



Research article

Comprehensive analysis of competitive endogenous RNAs network reveals potential prognostic lncRNAs in gastric cancer

Dengzhong Sun^a, Yongzhi Miao^a, Wu Xu^a, Weijun Shi^a, Lili Wang^a, Tianwen Chen^a, Hexin Wen^a, Huazhang Wu^b, Mulin Liu^{a,*}^a Department of Gastrointestinal Surgery, The First Affiliated Hospital of Bengbu Medical College, Bengbu, China^b School of Life Sciences, Anhui Province Key Laboratory of Translational Cancer Research, Bengbu Medical College, Bengbu, China

ARTICLE INFO

Keywords:

Bioinformatics
Cancer research
Biological sciences
CeRNA
Gastric cancer
LncRNA
Prognosis

ABSTRACT

Long non-coding RNAs (lncRNAs) are key regulators of a range of human diseases, including various cancers, with multiple previous studies having explored lncRNA dysregulation in the context of gastric cancer (GC). The present study sought to expand upon these previous results by downloading lncRNA, mRNA, and microRNA (miRNA) expression profiles derived from 180 GC tissues and 24 normal control tissues within the Cancer Genome Atlas (TCGA) database. These datasets were then interrogated to identify GC-related differentially expressed (DE) RNAs ($|\text{fold change}| \geq 2$, $\text{FDR} < 0.01$), leading to the identification of 1946 DE lncRNAs, 123 DE miRNAs, and 3159 DE mRNAs. These results were then used to generate a putative GC-related competitive endogenous RNA (ceRNA) network composed of 131 lncRNAs, 9 miRNAs, and 78 mRNAs. Subsequent survival analyses based upon this network revealed 17 of these lncRNAs to be significantly associated with GC patient survival ($P < 0.05$). Further multivariable Cox regression and lasso analyses allowed for the construction of an 8-lncRNA risk score that was able to effectively predict GC patient survival with good discriminative ability. The Kaplan-Meier Plotter database further confirmed that network hub genes that were related to these 8 lncRNAs were associated with GC patient prognosis ($P < 0.05$). As the ceRNA network in the present study was constructed with a focus on both disease stage and differential gene expression, it represents a key resource that will offer valuable insights into the mechanistic roles of ceRNA pathways in GC development and progression.

1. Introduction

Gastric cancer (GC) remains the second leading cause of cancer-related mortality globally [1]. As such, it is vital that novel approaches to diagnosing and treating GC be developed. While many advances in cancer treatment have been made in recent decades, GC patients still typically have a relatively poor prognosis [2]. Individuals with advanced GC typically suffer from tumor invasion and metastasis, leading to a marked reduction in the average duration of patient survival [3]. The mechanisms governing GC onset and progression are complex, and as such preventing and treating GC remains challenging. It is therefore essential that the mechanistic basis for GC invasion and metastasis be better understood, and it is equally important that novel diagnostic and prognostic GC-related biomarkers be identified.

It is possible to construct ceRNA gene interaction networks incorporating lncRNAs, miRNAs, and mRNAs. As they can bind directly to miRNAs, lncRNAs can function as competitive endogenous molecules

that can sequester these target miRNAs, thereby positively regulating the expression of miRNA-targeted mRNAs [4]. The roles of such indirect lncRNA-mRNA regulatory relationships have recently been shown to be directly related to cancer development and progression in multiple experimental contexts. For example, H19 has been shown to promote the proliferation and metastasis of GC cells via competitively binding miR-22-3p and thereby promoting the upregulation of Snail1, which is related to the epithelial-mesenchymal transition (EMT) in tumor cells [5]. At present, however, robust and reliable ceRNA network databases exploring tumor-associated lncRNA-miRNA-mRNA relationships are lacking, making it difficult to fully understand the pathological basis for many forms of cancer.

The Cancer Genome Atlas (TCGA) is a large public database containing data corresponding to over 30 forms of human cancer, including patient clinicopathological information [6]. Given its robust nature, the TCGA is ideal for data mining efforts aimed at exploring the mechanistic basis for tumor development. This data has been used to construct ceRNA

* Corresponding author.

E-mail address: liumulin66@aliyun.com (M. Liu).

networks corresponding to tumor types including lung cancer [7], breast cancer [8] and colorectal cancer [9]. The construction of such ceRNA networks is an invaluable means of highlighting complex genetic relationships and of identifying potential diagnostic, prognostic, or therapeutic biomarkers of disease.

This study was designed to initially compare the differential expression of lncRNAs, miRNAs, and mRNAs in GC patient tissue samples relative to normal tissue samples, with an additional focus on the expression of these genes as a function of GC stage. Those differentially expressed (DE) lncRNAs, miRNAs, and mRNAs that were shared among GC stages were then used to construct a GC-associated ceRNA network. The putative relationships between these different DE RNAs were predicted using the miRcode, TargetScan, miRDB, and miRTarBase databases in order to identify predicted interactions. These predictions

allowed for the construction of a ceRNA network composed of 131 lncRNAs, 9 miRNAs, and 78 mRNAs. The resultant network may offer value as a means of rapidly identifying key GC-related genes. In addition, we were able to leverage these results by using a lasso-penalized Cox regression analysis to generate an 8-lncRNA-based risk score that was able to independently predict GC patient survival outcomes, highlighting the value of this research approach.

2. Materials and methods

2.1. Data retrieval and processing

GC patient lncRNA, miRNA, and mRNA data and corresponding clinical data were downloaded from TCGA [10, 11] (<https://porta>

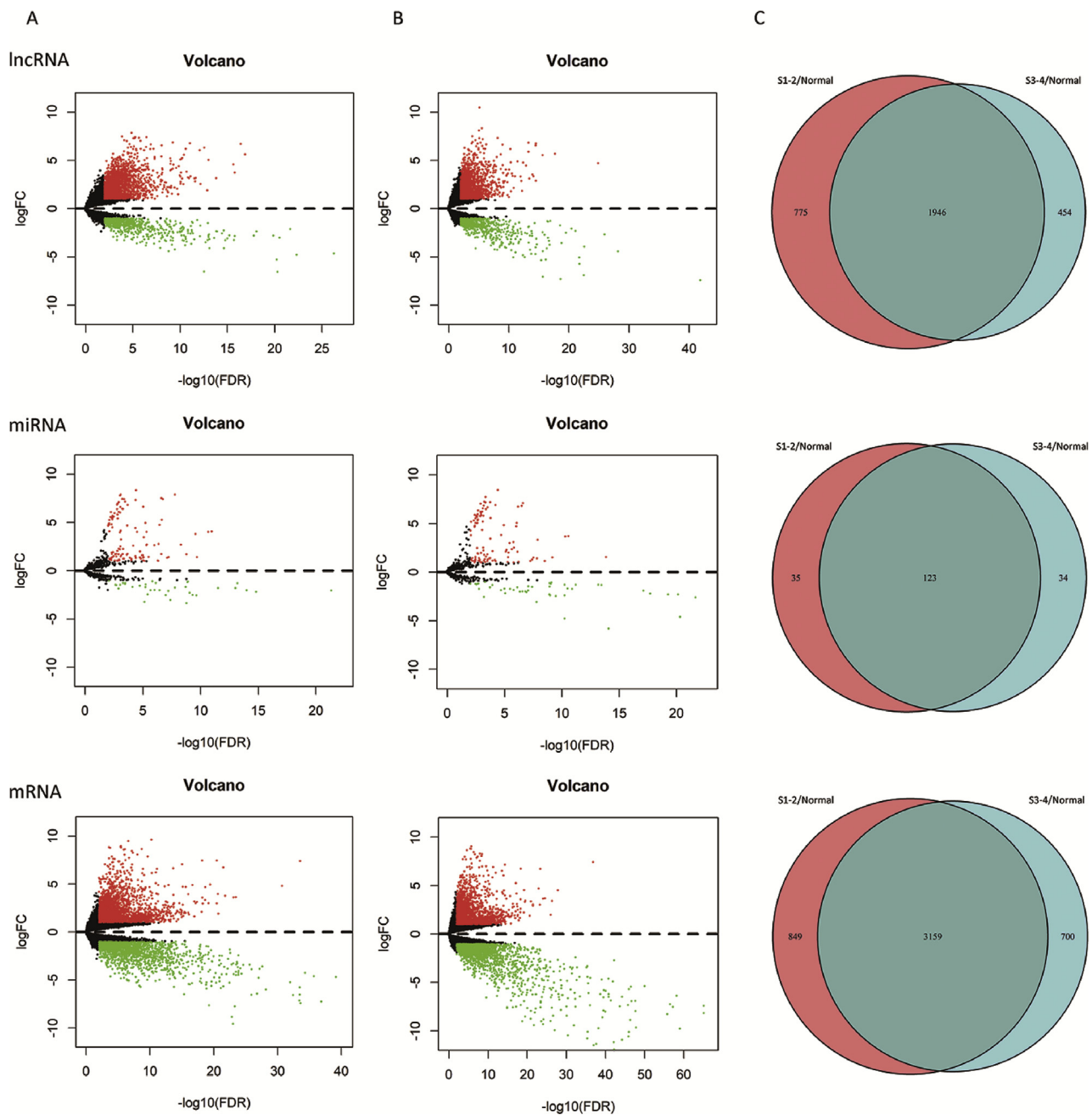


Figure 1. GC-related DE RNA identification. Differentially expressed lncRNAs (top), miRNAs (middle), and mRNAs (bottom) between normal tissue and either early-stage GC (stage I-II; A) or advanced GC (stage III-IV; B) are shown using volcano plots. Upregulated RNAs are shown in red, while downregulated RNAs are shown in green. Overlapping DE RNAs between these two comparisons were identified using Venn diagrams (C).

l.gdc.cancer.gov/). A total of 180 GC tumor tissue samples and 24 normal tissue samples were included in this cohort. The Ensemble database was then used to convert RNA sequences into gene symbols [12]. Samples were only included in this analysis if they contained data for all three RNA types of interest (lncRNA, miRNA and mRNA), as well as information regarding patient survival and tumor staging, which can ensure that every GC sample with explicit pathological stage contains not only the expression of lncRNA, miRNA and mRNA, but also survival information for subsequent prognostic analysis. The detailed sequences of GC clinical samples can be found in Table S1.

2.2. DE RNA identification

The edgeR package [13, 14] installed in R (version 3.6.1, www.r-project.org) was used to identify DE lncRNAs, DE miRNAs, and DE mRNAs in two separate comparisons: early-stage GC (stage I-II) versus normal tissue, and advanced GC (stage III-IV) versus normal tissue. The cutoff criteria for differential expression were $|\text{fold change}| \geq 2$ and $\text{FDR} < 0.01$ [15]. Identified DE RNAs were then arranged into volcano plots, and those DE genes (DEGs) that overlapped between these comparisons were further identified using Venn diagrams.

2.3. ceRNA network construction

The miRcode database [16] (<http://mircode.org/>) was used to predict lncRNA-mediated binding and sequestration of target miRNAs, while putative miRNA target genes were identified using the miRDB [17] (<http://mirdb.org/>), miRTarBase [18] (<http://mirtarbase.mbc.nc.tu.edu.tw>), and TargetScan databases [19] (<http://www.targetscan.org>). Putative miRNA target genes that were also differentially expressed at the mRNA level were then used to construct a ceRNA network which incorporated all consistently predicted interactions among DE lncRNAs, miRNAs, and mRNAs in the present study. This ceRNA network was then visualized using Cytoscape [20] (www.cytoscape.org).

2.4. DE mRNA functional and pathway analyses

GO and KEGG enrichment analyses of those DE mRNAs in the ceRNA network were used to explore the molecular basis for GC development. The clusterProfiler packages [21] were used for both gene ontology (GO) [22] and KEGG pathway enrichment analyses [23], as in previous studies, with $P < 0.05$ used as a significance threshold.

2.5. Survival analysis

As they represented the top regulatory layer in our ceRNA network, the associations between lncRNA expression and GC patient survival outcomes were assessed using a univariable Cox regression model [24]. These analyses were conducted using the R survival package, with hazard ratios and corresponding 95% confidence intervals (CIs) being calculated, and with $P < 0.05$ as a significance threshold.

2.6. Risk scoring

The DE lncRNAs which were determined to be significantly related to GC patient survival were considered as ideal diagnostic and prognostic biomarkers, and were therefore analyzed via a lasso-penalized Cox regression approach in order to eliminate confounding variables and extraneous lncRNAs [25]. The optimal lambda value necessary to minimize mean cross-validated error and regression coefficients (β) of this multivariable Cox regression model were determined via ten-fold cross-validation. The predictive accuracy of the resultant risk scoring system was then assessed using time-dependent receiver operating characteristic (ROC) curves.

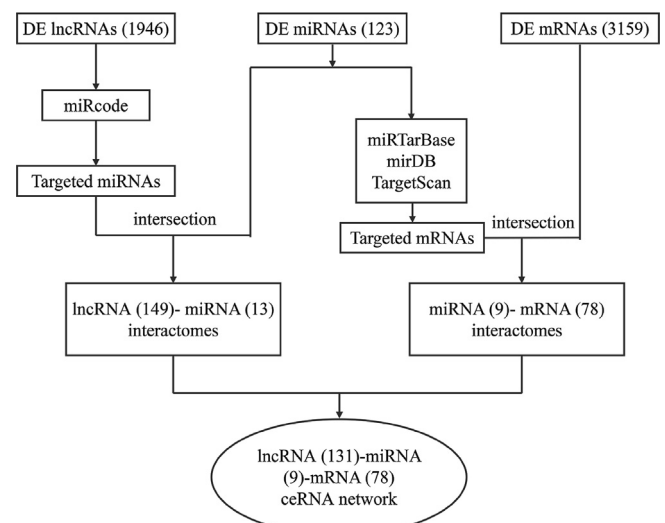


Figure 2. Study ceRNA regulatory network construction strategy.

2.7. Cox regression analyses

The relationships between specific clinical parameters (age, sex, tumor grade, TNM stage, pathologic stage) and GC patient survival were initially assessed via univariable Cox regression analysis. Variables that were significant in this initial analysis ($P < 0.05$) were incorporated as candidate variables in a multivariable analysis [26]. $P < 0.05$ was the significance threshold, and hazard ratios and 95% CIs were calculated for all variables.

2.8. PPI network construction and prognostic assessment

A GC patient prognosis-related subnetwork was constructed using the lncRNAs identified via Lasso regression, and a protein-protein interaction network (PPI) network was then constructed based upon the mRNAs within this sub-network. STRING (<https://string-db.org/>) was used for PPI network construction. A total of 5 hub genes in this network were then chosen for lncRNA-mRNA regression analyses which were conducted using the Cytohubba plugin in Cytoscape. The Kaplan-Meier Plotter database (<http://kmplot.com>) was then employed for survival analysis of those hub genes which had achieved significance, with $P < 0.01$ and $R > 0.3$ as the criteria for significance.

3. Results

3.1. Identification of GC-related differentially expressed lncRNAs, miRNAs, and mRNAs

In order to identify GC stage-related DE RNAs, we initially separated our downloaded sample cohort into three stage-based cohorts: normal samples ($n = 24$), early-stage GC (stage I-II; $n = 93$) and advanced GC (stage III-IV; $n = 87$). We then compared the two GC cohorts to normal sample controls (Figure 1A and Figure 1B) in order to identify GC-related DE lncRNAs, DE miRNAs, and DE mRNAs based on the use of $|\text{fold change}| \geq 2$ and $\text{FDR} < 0.01$ as criteria for differential expression. We then compared these two DE RNA datasets, identifying 1,946 shared DE lncRNAs, 123 shared DE miRNAs and 3,159 shared DE mRNAs (Figure 1C). It seems that differential expression of the genes in the overlapped part leads to both initiation and progression of cancer, but further identification of interacting lncRNA-miRNA and miRNA-mRNA pairs is necessary in order to better understand the regulatory relationship between these identified DE RNAs.

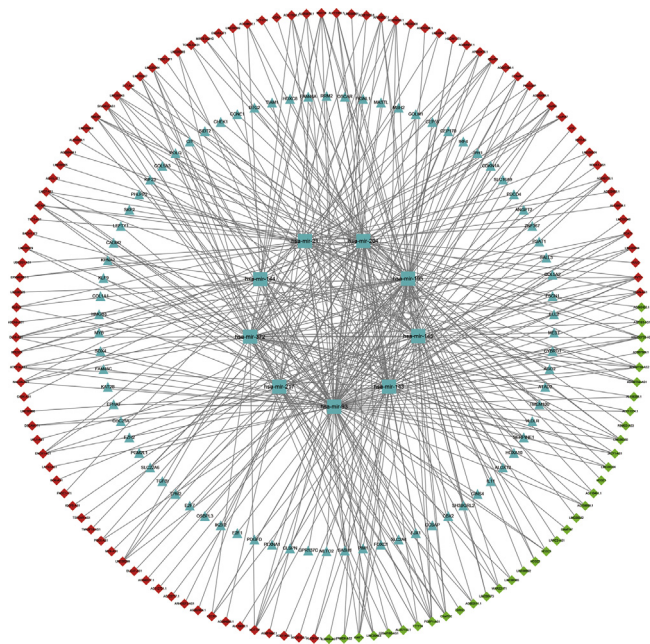


Figure 3. The constructed GC-related ceRNA network. In this network, a total of 218 nodes and 404 edges, there are 1–8 interactions between one lncRNA by other miRNAs; meanwhile, there are 3–29 interactions between one miRNA by other mRNA. Diamonds represent 131 lncRNAs; squares represent 9 miRNAs; triangles represent 78 mRNAs. Upregulated lncRNAs are shown in red, while downregulated lncRNAs are shown in green.

3.2. Identification of putative lncRNA-miRNA interactions

We next sought to identify lncRNA-miRNA interaction pairs among the DE RNAs identified in Figure 1. The process of ceRNA network construction is detailed in Figure 2. Initially, those miRNAs that were predicted to interact with the 1946 identified DE lncRNAs were selected using the miRcode database. We then determined which of these predicted miRNAs were represented in our DE miRNA dataset, leading us to identify 87 DE miRNAs of interest. We then ultimately identified 149 lncRNAs and 13 miRNAs that were predicted to undergo mutual interactions.

3.3. Identification of putative miRNA-mRNA interactions

Using a process similar to that detailed above, we next identified putative targets of the 13 DE miRNAs identified as lncRNA targets.

Potential miRNA targets were first selected using miRDB, miRtarBase, and TargetScan, and the resultant mRNA list was then compared to our list of 3,159 DE mRNAs. This led to the identification of 78 DE mRNAs that were targets of the 13 DE miRNAs of interest in the present study.

3.4. ceRNA network construction

As the above approach led to the indirect generation of a list of 78 DE mRNAs of interest from an initial list of 149 lncRNAs, we next revised this ceRNA network by omitting those lncRNA-miRNA and miRNA-mRNA pairs that did not form part of a predicted lncRNA-miRNA-mRNA relationship. This ultimately led to a final list of 131 lncRNAs, 9 miRNAs, and 78 mRNAs that were incorporated into a final GC-associated ceRNA network (Figure 3).

3.5. Functional enrichment analysis

The potential biological roles of the 78 DE mRNAs within our ceRNA network were next assessed via GO and KEGG functional enrichment analyses. A total of 105 GO biological process terms were found to be significantly enriched for these DE mRNAs (adjusted $P < 0.05$). The top terms identified were, in order, "DNA integrity checkpoint", "DNA damage checkpoint", "cell cycle checkpoint" and "mitotic DNA damage checkpoint" (Figure 4A; Table 1). In addition, 9 KEGG pathways were significantly enriched for these identified DE mRNAs (adjusted $P < 0.05$), including "microRNAs in cancer", "cell cycle", "PI3K-Akt signaling pathway" and "P53 signaling pathway" (Figure 4B; Table 2). These results suggest that the cell cycle, PI3K-Akt signaling, and P53 signaling are all important regulators of GC development.

3.6. The association between ceRNA network-associated genes and GC patient prognosis

We next conducted univariable Cox regression analyses for the 131 DE lncRNAs in our ceRNA network in order to identify lncRNAs significantly associated with GC patient survival. Of these lncRNAs, we found that 17 were significantly correlated with GC patient overall survival (OS) ($P < 0.05$; Table 3). A lasso-penalized Cox regression analysis was then used to determine whether these genes were independently related to GC patient outcomes.

3.7. lncRNA-related risk score generation

As lncRNAs exhibit defined expression patterns and were the top-level regulators in our ceRNA network, they represented ideal potential biomarkers for GC diagnostic and prognostic analyses. We therefore used

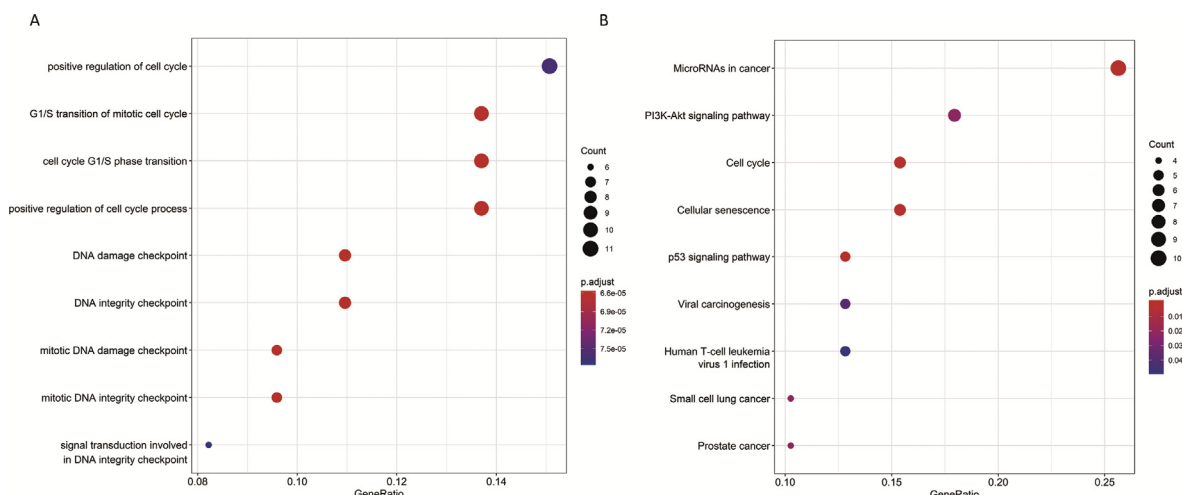


Figure 4. Functional enrichment analysis. GO (A) and KEGG (B) enrichment analyses for 78 DE mRNAs in the ceRNA network; show category = 9.

Table 1. GO biological process terms for DE mRNAs in the ceRNA network.

ID	Description	P adjust	q value	Gene symbols
GO:0000077	DNA damage checkpoint	0.000	0.000	BTG2/CLSPN/E2F1/CHEK1/CDKN1A/MSH2/SOX4/E2F7
GO:0044773	mitotic DNA damage checkpoint	0.000	0.000	BTG2/CLSPN/E2F1/CDKN1A/MSH2/SOX4/E2F7
GO:0000082	G1/S transition of mitotic cell cycle	0.000	0.000	BTG2/E2F1/CDKN1A/SOX4/E2F7/CCNE1/EZH2/SKP2/CDC25A/RRM2
GO:0031570	DNA integrity checkpoint	0.000	0.000	BTG2/CLSPN/E2F1/CHEK1/CDKN1A/MSH2/SOX4/E2F7
GO:0044774	mitotic DNA integrity checkpoint	0.000	0.000	BTG2/CLSPN/E2F1/CDKN1A/MSH2/SOX4/E2F7
GO:0044843	cell cycle G1/S phase transition	0.000	0.000	BTG2/E2F1/CDKN1A/SOX4/E2F7/CCNE1/EZH2/SKP2/CDC25A/RRM2
GO:0090068	positive regulation of cell cycle process	0.000	0.000	BTG2/KIF23/E2F1/CDKN1A/SOX4/E2F7/CIT/CCNE1/EZH2/CDC25A
GO:0045787	positive regulation of cell cycle	0.000	0.000	BTG2/KIF23/E2F1/CHEK1/CDKN1A/SOX4/E2F7/CIT/CCNE1/EZH2/CDC25A
GO:0072401	signal transduction involved in DNA integrity checkpoint	0.000	0.000	BTG2/E2F1/CHEK1/CDKN1A/SOX4/E2F7

Table 2. KEGG pathways analyses for the ceRNA network based on DE mRNAs.

ID	Description	P adjust	q value	Gene symbols
hsa05206	MicroRNAs in cancer	0.000	0.000	KIF23/E2F1/CDKN1A/SOX4/PIM1/CCNE1/PDCD4/EZH2/CDC25A/FSCN1
hsa04115	p53 signaling pathway	0.001	0.001	CHEK1/CDKN1A/CCNE1/SERPINE1/RRM2
hsa04110	Cell cycle	0.001	0.001	E2F1/CHEK1/CDKN1A/CCNE1/SKP2/CDC25A
hsa04218	Cellular senescence	0.003	0.003	E2F1/CHEK1/CDKN1A/CCNE1/SERPINE1/CDC25A
hsa05222	Small cell lung cancer	0.022	0.018	E2F1/CDKN1A/CCNE1/SKP2
hsa05215	Prostate cancer	0.022	0.018	E2F1/CDKN1A/PDGFD/CCNE1
hsa04151	PI3K-Akt signaling pathway	0.022	0.019	COL1A1/ANGPT2/CDKN1A/PDGFD/CCNE1/PHLPP2/MYB
hsa05203	Viral carcinogenesis	0.038	0.031	KAT2B/CHEK1/CDKN1A/CCNE1/SKP2
hsa05166	Human T-cell leukemia virus 1 infection	0.048	0.040	KAT2B/E2F1/CHEK1/CDKN1A/CCNE1

lasso-penalized Cox regression analyses in order to determine which of the 17 survival-related lncRNAs identified above were significantly associated with patient prognosis, with the contribution of each lncRNA being weighted using relative coefficients (Figure 5B). This led to the exclusion of 11 lncRNAs (Figure 5A), yielding the final risk score formula: risk score = (0.0040 × LINC00330 expression level) + (0.0016 × AC061975.6 expression level) + (0.0012 × ST7-AS2 expression level) + (6.9102 × LINC00346 expression level) + (0.0004 × LINC00473 expression level) + (9.5852 × AL158206.1 expression level) + (0.0062 × AC110491.1 expression level) + (0.0001 × LINC00460 expression level). We then determined the risk scores for all patients in our cohort using the above formula, and stratified patients into high (n = 75) and low-risk groups (n = 95) based upon whether their risk score was greater than or below 1.19 based on the maximally selected rank statistics,

respectively. Patient samples with a 0 day survival period were omitted from this analysis. This stratification approach was able to effectively separate patients based upon their survival outcomes in Kaplan–Meier curve analyses (Figure 5D). We then conducted a univariable analysis in order to identify OS-related predictive factors in GC patients (Figure 5F), revealing this 8-lncRNA-based risk score to be the only independent predictor of GC patient survival in a subsequent multivariable Cox regression analysis (Figure 5G).

3.8. PPI network construction and prognostic assessment

Lastly, we constructed a prognostic sub-network based upon the 8 lncRNAs that were included in our risk score model (Figure 6A), with mRNAs in this sub-network then being used for PPI network construction

Table 3. Survival analysis for DE lncRNAs involved in the ceRNA network.

id	Hazard ratio	Low 95	High 95	P value
LINC00330	1.117	1.064	1.173	0.000
AC061975.6	1.067	1.030	1.106	0.000
AP002478.1	1.070	1.023	1.119	0.003
ST7-AS2	1.096	1.030	1.165	0.004
AC123777.1	1.822	1.196	2.778	0.005
LINC00346	1.019	1.005	1.033	0.007
LINC00473	1.047	1.008	1.086	0.017
AC007389.1	1.573	1.070	2.311	0.021
AL158206.1	1.005	1.001	1.009	0.022
LINC00365	1.033	1.004	1.062	0.024
PVT1	0.990	0.981	0.999	0.028
TM4SF19-AS1	1.072	1.006	1.142	0.031
AC110491.1	1.363	1.011	1.837	0.042
DSCR4-IT1	1.333	1.009	1.761	0.043
LINC00460	1.020	1.001	1.040	0.044
AC011374.1	1.202	1.004	1.439	0.045
HCG22	1.009	1.000	1.018	0.049

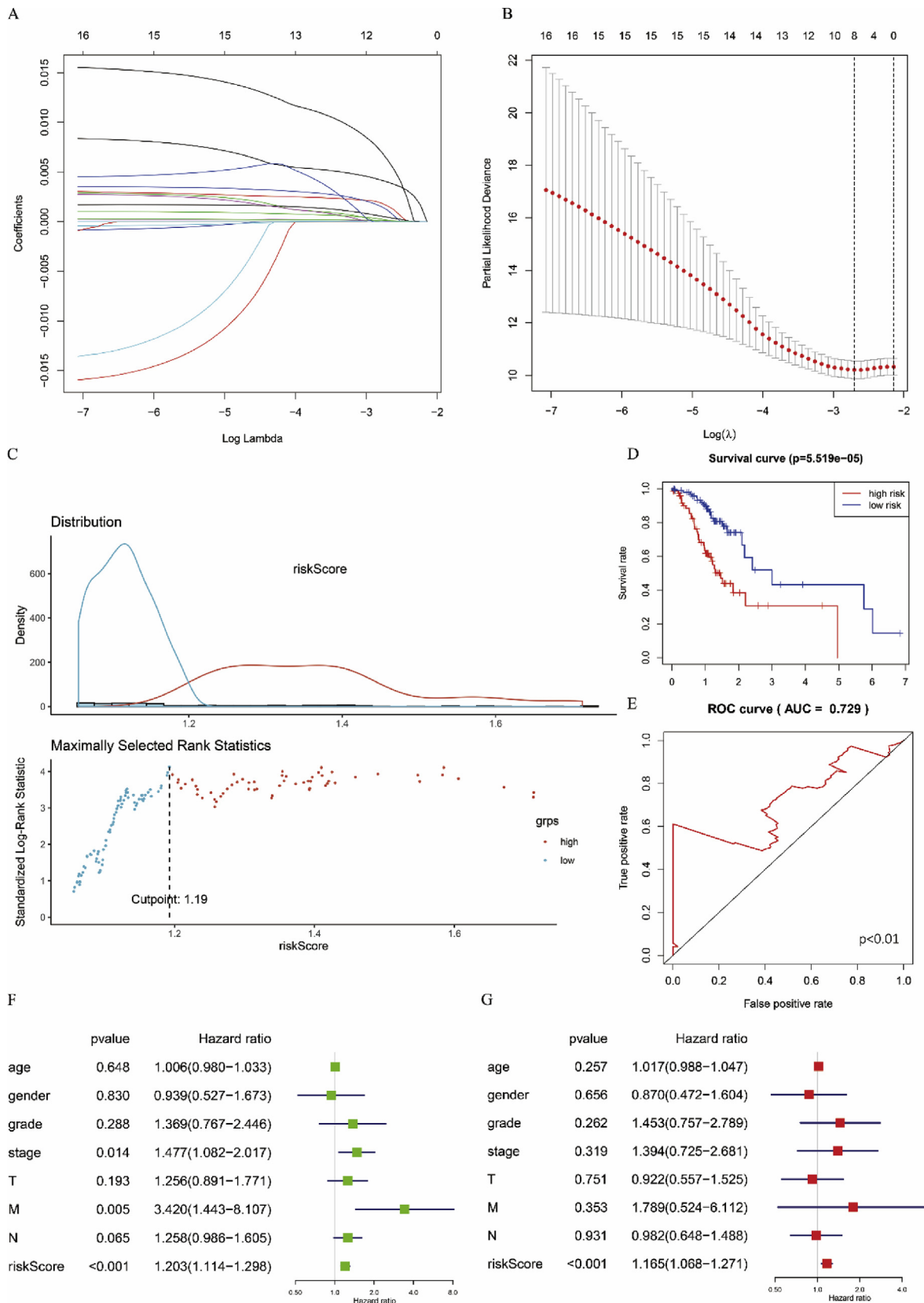


Figure 5. Risk score system. The 17 survival-related DE lncRNAs of interest were subjected to a Lasso-penalized Cox regression analysis of 17 DE lncRNAs, with each lncRNA being shown in a separate curve (A). The optimal lambda leading to the minimum cross-validation error was determined via ten-fold cross-validation (B). Distribution and selection of cutoff value of risk score (C). Kaplan–Meier survival analyses for high- and low-risk score patients were performed (D). Risk score-based time-dependent ROC curves were constructed (E). A risk score analysis for the resultant 8 DE lncRNAs was performed, and both univariable (F) and multivariable (G) analyses of the relationship between clinical parameters and GC patient OS were performed.

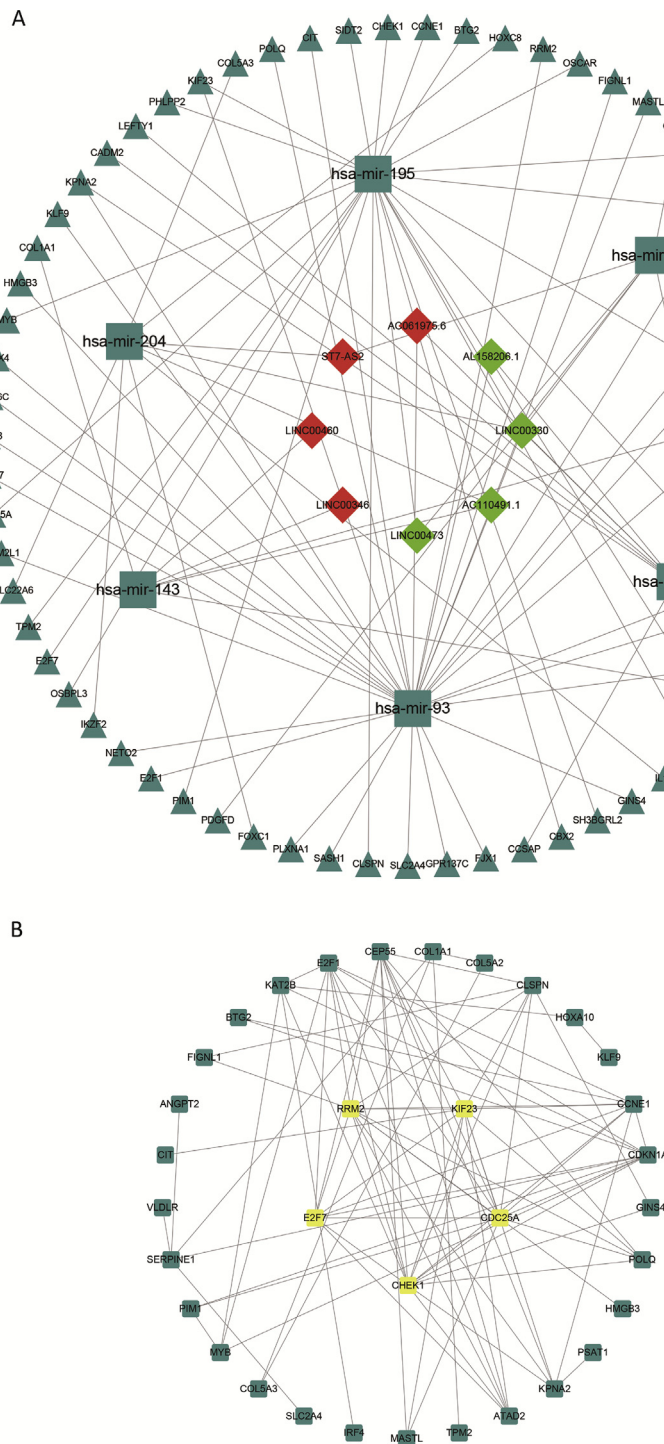


Figure 6. A prognostic sub-network incorporating 8 risk score-related lncRNAs was constructed (A), and the mRNAs in this network were then used for PPI network construction (B). Upregulated lncRNAs are shown in red, while downregulated lncRNAs are shown in green and hub genes are shown in yellow.

(Figure 6B). A total of 5 hub genes within this network were then selected for further analysis. Theoretically, lncRNAs involved in a lncRNA-miRNA-mRNA relationship should positively regulate mRNA expression levels. To validate this mechanism in the context of GC, we conducted regression analyses for the abovementioned 8 lncRNAs and 5 hub mRNAs in GC. This analysis revealed positive correlations between 4 lncRNA-mRNA pairs ($R > 0.30$), with only both CHEK1 and E2F7 being significantly correlated with these 8 risk score-related lncRNAs (Figure 7). We then utilized the Kaplan-Meier Plotter database in order to

demonstrate that CHEK1 and E2F7 were related to GC patient OS, first progression (FP), and post-progression survival (PPS) (Figure 8).

4. Discussion

GC remains the 5th most common form of cancer globally, with incidence rates steadily rising in East Asia, and with increasingly high mortality rates in younger individuals [27]. While the surgical resection of early-stage GC can substantially improve outcomes in certain individuals, a large proportion of patients are either not candidates for

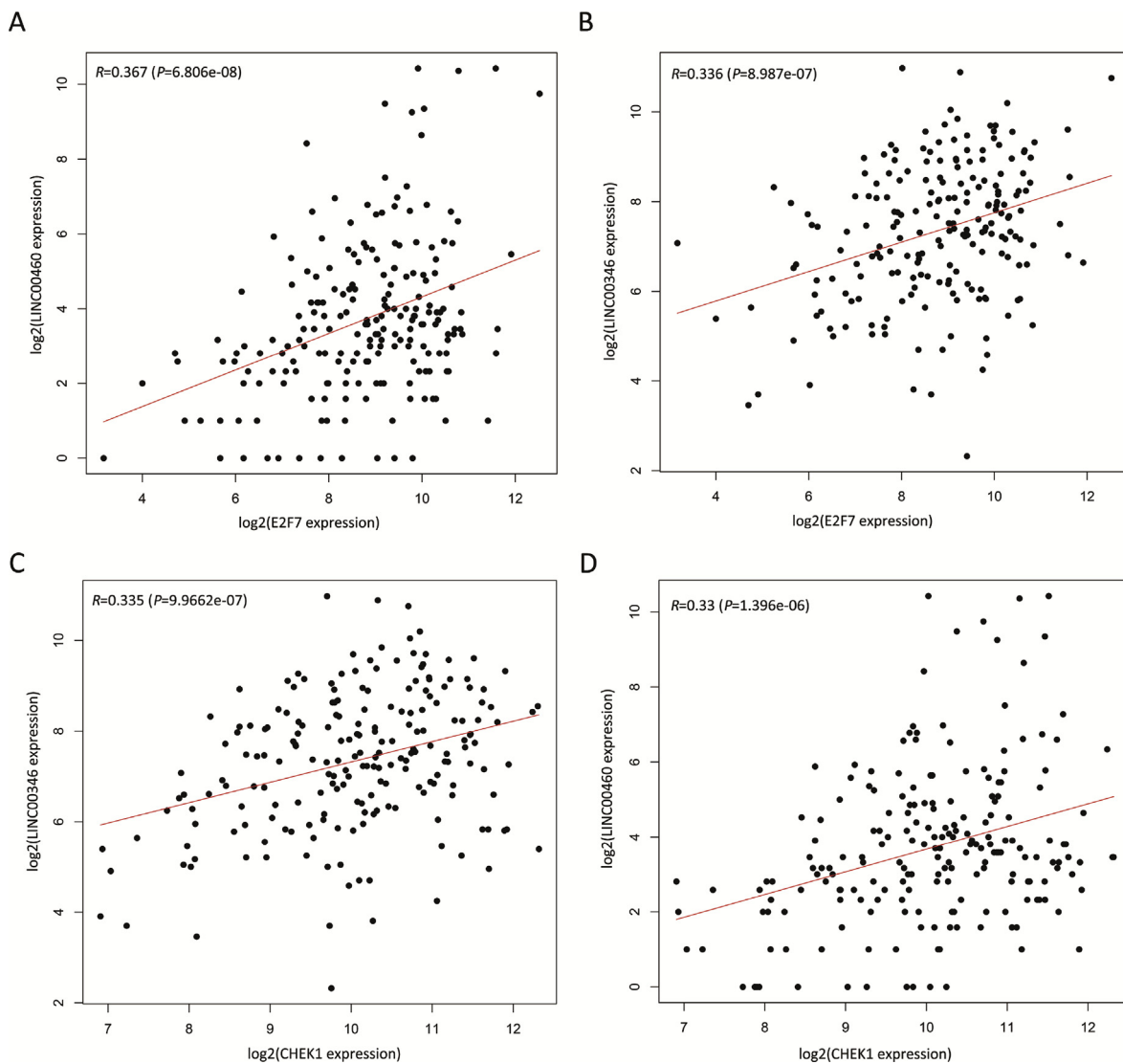


Figure 7. Regression analyses for 8 risk score-associated lncRNAs and 5 hub genes. E2F7 correlates with LINC00460 (A) and LINC00346 (B), and CHEK1 correlates with LINC00346 (C) and LINC00460 (D).

surgical resection or suffer from tumor recurrence or metastasis following treatment [28]. As such, it is vital that the mechanistic basis for GC progression and development be better understood in order to identify novel treatment modalities capable of modulating the regulation of GC tumors. High-throughput sequencing studies have highlighted the vital role played by lncRNAs in myriad biological processes wherein they function as key regulators of gene expression. In particular, the importance of lncRNAs in ceRNA networks is being increasingly well understood, and there is a growing body of evidence suggesting that ceRNA-associated target genes can have a significant impact on cancer patient prognosis and therapeutic resistance [29, 30, 31].

Tumor staging is often conducted based on the TNM staging, which considered primary tumor condition (T), regional lymph node status (N), and whether or not distant metastases are present (M). In the present study, we sought to identify GC stage-related DE lncRNAs, miRNAs and mRNAs by comparing the expression patterns of these different RNA types in normal tissue samples to those in GC samples derived from patients with either early-stage or advanced disease. By identifying putative lncRNA-miRNA and miRNA-mRNA interaction pairs, we were then further able to construct a GC-associated ceRNA regulatory network. The mRNAs within this network were then subjected to GO and KEGG Pathway enrichment analyses, while survival analyses suggested that a

subset of the identified DE lncRNAs were significantly associated with GC patient OS. We then used a lasso-penalized Cox regression analysis to identify 8 lncRNAs that were independently associated with GC patient prognosis, and we used these lncRNAs to develop a risk score model that was capable of independently predicting the survival duration of GC patients within our cohort, a time-dependent ROC curve analysis clearly demonstrated the predictive power of the risk score model that was developed in the present study (Figure 5E). These 8 lncRNAs were then used to construct a ceRNA sub-network from which a PPI network was subsequently constructed, and regression analyses of these lncRNAs and of 5 hub genes within this network allowed us to identify a significant relationship between a subset of these hub genes and lncRNAs. Lastly, a survival analysis conducted using an independent patient cohort revealed these identified hub genes to be significantly associated with GC patient survival. GO terms that were significantly enriched for the DE mRNAs in our ceRNA network included "DNA integrity checkpoint", "DNA damage checkpoint", "cell cycle checkpoint" and "mitotic DNA damage checkpoint", suggesting that the development of GC may be heredity-related. In addition, KEGG terms enriched for these DE mRNAs included "PI3K-Akt signaling pathway", "cell cycle", and "P53 signaling pathway", which is consistent with the frequent identification of the dysregulation of these pathways in many cancer types [32, 33].

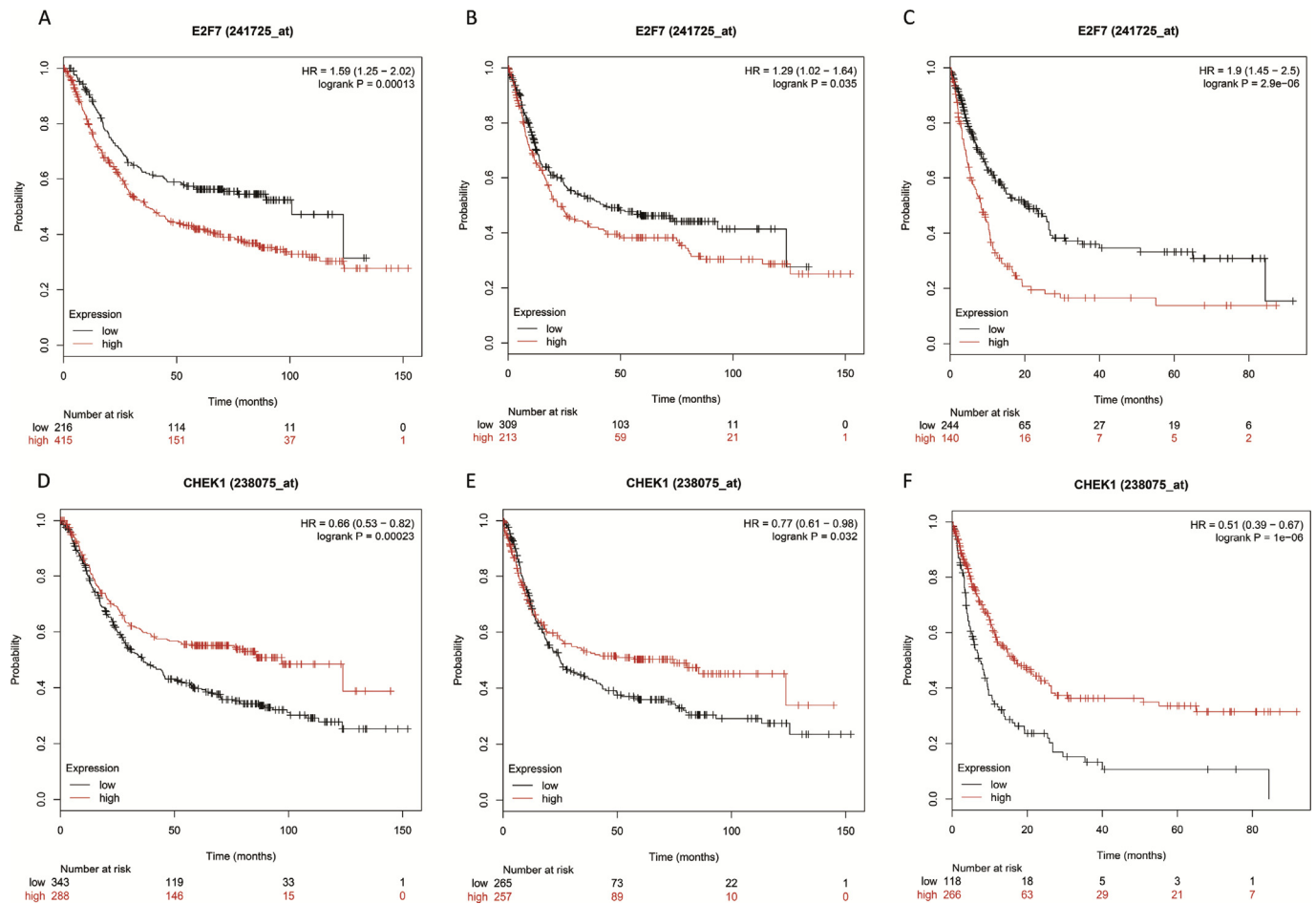


Figure 8. The relationship between hub gene expression and GC patient prognosis. E2F7 is associated with OS (A), FP (B), and PPS (C) in GC patients. CHEK1 is also associated with OS (D), FP (E), and PPS (F) in GC patients.

Importantly, in a univariable Cox regression survival analysis of ceRNA-related genes, we determined that 17 lncRNAs were significantly related to GC patient OS ($P < 0.05$), thus confirming the value of our ceRNA network as a means of identifying putative biomarkers of GC patient prognosis. A subsequent lasso-penalized Cox regression analysis excluded 11 of these 17 lncRNAs from incorporation into our risk score model.

The remaining 8 lncRNAs included 5 unidentified lncRNAs (LINC00330, AC061975.6, ST7-AS2, AL158206.1, AC110491.1) and 3 already reported lncRNAs (LINC00460, LINC00346, LINC00473). LINC00460, which is among the most well-understood oncogenic lncRNAs. It can participate in the occurrence and development of GC through competitively binding miR-342-3p to up-regulate KDM2A expression [34]. This lncRNA has been shown to promote other types of tumor cell migration, proliferation, and metastasis including breast cancer [35], lung cancer [36, 37], colorectal cancer [38, 39] and hepatocellular carcinoma [40], and it is closely linked to a poorer patient prognosis in these tumor types. LINC00346 regulated the expression of CD44 and NOTCH1 by antagonizing miR-34a-5p, served as a critical effector in GC tumorigenesis and progression [41]. It can also promote cancer cell proliferation, migration and invasion in hepatocellular carcinoma [42], bladder cancer [43] and pancreatic cancer [44]. Similarly, LINC00473 is one of the most studied oncogenic lncRNAs, which regulated the migration and invasion of GC cells and related to the poor prognosis of GC patients [45]. Besides, it can also be used as ceRNA to participate in the development and chem-radiotherapy resistance of cervical cancer [46], breast cancer [47], glioma [48], colorectal cancer [49], lung cancer [50], wilms tumour [51], head and neck squamous cell

carcinoma [52], hepatocellular carcinoma [53] and esophageal squamous cell carcinoma [54]. As such, our ceRNA network was able to both successfully identify previously characterized GC-related lncRNAs such as LINC00460, LINC00346, LINC00473, and to identify less well-understood lncRNAs such as LINC00330, AC061975.6, ST7-AS2, AL158206.1, AC110491.1. Interestingly, subsequent analysis also found that hub genes E2F7 and CHEK1 positively related to the 8 lncRNAs were also involved in the occurrence and development of tumors and radiotherapy and chemotherapy resistance [55, 56].

In conclusion, we were able to generate a novel GC-related lncRNA-miRNA-mRNA ceRNA network using data from GC patient tissue samples in various stages of disease progression. This comprehensive network both offers insight into the mechanistic basis of GC and highlights potential targets for future therapeutic and/or diagnostic research, underscoring the value of studying lncRNAs as biomarkers of GC patient prognosis and therapeutic outcomes. As previous studies constructing GC-related lncRNA databases are lacking, there are limitations to our study. Notably, we indirectly conducted external validation survival analyses for 8 lncRNAs by analyzing the survival of two hub genes. In addition, the mechanistic roles of the 8 lncRNAs in our risk score model were not directly assessed. Further work will therefore be needed to both validate and expand upon our findings.

Declarations

Author contribution statement

M. Liu: Conceived and designed the experiments.

D. Sun: Performed the experiments; Wrote the paper.

H. Wu: Analyzed and interpreted the data.

Y. Miao, W. Xu, W. Shi, L. Wang, T. Chen and H. Wen: Contributed reagents, materials, analysis tools or data.

Funding statement

This work was supported by the National Natural Science Foundation of China (21707002); Foundation for the Scientific Research Innovation Team and the Translational Medicine of Bengbu Medical College (BYKC201909, BYTM2019008); the Natural Science Foundation of Anhui Province (1908085MH257).

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at <https://doi.org/10.1016/j.heliyon.2020.e03978>.

References

- Bray, F., Ferlay, J., Soerjomataram, R.L., Siegel, L.A., Torre, A., Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *Ca - Cancer J. Clin.* 68 (6) (2018) 394–424.
- Sun, Q., Sun, J., Li, F., Li, H., Li, S., Bei, X., Zhang, L., Feng, LncRNA LOXL1-AS1 facilitates the tumorigenesis and stemness of gastric carcinoma via regulation of miR-708-5p/USF1 pathway, *Cell Prolif* 2019 (2019), e12687.
- Chen, J., Ren, G., Cai, R., Cai, X., Wu, T., Gui, J., Zhao, C., Guo, Cellular magnetic resonance imaging: in vivo tracking of gastric cancer cells and detecting of lymph node metastases using microparticles of iron oxide in mice, *Canc. Manag. Res.* 11 (2019) 7317–7326.
- D.W. Thomson, M.E. Dinger, Endogenous microRNA sponges: evidence and controversy, *Nat. Rev. Genet.* 17 (5) (2016) 272–283.
- Gan, L., Lv, S., Liao, S., Long non-coding RNA H19 regulates cell growth and metastasis via the miR-22-3p/Snai1 axis in gastric cancer, *Int. J. Oncol.* 54 (6) (2019) 2157–2168.
- Tomczak, P., Czerwińska, M., Wiznerowicz, The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge, *Contemp. Oncol.* 19 (1a) (2015) A68–77.
- Zeng, F., Zeng, Q., Wang, S., Wang, S., Liang, W., Huang, Y., Guo, L., Guo, Linc00173 promotes chemoresistance and progression of small cell lung cancer by sponging miR-218 to regulate Etk expression, *Oncogene* (2019) 1–15.
- Yao, N., Yao, Y., Fu, L., Chen, Z., Liu, J., He, Y., Zhu, S., Wang, Long non-coding RNA NONHSAT101069 promotes epirubicin resistance, migration, and invasion of breast cancer cells through NONHSAT101069/miR-129-5p/Twist1 axis, *Oncogene* 38 (47) (2019) 7216–7233.
- Liu, L., Liu, H.J., Wang, T., Meng, C., Lei, X.H., Yang, Q.S., Wang, J.F., Zhu, LncRNA GAS5 inhibits cell migration and invasion and promotes autophagy by targeting miR-222-3p via the GAS5/PTEN-signaling pathway in CRC, *Mol. Ther. Nucleic Acids* 17 (2019) 644–656.
- Q.R. Huang, X.B. Pan, Prognostic lncRNAs, miRNAs, and mRNAs form a competing endogenous RNA network in colon cancer, *Front Oncol.* 9 (2019) 712.
- Long, J., Xiong, J., Mao, H., Zhang, Y., Bai, J., Lin, H., Zhao, Construction and investigation of a lncRNA-associated ceRNA regulatory network in cholangiocarcinoma, *Front Oncol.* 9 (2019) 649.
- Rinne, M., Khan, A., Khan, H., Xhaard, Cartography of rhodopsin-like G protein-coupled receptors across vertebrate genomes, *Sci. Rep-UK* 9 (1) (2019) 7058.
- J.H. Zeng, W. Lu, L. Liang, G. Chen, H.H. Lan, X.Y. Liang, X. Zhu, Prognosis of clear cell renal cell carcinoma (ccRCC) based on a six-lncRNA-based risk score: an investigation based on RNA-sequencing data, *J. Transl. Med.* 17 (1) (2019) 281.
- Cai, J., Cai, Y., Yu, Y., Xu, H., Liu, J., Shou, L., You, W., Han, Exploring the role of Mir204/211 in HNSCC by the combination of bioinformatic analysis of ceRNA and transcription factor regulation, *Oral Oncol.* 96 (2019) 153–160.
- Jiang, P., Jiang, J., Wu, X., Chen, B., Ning, Q., Liu, Z., Li, S., Wang, Quantitative proteomics analysis of differentially expressed proteins in ruptured and unruptured cerebral aneurysms by iTRAQ, *J. Proteom.* 182 (2018) 45–52.
- Chen, F., Chen, Z., Li, C., Deng, H., Yan, Integration analysis for novel lncRNA markers predicting tumor recurrence in human colon adenocarcinoma, *J. Transl. Med.* 17 (1) (2019) 299.
- Xu, F., Tu, N., Xiu, X., Li, microRNA-126-3p attenuates intracerebral hemorrhage-induced blood-brain barrier disruption by regulating VCAM-1 expression, *Front Neurosci-Switz* 13 (2019) 866.
- D.B. Ma, M.M. Qin, L. Shi, X.M. Ding, MicroRNA-6077 enhances the sensitivity of patients-derived lung adenocarcinoma cells to anlotinib by repressing the activation of glucose transporter 1 pathway, *Cell. Signal.* 64 (2019) 109391.
- Cheng, Y., Cheng, T., Cheng, Y., Zhao, Y., Qu, HMG1 exacerbates tumor progression by activating miR-222 through PI3K/Akt/MMP-9 signaling pathway in uveal melanoma, *Cell. Signal.* 63 (2019) 109386.
- Otasek, J.H., Morris, J., Bouças, A.R., Pico, B., Demchak, Cytoscape Automation: empowering workflow-based network analysis, *Genome Biol.* 20 (1) (2019) 1–15.
- J.B. Zou, H.B. Chai, X.F. Zhang, D.Y. Guo, J. Tai, Y. Wang, Y.J. Shi, Reconstruction of the lncRNA-miRNA-mRNA network based on competitive endogenous RNA reveal functional lncRNAs in Cerebral Infarction, *Sci. Rep-UK* 9 (1) (2019) 1–13.
- Z. Wang, B. Qin, Prognostic and clinicopathological significance of long noncoding RNA CTD-2510F5. 4 in gastric cancer, *Gastric Cancer* 22 (4) (2019) 692–704.
- M. Qu, L. Luo, Y. Yang, Y. Kong, D. Wang, Nanopolystyrene-induced microRNAs response in *Caenorhabditis elegans* after long-term and lose-dose exposure, *Sci. Total Environ.* 697 (2019) 134131.
- R.C. Chai, F. Wu, Q.X. Wang, S. Zhang, K.N. Zhang, Y.Q. Liu, C.S. Kang, m6A RNA methylation regulators contribute to malignant progression and have clinical prognostic impact in gliomas, *Aging (Albany NY)* 11 (4) (2019) 1204–1225.
- Y. Bai, J. Long, Z. Liu, J. Lin, H. Huang, D. Wang, H. Zhao, Comprehensive analysis of a ceRNA network reveals potential prognostic cytoplasmic lncRNAs involved in HCC progression, *J. Cell. Physiol.* 234 (10) (2019) 18837–18848.
- A. Santoro, G. Angelico, A. Piermattei, F. Inzani, M. Valente, D., Arciuolo, G. Scambia, Pathological chemotherapy response score in patients affected by high grade serous ovarian carcinoma: the prognostic role of omental and ovarian residual disease, *Front Oncol.* 9 (2019) 778.
- D.T. Quach, T. Hiyama, T. Gotoda, Identifying high-risk individuals for gastric cancer surveillance from western and eastern perspectives: lessons to learn and possibility to develop an integrated approach for daily practice, *World J. Gastroenterol.* 25 (27) (2019) 3546–3562.
- A. Biagioni, I. Skalamera, S. Peri, N. Schiavone, F. Cianchi, E. Giommoni, L. Papucci, Update on gastric cancer treatments and gene therapies, *Canc. Metastasis Rev.* 38 (3) (2019) 537–548.
- Z. Zhao, M. Ji, Q. Wang, N. He, Y. Li, Circular RNA Cdr1as upregulates SCAI to suppress cisplatin resistance in ovarian cancer via miR-1270 suppression, *Mol. Ther. Nucleic Acids* 18 (2019) 24–33.
- D. Zhang, X.J. Yang, Q.D. Luo, D.L. Fu, Z.L. Li, P. Zhang, T. Chong, Down-regulation of circular RNA_000926 attenuates renal cell carcinoma progression through miRNA-411-dependent CDH2 inhibition, *Am. J. Pathol.* 189 (12) (2019) 2469–2486.
- X. Dong, Z. Fang, M. Yu, L. Zhang, R. Xiao, X. Li, J. Liu, Knockdown of long noncoding RNA HOXA-AS2 suppresses chemoresistance of acute myeloid leukemia via the miR-520c-3p/S100A4 axis, *Cell. Physiol. Biochem.* 51 (2) (2018) 886–896.
- T. Lin, P.F. Hou, S. Meng, F. Chen, T. Jiang, M.L. Li, J. Bai, Emerging roles of p53 related lncRNAs in cancer progression: a systematic review, *Int. J. Biol. Sci.* 15 (6) (2019) 1287–1298.
- B. Shamlou, S. Usluer, p21 in cancer research, *Cancers* 11 (8) (2019) 1178.
- F. Wang, S. Liang, X. Liu, L. Han, J. Wang, Q. Du, LINC00460 modulates KDM2A to promote cell proliferation and migration by targeting miR-342-3p in gastric cancer, *OncoTargets Ther.* 11 (2018) 6383–6394.
- Y. Zhu, L. Yang, Q.Y. Chong, H. Yan, W. Zhang, W. Qian, T. Zhu, Long noncoding RNA Linc00460 promotes breast cancer progression by regulating the miR-489-5p/FGF7/AKT axis, *Canc. Manag. Res.* 11 (2019) 5983–6001.
- J.J. Ye, Y.L. Cheng, J.J. Deng, W.P. Tao, L. Wu, Lncrna Linc00460 promotes tumor growth of human lung adenocarcinoma by targeting Mir-302c-5p/Foxa1 axis, *Gene* 685 (2019) 76–84.
- K. Li, D. Sun, Q. Gou, X. Ke, Y. Gong, Y., Zuo, Q. Li, Long non-coding RNA linc00460 promotes epithelial-mesenchymal transition and cell migration in lung cancer cells, *Canc. Lett.* 420 (2018) 80–90.
- Y. Zhang, X. Liu, Q. Li, Y. Zhang, lncRNA LINC00460 promoted colorectal cancer cells metastasis via miR-939-5p sponging, *Canc. Manag. Res.* 11 (2019) 1779–1789.
- X. Wang, F.M. Mo, H. Bo, L. Xiao, G.Y. Chen, P.W. Zeng, Z.H. Chen, Upregulated expression of long non-coding RNA, LINC00460, suppresses proliferation of colorectal cancer, *J. Canc.* 9 (16) (2018) 2834–2843.
- Y. Lian, C. Yan, H. Xu, J. Yang, Y. Yu, J. Zhou, K. Wang, A novel lncRNA, LINC00460, affects cell proliferation and apoptosis by regulating KLF2 and CUL4A expression in colorectal cancer, *Mol. Ther. Nucleic Acids* 12 (2018) 684–697.
- T.P. Xu, P. Ma, W.Y. Wang, Y. Shuai, Y.F. Wang, T. Yu, Y.Q. Shu, KLF5 and MYC modulated LINC00346 contributes to gastric cancer progression through acting as a competing endogenous RNA and indicates poor outcome, *Cell Death Differ.* 26 (2019) 2179–2193.
- Y.Z. Yin, W.H. Zheng, X. Zhang, Y.H. Chen, Y.H. Tuo, LINC00346 promotes hepatocellular carcinoma progression via activating the JAK-STAT3 signaling pathway, *J. Cell. Biochem.* 121 (2020) 735–742.
- T. Ye, W. Ding, N. Wang, H. Huang, Y. Pan, A. Wei, Long noncoding RNA linc00346 promotes the malignant phenotypes of bladder cancer, *Biochem. Biophys. Res. Commun.* 491 (2017) 79–84.
- W. Shi, C. Zhang, Z. Ning, Y. Hua, Y. Li, L. Chen, Z. Meng, Long non-coding RNA LINC00346 promotes pancreatic cancer growth and gemcitabine resistance by sponging miR-188-3p to derepress BRD4 expression, *J. Exp. Clin. Oncol.* 38 (2019) 60.
- W. Zhang, Y. Song, LINC00473 predicts poor prognosis and regulates cell migration and invasion in gastric cancer, *Biomed. Pharmacother.* 107 (2018) 1–6.
- C. Shi, Y. Yang, J. Yu, F. Meng, T. Zhang, Y. Gao, The long noncoding RNA LINC00473, a target of microRNA 34a, promotes tumorigenesis by inhibiting ILF2 degradation in cervical cancer, *Am. J. Cancer Res.* 7 (2017) 2157–2168.
- J. Bai, W.Y. Zhao, W.J. Li, Z.W. Ying, D.Q. Jiang, Long noncoding RNA LINC00473 indicates a poor prognosis of breast cancer and accelerates tumor carcinogenesis by

- competing endogenous sponging miR-497, *Eur. Rev. Med. Pharmacol. Sci.* 23 (2019) 3410–3420.
- [48] Q. Zhang, G. Wang, L. Xu, Z. Yao, L. Song, Long non-coding RNA LINC00473 promotes glioma cells proliferation and invasion by impairing miR-637/CDK6 axis, *Artif Cell Nanomed Biotechnol.* 47 (2019) 3896–3903.
- [49] L. Wang, X. Zhang, L. Sheng, C. Qiu, R. Luo, LINC00473 promotes the Taxol resistance via miR-15a in colorectal cancer, *Biosci. Rep.* 38 (2018). BSR20180790.
- [50] Z. Chen, J.L. Li, S. Lin, C. Cao, N.T. Gimbrone, R. Yang, L. Wu, cAMP/CREB-regulated LINC00473 marks LKB1-inactivated lung cancer and mediates tumor growth, *J. Clin. Invest.* 126 (2016) 2267–2279.
- [51] S. Zhu, W. Fu, L. Zhang, K. Fu, J. Hu, W. Jia, G. Liu, LINC 00473 antagonizes the tumour suppressor miR-195 to mediate the pathogenesis of Wilms tumour via IKK α , *Cell Prolif* 51 (2018), e12416.
- [52] P.B. Han, X.J. Ji, M. Zhang, L.Y. Gao, Upregulation of lncRNA LINC00473 promotes radioresistance of HNSCC cells through activating Wnt/beta-catenin signaling pathway, *Eur. Rev. Med. Pharmacol. Sci.* 22 (2018) 7305–7313.
- [53] J. Mo, B. Li, Y. Zhou, Y. Xu, H. Jiang, X. Cheng, Y. Zhang, LINC00473 promotes hepatocellular carcinoma progression via acting as a ceRNA for microRNA-195 and increasing HMGGA2 expression, *Biomed. Pharmacother.* 120 (2019) 109403.
- [54] W. Chen, Y. Zhang, H. Wang, T. Pan, Y. Zhang, C. Li, LINC00473/miR-374a-5p regulates esophageal squamous cell carcinoma via targeting SPIN1 to weaken the effect of radiotherapy, *J. Cell. Biochem.* 120 (2019) 14562–14572.
- [55] J. Liu, X. Li, M. Wang, G. Xiao, G. Yang, H. Wang, Y. Peng, A miR-26a/E2F7 feedback loop contributes to tamoxifen resistance in ER-positive breast cancer, *Int. J. Oncol.* 53 (4) (2018) 1601–1612.
- [56] X. Bai, J. Wang, L. Huo, Y. Xie, W. Xie, G. Xu, M. Wang, Serine/threonine kinase CHEK1-dependent transcriptional regulation of RAD54L promotes proliferation and radio resistance in glioblastoma, *Trans Oncol.* 11 (2018) 140–146.