	<0.05
LINC00467	1.58 (0.41-2.49)	0.81 (0.27-1.83)	<0.01
MEG3	1.22 (0.33-2.41)	0.69 (0.36-1.84)	<0.05
LOC100129917	0.89 (0.31-2.77)	0.63 (0.24-1.68)	0.09
CASC2	0.81 (0.23-1.69)	1.27 (0.46-2.95)	0.07
GAS5	1.06 (0.48-1.63)	1.10 (0.60-2.34)	0.32

generated using RNA fragments as templates and 6-bp random primers. Second chain of the cDNA was synthesized according to the kit's instruction (TakaRa Co., Ltd., Dalian, China). After purification, end repair, base A and sequencing joint adding, the generated cDNA was fragmented using uracil-N-glycosylase (UNG). cDNA fragments were chosen according to size, then PCR amplification was performed to establish the complete sequencing cDNA library. LncRNAs were sequenced using the high-throughput, high-sensitivity HiSeq 2500 sequencing platform (Illumina Company, San Diego, USA). Sequencing results were analyzed and treated using Trim Galore software to dynamically remove joint sequence fragments and low-quality segments from the 3' end. FastQC software was used for quality control of the pretreated data.

Quantification of lncRNAs by RT-qPCR analysis

Total RNA from serum samples were extracted using TRIzol LS reagents (Invitrogen, Carlsbad, CA), according to the manufacturer's protocol. 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_2O 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 Inc, Dalian). 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.

Statistical analysis

Kolmogorov-Smirnov test was used to determine the distribution of the samples of each group. Data were presented as median (interquartile range). Nonparametric Mann-Whitney U-tests were performed to compare the differences in concentrations of serum lncRNAs between the GC group and the control group. Receiver operating characteristic (ROC) curves were established to discriminate the patients with GC from controls. Area under the receiver operating characteristic curve (AUC) was employed as an accuracy index for evaluating the diagnostic performance of the selected lncRNA panel [20]. Kaplan-Meier method was utilized to estimate survival curves, and the log-rank test was used to make comparisons.

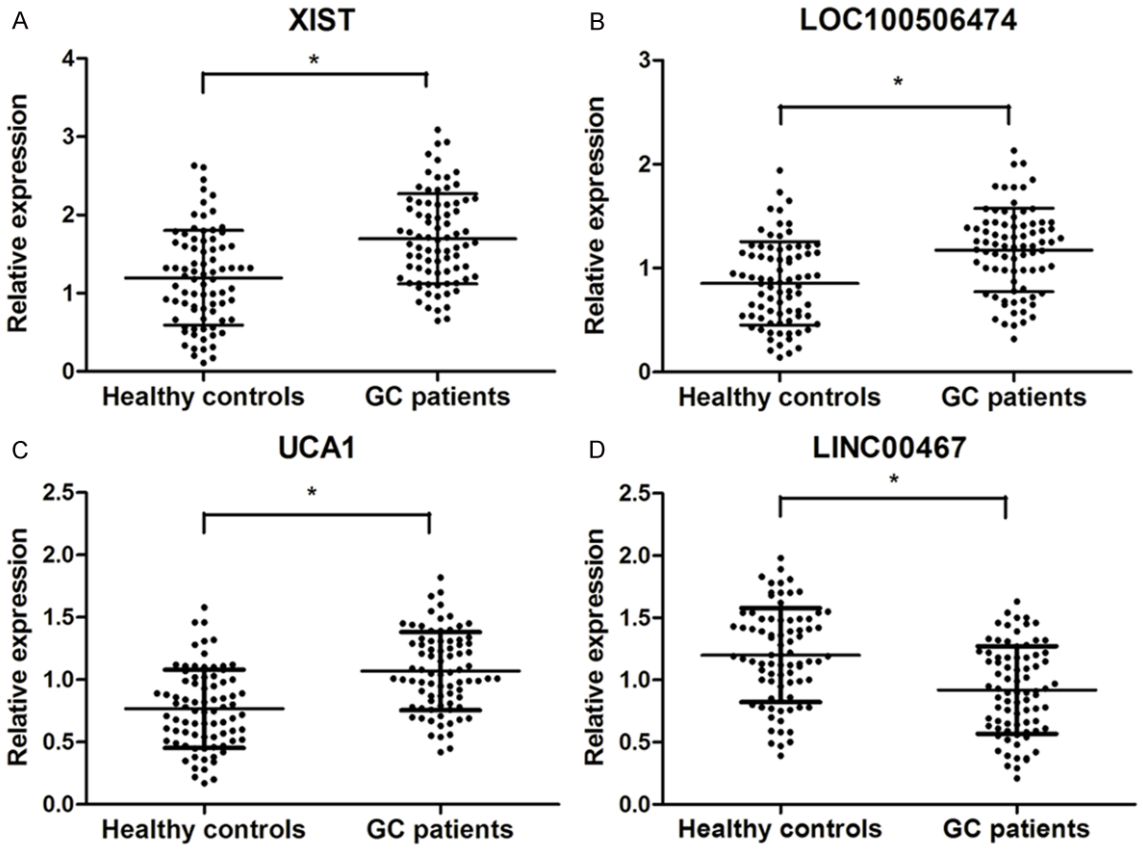

Figure 2. Concentrations of four identified serum lncRNAs in patients with GC (n=80) and control individuals (n=80) using RT-qPCR assay in training set (A-D), *P<0.001.

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 and others were calculated using SPSS version 19.0 software (SPSS, Chicago, IL). A *P*-value of <0.05 was considered statistically significant.

Results

Identification of candidate lncRNAs by high-throughput Hiseq sequencing

The Hiseq sequencing with six serum samples pooled from GC patient and six from healthy donors were conducted. In total, 418 lncRNAs were identified with significant differential expression (fold change ≥2.0). To identify the lncRNAs that were potential biomarkers, we concentrated on the top 40 most up- and down-regulated lncRNAs that were differentially expressed between GC patients and healthy donors (**Figure 1**). Starting from those lncRNAs with the greatest fold changes, we filtered appropriate candidate lncRNAs in descending order. Candidates should be plausible for primer designing, and only those have steady expressions in serum samples were selected. Finally, we chose three candidate lncRNAs from the up-regulated group and three from the down-regulated group as well (**Table 3**). Another four lncRNAs were also tested by RT-qPCR because they had been shown dysregulation in GC patients [12-18]. Thus, 10 lncRNAs were selected as candidates for further testing via RT-qPCR.

Evaluation of selected lncRNA expression by RT-qPCR

The expression of all 10 candidate lncRNAs was firstly evaluated by RT-qPCR using serum samples from 40 patients with GC and 40 controls in the training phase. Among these, six lncRNAs (XIST, LOC100506474, UCA1, H19, LINC00467 and MEG3) were found significantly dysregulated in GC patients (**Table 4**). Subsequ-

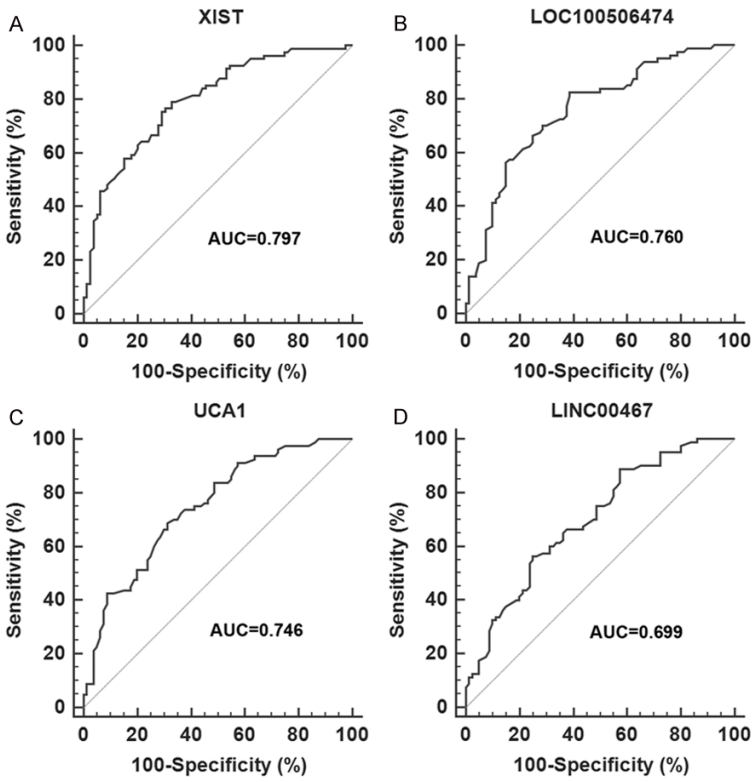

Figure 3. ROC curve analysis for the detection of GC using XIST (A), LOC100506474 (B), UCA1 (C) and LINC00467 (D) in patients with GC (n=80) and control individuals (n=80) in training set.

ently, these six lncRNAs were further validated in an independent cohort of 160 serum samples from 80 GC patients and 80 healthy individuals. Among the six candidate lncRNAs, four lncRNAs were finally identified to have differential expression patterns between GC group and control group. Three lncRNAs (XIST, LOC100506474 and UCA1) were upregulated and one (LINC00467) was downregulated (**Figure 2A-D**). The corresponding AUCs of the four lncRNAs (XIST, LOC100506474, UCA1 and LINC00467) were 0.797, 0.760, 0.746 and 0.699, respectively (**Figure 3A-D**).

Establishment of the predictive lncRNA panel

A stepwise logistic model was constructed for GC diagnosis using the 240 serum samples enrolled in the training phase. The predicted probability of being diagnosed with GC from the model based on the 4-lncRNA panel was calculated using the equation as follows: Logit(P)=-1.0396-0.7424 × XIST - 0.5625 × LOC100506474 - 0.6297 × UCA1 + 0.3954 × LINC00467. ROC analysis was used to evaluate the diagnostic performance of the established lncRNA panel. The AUC for the 4-lncRNA panel was 0.886 (95% confidence interval [CI]=0.827-0.931) and the optimal cut-off value was -2.84, providing a sensitivity of 83.95% and a specificity of 81.01% (**Figure 4A**).

Validation of the lncRNA panel

The parameters estimated from the training set were used in a blind fashion to predict the probability of being diagnosed with GC for the independent validation data set. Using the classification threshold score of <-2.84 derived above, 125 samples were identified as GC patients and 95 samples were identified as healthy individuals. After unblinding 81 of the 110 healthy controls [specificity, 73.64% (95% CI, 62.7 to 83.0)] and 96 of the 110 GC patients [sensitivity, 87.27% (95% CI, 78.2 to 93.8)] were correctly identified resulting in an AUC of 0.888 (95% CI, 0.828 to 0.932, **Figure 4B**), which was significantly better than that of CA19-9 (AUC: 0.704, 95% CI=0.621-0.778, sensitivity =74.29% and specificity =58.57%, P<0.001, **Figure 4C**).

Furthermore, we then compared the diagnostic performance of this 4-lncRNA panel with CA19-9, in discriminating GC patients from control individuals at different TNM stages on validation set. The AUCs of the 4-lncRNA panel for patients with TNM stage I, II and III were 0.784, 0.851 and 0.931, respectively (**Figure 4D-F**), and were all higher than those of CA19-9, which were 0.614, 0674 and 0.823, respectively (**Figure 4G-I**).

Correlation between the four lncRNAs and clinicopathological characteristics

The data summarized in **Table 5** show the relationship between the four lncRNAs and the clinicopathological characteristics of the patients with GC in the validation set. Higher levels of serum XIST, LOC100506474, UCA1 and

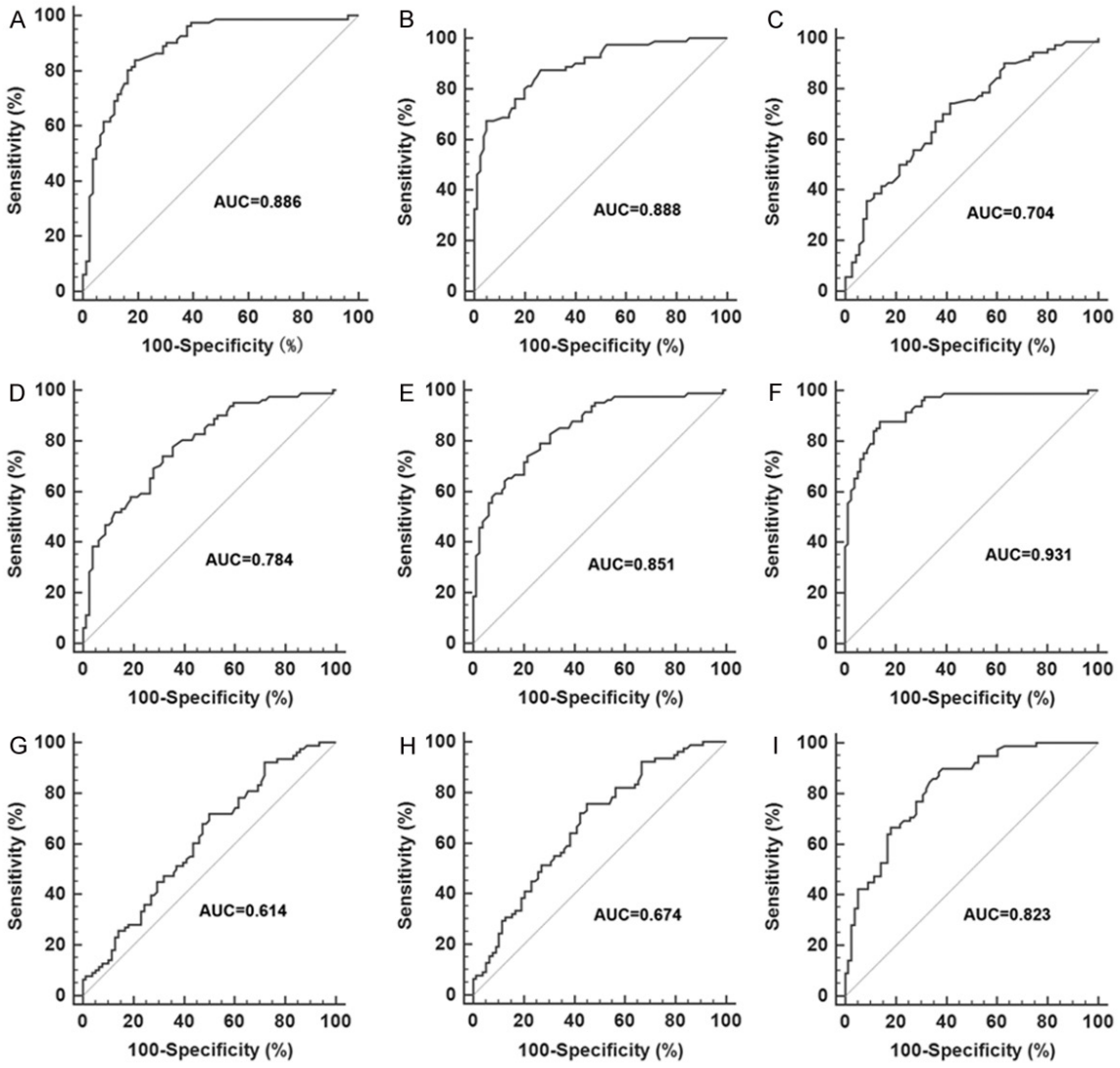

Figure 4. (A, B) ROC curves for the detection of GC using 4-lncRNA panel in training set (A) and validation set (B); (C) ROC curve analysis using CA19-9 for the detection of GC in validation set; (D-F) ROC curves using the 4-lncRNA panel for the detection of GC patients with TNM stage I (D), II (E) and III (F) in validation set; (G-I) ROC curve analysis using CA19-9 for the detection of GC stage I (G), II (H) and III (I) in validation set.

lower levels of LINC00467 significantly correlated with advanced TNM stage (all at $P<0.05$) and lymph node metastasis (all at $P<0.01$). Higher levels of UCA1 correlated with poorer tumor differentiation ($P<0.05$). However, no significant associations were found between the four lncRNAs with age, gender, tumor location, size or local invasion (all at $P\geq0.05$).

Correlation between lncRNAs expression levels and patient recurrence

Survival analysis has been finally carried on 96 patients on validation set since 14 patients were lost to follow up. Kaplan-Meier survival analysis revealed that GC patients with high XIST and LOC100506474 expression levels showed significantly reduced RFS than those with low XIST and LOC100506474 levels ($P=0.008$ and $P=0.019$, respectively) (**Figure 5**). Univariate Cox proportional hazards regression model analysis revealed that RFS was significantly correlated with XIST level ($P<0.001$), LOC100506474 level ($P=0.009$), tumor stage ($P=0.010$) and lymph node status ($P=0.037$). Parameters significantly related to RFS in the univariate analysis were put into the multivariate analysis to identify independent factors for

Table 5. Correlation between serum lncRNA concentrations and clinicopathological characteristics of patients with GC in validation set [median (interquartile range)]

Parameters	Total cases	XIST expression	P	LOC100506474 expression	P	UCA1 expression	*P*	LINC00467 expression	P
Sex			0.77		0.43		0.96		0.49
Male	64	1.75 (1.23-1.90)		1.49 (1.01-1.70)		1.34 (1.16-1.89)		0.83 (0.59-1.28)	
Female	46	1.73 (1.33-1.86)		1.25 (1.05-1.32)		1.36 (1.26-1.89)		0.79 (0.51-1.11)	
Age			0.57		0.20		0.51		0.38
≤60 years	55	1.77 (1.23-1.90)		1.49 (1.03-1.70)		1.44 (1.05-1.90)		0.75 (0.64-1.32)	
>60 years	55	1.71 (1.22-1.82)		1.36 (1.05-1.31)		1.34 (1.14-1.88)		0.81 (0.59-1.44)	
Tumor size			0.28		0.96		0.37		0.13
<6 cm	72	1.73 (1.22-1.87)		1.38 (1.01-1.62)		1.34 (1.04-1.90)		0.73 (0.58-1.14)	
≥6 cm	38	1.76 (1.24-1.96)		1.39 (1.06-1.66)		1.47 (1.03-1.87)		0.80 (0.62-1.09)	
Differentiation			0.61		0.75		<0.05		0.07
Poor	21	1.75 (143-1.84)		1.42 (1.02-1.74)		1.67 (1.18-2.04)		0.76 (0.55-1.19)	
Moderate	77	1.74 (1.33-1.87)		1.37 (1.05-1.71)		1.34 (1.03-1.87)		0.70 (0.62-0.92)	
Well	12	1.63 (1.22-1.93)		1.39 (0.92-1.78)		1.43 (1.17-1.99)		0.84 (0.60-1.08)	
Local invasion			0.65		0.87		0.09		0.09
T1-T2	44	1.75 (1.36-1.90)		1.41 (0.97-1.77)		1.54 (1.03-1.96)		0.87 (0.62-1.10)	
T3-T4	66	1.74 (1.43-1.96)		1.40 (1.09-1.86)		1.41 (1.04-0.88)		0.80 (0.59-0.99)	
Lymph node metastasis			<0.01		<0.01		<0.01		<0.01
No	62	1.57 (1.51-1.85)		1.19 (0.90-1.65)		1.22 (1.08-1.94)		0.84 (0.68-1.24)	
Yes	48	1.94 (1.64-2.26)		1.56 (1.13-2.13)		1.83 (1.21-2.25)		0.63 (0.53-0.83)	
TNM stage			<0.01		<0.01		0.02		<0.01
I	16	1.57 (1.31-1.75)		1.06 (0.50-1.47)		1.34 (1.06-1.99)		0.88 (0.69-1.26)	
II	41	1.82 (1.68-1.97)		1.35 (0.91-1.60)		1.29 (1.09-1.91)		0.81 (0.67-1.17)	
III	53	1.87 (1.70-1.98)		1.78 (1.13-1.94)		1.52 (1.26-2.13)		0.74 (0.53-0.93)	

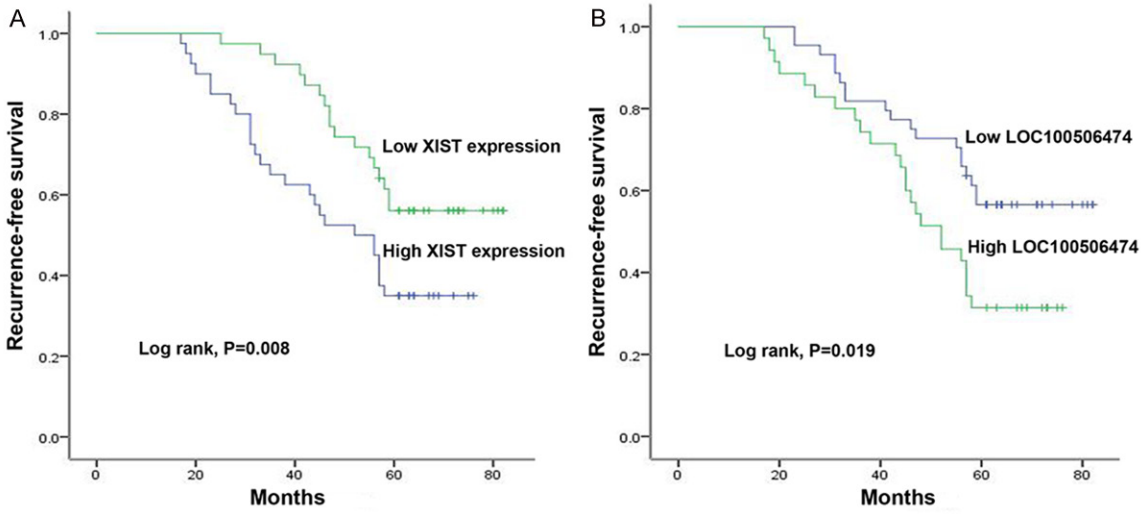

Figure 5. Kaplan-Meier curves for recurrence-free survival rate according to the serum levels of XIST (A) and LOC100506474 (B) in patients with GC in validation set.

prognoses. The results showed that XIST level (P=0.003), LOC100506474 level (P=0.010) and tumor stage (P=0.008) maintained their significance as independent prognostic factors for RFS of GC (**Table 6**).

Discussion

LncRNAs represent a novel class of gene regulators in cancer [21, 22]. They are involved in a variety of tumorigenesis process such as cell

Table 6. Univariate and multivariate Cox proportional hazards regression model analysis of factors for RFS in patients with GC in validation cohort

Characteristics	Univariate analysis			Multivariate analysis		
	HR	95% CI	*P* value	HR	95% CI	*P* value
Gender	1.023	0.617-2.012	0.417			
Age	1.492	0.709-2.716	0.215			
Tumor size	1.405	0.551-2.538	0.369			
Differentiation	1.848	1.036-3.297	0.092			
Lymphnode metastasis	2.264	1.146-3.683	0.037	2.185	1.007-3.875	0.059
Local Invasion	1.683	0.892-2.883	0.178			
TNM stage	2.538	1.432-3.875	0.010	2.568	1.243-5.102	0.008
XIST level	1.694	0.521-2.686	<0.001	1.590	0.424-2.598	0.003
LOC100506474 level	1.337	0.382-1.937	0.009	1.342	0.401-1.993	0.010
UCA1 level	1.412	0.293-1.885	0.066			
LINC00467 level	0.683	0.112-1.308	0.239			

proliferation [23], invasion [24] and apoptosis [25] by acting as tumor suppressors or oncogenes. The aberrant expressions of specific lncRNAs in cancer can mark the spectrum of disease progression and may serve as independent biomarkers for diagnosis and prognosis [26]. In GC, identification of noninvasive and invasive phenotypes is vital to rational clinical management [27]. Previously, Ren et al identified a panel of lncRNAs from GC tissues that seemed promising in predicting GC [28]. Yet, little was known about noninvasive lncRNA biomarkers that can effectively accomplish this task.

In our study, Hiseq sequencing was firstly employed to provide basic information of lncRNAs significantly dysregulated in serum samples of GC patients. Candidate lncRNAs were selected, compiled the Hiseq sequencing result and previous studies, and then evaluated by RT-qPCR in serum samples to validate their consistent pattern of dysregulation in these clinical materials. Four lncRNAs (XIST, LOC100506474, UCA1 and LINC00467), which showed considerable discriminating potential to identify GC 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 GC patients with higher accuracy in comparison with traditional diagnostic biomarker like CA-19-9. In addition, we also identified LOC100-506474 and UCA1 as independent factors for GC patient recurrence. Thus, a serum four-lncRNA panel was finally identified in our study, which can serve as noninvasive biomarkers for diagnosis and prognosis of GC.

It is widely accepted that the effective ways to improve the diagnosis and the prognosis of cancer patients are early detection and early treatment [29]. Despite the advances in diagnostic method, such as fecal occult blood testing and stool DNA test, early diagnosis for GC still remains difficult and the overall survival rate of GC patients has not changed dramatically [30]. Therefore, it is very important to search for cell-free markers to improve GC management. Utilizing lncRNA expression level in peripheral blood to diagnose tumors early is effective and deserves to be explored further because lncRNA is very stable in blood plasma and serum. The previous studies on searching for serum lncRNA based cancer biomarkers generally focused on individual cancer-specific lncRNAs [31, 32]. However, a single lncRNA may not be a reliable tumor biomarker because of the complex pathogenesis during the initiation and development of a severe malignancy. Simultaneous assessment of a panel of tumor-specific lncRNAs in serum may improve the sensitivity and specificity for cancer diagnosis and prognosis [33]. In our study, we screened the whole lncRNA profile in both GC and control serum samples via Hiseq sequencing, which enabled us to have better chance to identify potential diagnostic biomarkers. Hiseq sequencing is a high-throughput assay to initially screen lncRNAs and could exclude possible contamination by other small RNA and DNA fragments. However, the Hiseq results from

pooled serum samples might include inaccurate information owing to the individual variation. For this reason, candidate lncRNAs revealed by Hiseq sequencing were evaluated by two phases of RT-qPCR assays using a large number of individual samples. Finally, a four-lncRNA panel from the logistic model was identified for the predicting of GC. The high diagnostic accuracy in the training and validation set indicated that the expression profile of the four lncRNAs could serve as an accurate biomarker for the detection of GC. In addition, we performed a direct comparison of our results with traditional marker CA19-9 in the same cohort. Our data clearly demonstrated that the panel can more effectively discriminate the patients with GC from controls with better sensitivity than CA19-9, especially in early stage tumors. Based on these findings, the serum lncRNA panel provides a much more sensitive detection of GC. Furthermore, technically speaking, serum test is more convenient and noninvasive, and thus being an ideal for the investigation of a panel containing a small number of lncRNAs.

Functional studies of lncRNAs in tumor tissue may be helpful for evaluating serum lncRNAs as biomarkers for various types of cancer. Among four lncRNAs revealed in this study, some are reported to be involved in genesis and development of GC. Ma et al demonstrated that XIST promotes cell growth and invasion through regulating miR-497/MACC1 axis in gastric cancer [34]. Chen et al also found that XIST is up-regulated and is associated with aggressive tumor phenotypes and patient survival in gastric cancer by acting as a molecular sponge of miR-101 to modulate EZH2 expression [35]. Our date indicates that XIST is upregulated and serve as a oncogene in GC patients, which is consistent with the previous research. The UCA1 lncRNA has been widely reported to be involved in tumorigenesis of GC. Shang et al showed that silence of UCA1 inhibits malignant proliferation and chemotherapy resistance to adriamycin in gastric cancer [36]. In study by Gao et al, it was demonstrated that UCA1 may be a novel diagnostic and predictive biomarker in plasma for early gastric cancer [37]. There have been no functional studies about the role of LINC-00467 in development of GC, but LINC00467 has been found to be downregulated by N-Myc and promote cell survival in neuroblastoma [38]. Expression level of LOC100506474 was firstly reported in our study.

Considering one of the most urgent needs of clinicians, to find adequate predictive biomarker that could discriminate GC patients with high risk and poor prognosis, we investigated the role of these four-lncRNA panel as prognostic biomarkers. GC patients with high LOC100506474 and UCA1 expression levels showed significantly reduced RFS than those with low LOC100506474 and UCA1 levels. The Cox proportional hazards regression model analysis showed that serum expression level of LOC100506474 and UCA1 was independent factors for recurrence-free survival rate of GC patients, suggesting that they may be employed as biomarkers for GC prognosis.

There are limitations in our manuscript. First, origin of circulating lncRNAs was not fully understood. Some investigators suggested serum lncRNA profiles was not simply a default product of circulating blood cells but might derive from tissues affected by diseases such as cancer [39]. More focus on release mechanisms of identified lncRNAs in tumorigenesis and progression of GC may be a valuable avenue for increasing diagnostic specificity. Moreover, although we have constructed a promising four-lncRNA panel for GC detection in serum, it is uncertain if this panel is only specific for GC. Thus, additional studies will be required to further examine the expression changes of these four lncRNAs in other tumors. Finally, confirmation of our findings in a multicenter trial of larger independent samples is the objective of ongoing work.

In conclusion, we have defined a distinctive serum lncRNA signature for the detection of GC and identified LOC100506474 and UCA1 as an independent predictor of GC recurrence. Although further multicenter studies are needed to confirm the results of our study, our findings may provide a foundation for development of a novel noninvasive test to predict GC and determination of innovative therapeutic strategies.

Acknowledgements

This work was supported by National 863 program project (2014AA022304), Natural Science Foundation of Shandong (ZR2015PH049, 2013GSF11824), Science Foundation of Qilu Hospital of Shandong University, and the Fundamental Research Funds of Shandong University.

Disclosure of conflict of interest

None.

Address correspondence to: Qian Wang, Department of Clinical Laboratory, Qilu Hospital, Shandong University, 107 Wenhua Xi Road, Jinan 250-012, Shandong Province, China. Tel: 86-0531-68697299; Fax: 86-0532-66850879; E-mail: sd.wangqian@163.com

References

[1] Sugano K. Screening of gastric cancer in Asia. Best Pract Res Clin Gastroenterol 2015; 29: 895-905.

[2] Liu Z, Shao Y, Tan L, Shi H, Chen S and Guo J. Clinical significance of the low expression of FER1L4 in gastric cancer patients. Tumour Biol 2014; 35: 9613-9617.

[3] Sun Z and Zhang N. Clinical evaluation of CEA, CA19-9, CA72-4 and CA125 in gastric cancer patients with neoadjuvant chemotherapy. World J Surg Oncol 2014; 12: 397.

[4] Ponting CP, Oliver PL and Reik W. Evolution and functions of long noncoding RNAs. Cell 2009; 136: 629-641.

[5] Chandra Gupta S, Nandan Tripathi Y. Potential of long non-coding RNAs in cancer patients: from bio-markers to therapeutic targets. Int J Cancer 2017; 140: 1955-1967.

[6] Sun W, Yang Y, Xu C, Xie Y and Guo J. Roles of long noncoding RNAs in gastric cancer and their clinical applications. J Cancer Res Clin Oncol 2016; 142: 2231-2237.

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

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

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

[10] Lai Y, Xu P, Li Q, Ren D, Wang J, Xu K and Gao W. Downregulation of long noncoding RNA ZMAT1 transcript variant 2 predicts a poor prognosis in patients with gastric cancer. Int J Clin Exp Pathol 2015; 8: 5556-5562.

[11] Hu Y, Pan J, Wang Y, Li L and Huang Y. Long noncoding RNA linc-UBC1 is negative prognostic factor and exhibits tumor pro-oncogenic activity in gastric cancer. Int J Clin Exp Pathol 2015; 8: 594-600.

[12] Sun W, Yang Y, Xu C, Xie Y and Guo J. Roles of long noncoding RNAs in gastric cancer and their clinical applications. J Cancer Res Clin Oncol 2016; 142: 2231-2237.

[13] Liu FT, Qiu C, Luo HL, Zhang Y, Xia GF, Hao TF and Zhu PQ. The association of HOTAIR expression with clinicopathological features and prognosis in gastric cancer patients. Panminerva Med 2016; 58: 167-174.

[14] Chen JS, Wang YF, Zhang XQ, Lv JM, Li Y, Liu XX and Xu TP. H19 serves as a diagnostic biomarker and up-regulation of H19 expression contributes to poor prognosis in patients with gastric cancer. Neoplasma 2016; 63: 223-230.

[15] Zhou X, Yin C, Dang Y, Ye F and Zhang G. Identification of the long non-coding RNA H19 in plasma as a novel biomarker for diagnosis of gastric cancer. Sci Rep 2015 22; 5: 11516.

[16] Li P, Xue WJ, Feng Y and Mao QS. Long non-coding RNA CASC2 suppresses the proliferation of gastric cancer cells by regulating the MAPK signaling pathway. Am J Transl Res 2016; 8: 3522-3529.

[17] Liu Y, Zhao J, Zhang W, Gan J, Hu C, Huang G and Zhang Y. LncRNA GAS5 enhances G1 cell cycle arrest via binding to YBX1 to regulate p21 expression in stomach cancer. Sci Rep 2015; 5: 10159.

[18] Sun M, Jin FY, Xia R, Kong R, Li JH, Xu TP, Liu YW, Zhang EB, Liu XH and De W. Decreased expression of long noncoding RNA GAS5 indicates a poor prognosis and promotes cell proliferation in gastric cancer. BMC Cancer 2014; 14: 319.

[19] Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, Zhao S, Poliakov A, Cao X, Dhanasekaran SM, Wu YM, Robinson DR, Beer DG, Feng FY, Iyer HK and Chinnaiyan AM. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet 2015; 47: 199-208.

[20] Hanley JA and McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983; 148: 839-843.

[21] Gibb EA, Brown CJ and Lam WL. The functional role of long non-coding RNA in human carcinomas. Mol Cancer 2011; 10: 38.

[22] Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 2010; 464: 1071-1076.

[23] Wang Y, Xue K, Guan Y, Jin Y, Liu S, Wang Y, Liu S, Wang L and Han L. Long non-coding RNA LINC00261 suppresses cell proliferation and invasion and promotes cell apoptosis in human choriocarcinoma. Oncol Res 2016; [Epub ahead of print].
[24] Meng Q, Ren M, Li Y and Song X. LncRNA-RMRP acts as an oncogene in lung cancer. PLoS One 2016; 11: e0164845.
[25] Han Y, Xu H, Cheng J, Zhang Y, Gao C, Fan T, Peng B, Li B, Liu L and Cheng Z. Down-regulation of long non-coding RNA H19 promotes P19CL6 cells proliferation and inhibits apoptosis during late-stage cardiac differentiation via miR-19b-modulated Sox6. Cell Biosci 2016; 6: 58.
[26] De Kok JB, Verhaegh GW, Roelofs RW, Hessels D, Kiemeney LA, Aalders TW, Swinkels DW and Schalken JA. DD3(PCA3), a very sensitive and specific marker to detect prostate tumors. Cancer Res 2002; 62: 2695-2698.
[27] Abramowicz A, Wojakowska A, Gdowicz-Klosok A, Polanska J, Rodziewicz P, Polanowski P, Namysl-Kaletka A, Pietrowska M, Wydmanski J, Widlak P. Identification of serum proteome signatures of locally advanced and metastatic gastric cancer: a pilot study. J Transl Med 2015; 13: 304.
[28] Ren W, Zhang J, Li W, Li Z, Hu S, Suo J and Ying X. A tumor-specific prognostic long non-coding RNA signature in gastric cancer. Med Sci Monit 2016; 22: 3647-3657.
[29] Shi T, Gao G and Cao Y. Long noncoding RNAs as novel biomarkers have a promising future in cancer diagnostics. Dis Markers 2016; 2016: 9085195.
[30] Lu YF, Liu ZC, Li ZH, Ma WH, Wang FR, Zhang YB and Lu JB. Esophageal/gastric cancer screening in high-risk populations in Henan Province China. Asian Pac J Cancer Prev 2014; 15: 1419-1422.
[31] Zhuang JJ, Yue M, Zheng YH, Li JP and Dong XY. Long non-coding RNA MVIH acts as a prognostic marker in glioma and its role in cell migration and invasion. Eur Rev Med Pharmacol Sci 2016; 20: 4898-4904.
[32] Wang Q, Cheng N, Li X, Pan H, Li C, Ren S, Su C, Cai W, Zhao C, Zhang L and Zhou C. Correlation of long non-coding RNA H19 expression with cisplatin-resistance and clinical outcome in lung adenocarcinoma. Oncotarget 2017; 8: 2558-2567.
[33] Wang R, Du L, Yang X, Jiang X, Duan W, Yan S, Xie Y, Zhu Y, Wang Q, Wang L, Yang Y and Wang C. Identification of long noncoding RNAs as potential novel diagnosis and prognosis biomarkers in colorectal cancer. J Cancer Res Clin Oncol 2016; 142: 2291-2301.
[34] Ma L, Zhou Y, Luo X, Gao H, Deng X and Jiang Y. Long non-coding RNA XIST promotes cell growth and invasion through regulating miR-497/MACC1 axis in gastric cancer. Oncotarget 2017; 8: 4125-4135.
[35] Chen DL, Ju HQ, Lu YX, Chen LZ, Zeng ZL, Zhang DS, Luo HY, Wang F, Qiu MZ, Wang DS, Xu DZ, Zhou ZW, Pelicano H, Huang P, Xie D, Wang FH, Li YH and Xu RH. Long non-coding RNA XIST regulates gastric cancer progression by acting as a molecular sponge of miR-101 to modulate EZH2 expression. J Exp Clin Cancer Res 2016; 35: 142.
[36] Shang C, Guo Y, Zhang J and Huang B. Silence of long noncoding RNA UCA1 inhibits malignant proliferation and chemotherapy resistance to adriamycin in gastric cancer. Cancer Chemother Pharmacol 2016; 77: 1061-1067.
[37] Gao J, Cao R and Mu H. Long non-coding RNA UCA1 may be a novel diagnostic and predictive biomarker in plasma for early gastric cancer. Int J Clin Exp Pathol 2015; 8: 12936-12942.
[38] Atmadibrata B, Liu PY, Sokolowski N, Zhang L, Wong M, Tee AE, Marshall GM and Liu T. The novel long noncoding RNA linc00467 promotes cell survival but is down-regulated by N-Myc. PLoS One 2014; 9: e88112.
[39] Kitow J, Derda AA, Beermann J, Kumarswarmy R, Pfanne A, Fendrich J, Lorenzen JM, Xiao K, Bavendiek U, Bauersachs J and Thum T. Mitochondrial long noncoding RNAs as blood based biomarkers for cardiac remodeling in patients with hypertrophic cardiomyopathy. Am J Physiol Heart Circ Physiol 2016; 311: H707-H712.