

# Comprehensive Analysis of a ceRNA Network Reveals Potential Prognostic lncRNAs Involved in Progression of Bladder Cancer

**Dehua Ou**

The First Affiliated Hospital of Shantou University Medical College <https://orcid.org/0000-0002-1144-3734>

**Zesong Wu**

The First Affiliated Hospital of Shantou University Medical College

**Peilin Shen**

The First Affiliated Hospital of Shantou University Medical College

**Yingkai Hong**

The First Affiliated Hospital of Shantou University Medical College

**Si Chen**

The First Affiliated Hospital of Shantou University Medical College

**Yichuan Huang**

The First Affiliated Hospital of Shantou University Medical College

**Ming-en Lin** (✉ [me\\_lin20@163.com](mailto:me_lin20@163.com))

The First Affiliated Hospital of Shantou University Medical College <https://orcid.org/0000-0001-7801-6780>

---

## Research

**Keywords:** Bladder cancer, ceRNA, lncRNA, risk score, overall survival

**Posted Date:** May 4th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-434872/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** The aberrant expression of long non-coding RNAs (lncRNAs) has attracted more and more attention in the biological field of bladder cancer(BC). We aim to construct a competing endogenous RNA(ceRNA) network reveals potential prognostic lncRNAs involved in progression of BC.

**Results:** Expression profiles of messenger RNA(mRNA), micro RNA(miRNA) and lncRNA of 397 BC samples and 19 non-tumor tissues were downloaded from The Cancer Genome Atlas(TCGA) database. Patients with BC were randomly divided into training group (n=198) and validation group (n=199). Then, 130 lncRNAs, 159 miRNAs and 2,048 mRNAs were identified as differentially expressed genes (DEGs,  $|\log FC| > 1$ ,  $FDR < 0.01$ ) related to BC progression. Next, we constructed an BC associated deregulated competing endogenous RNA(ceRNA) network with 70 lncRNAs, 30 miRNAs, and 62 mRNAs involved in. Subsequently, a seven-lncRNA signature was constructed by establishing a LASSO Cox model with 13 lncRNAs associated with survival from the ceRNA network. This signature can well distinguish high-risk patients from low-risk patients in training group and verification group. Furthermore, we combined the risk score model with other clinical features to estimate the ability of survival prediction. The result suggested that the risk score can be selected as an independent prognostic factor for overall survival(OS) rate.

**Conclusion:** In this study, we construct a ceRNA network related to progression of BC and established an seven-lncRNA signature, which was a candidate prognostic biomarker for prognostic prediction of BC patients .

## Background

In developed countries, BC is the first most incident neoplasm of urinary system. In addition, the incidence of BC is steadily rising worldwide and four-fifths of all BC patients are men[1, 2]. Although the endoscopic treatment, radical surgery, chemotherapy and immunity therapy has progressed vastly, high recurrence rate of BC(30–70%) and poor prognosis, especially for patients with metastatic BC, are still troubling[3]. The 5-year relative survival will decrease from 69.2–36.5% once the tumor transforms localized stage(confined to primary site) to regional stage(spread to regional lymph nodes)[4]. Therefore, exploration of potential molecular mechanisms that may trigger BC and promote BC progression is crucial. And it will help to identify novel prognostic molecular biomarkers and and effective therapeutic targets.

The occurrence of BC is related to the mutation of a variety of genes, like TP53, MLL2, HER-2, HRAS, Bcl-2, etc[5, 6]. In recent years, the important function of non-coding RNAs (ncRNAs) in the occurrence and progression of BC has been demonstrated in accumulating evidence[7]. lncRNAs are ribonucleic acid molecules with a length of more than 200 nucleotides, which have no protein-coding function. Originally, they were thought to be the by-products of RNA polymerase II transcription and noises of gene transcription[8]. However, recent studies suggested that numerous lncRNAs are involved in all essential biological processes in living cells[9]. Aberrantly expressed lncRNAs are related to cell proliferation,

migration, apoptosis, invasion and metastasis in various tumors. Specific lncRNAs are digged out during tumorigenesis and metastasis of BC[10]. For example, experiments in vivo showed that knocking down metastasis-related lung adenocarcinoma transcript 1 (MALAT1) can inhibit the metastasis of BC[11]. In addition, the silencing of taurine up-regulated gene 1 (TUG1) inhibits the proliferation and initiates apoptosis of BC cells[12]. lncRNA-ATB activates PI3K/AKT and mTOR signaling pathways in bladder cancer so that the cell viability, migration and invasion is promoted[13]. All these findings strongly indicated that some specific lncRNAs could be used as prognostic biomarkers and therapeutic targets for BC. The study of lncRNAs in BC and normal bladder epithelial cells will help to expound the pathogenesis of BC.

According to the hypothesis of ceRNA, circular RNAs(circRNAs), lncRNAs, miRNAs and mRNAs are involved in the ceRNA interaction network. miRNAs can bind mRNAs to induce gene silencing, while lncRNAs or circRNAs regulate gene expression by competitively binding miRNAs. CeRNAs can be combined with miRNA through response elements (micro RNA response elements, MREs) to up-regulate genes suppressed by microRNA[14]. CeRNAs usually contain multiple MREs, so each ceRNA can regulate multiple miRNAs, and each miRNA can be also regulated by multiple ceRNAs[15].

Experiments and bioinformatics have identified lncRNA-miRNA-mRNA ceRNA networks in various physiological and pathological processes to initiate and develop carcinoma[14, 16]. However, the lack of large samples and research approaches hinder integrative analyses of the ceRNA mechanisms of lncRNAs in BC progression. Encouragingly, the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) co-established a cancer research project, The Cancer Genome Atlas (TCGA, <https://www.cancer.gov/>) platform, which provides a large, free cancer research reference database by collecting and sorting out various cancer-related omics data. At present, a total of 33 cancer types are collected, more than 2 PB of clinical pathology and corresponding bioinformatics data. The data is free and public, which greatly helps cancer researchers to better investigate the tumorigenesis, recurrence and metastasis of cancer[17].

In the current study, RNA sequencing data of 397 BC tissue samples and 19 adjacent nontumor bladder tissue samples were obtained from TCGA database. Differentially expressed lncRNAs, miRNAs, and mRNAs between early tumor stages (I/II) and normal tissues were screened, and those differentially expressed genes (DEGs) between late tumor stages (III/IV) and normal tissues were identified as well. Subsequently, based on the miRcode, TargetScan, starBase and miRTarBase databases, a ceRNA regulatory network for BC progression was constructed by the intersecting 130 lncRNAs, 159 miRNAs, and 2,048 mRNAs of the two comparison above. Finally, a risk score system containing 7 lncRNAs significantly affecting BC patient prognosis is constructed by means of Lasso penalized Cox proportional hazards regression analysis. Moreover, we analyzed the correlations between the risk score system and clinical characteristics of BC. Our understanding of the function of lncRNAs in the ceRNA network of BC progression will be improved and potential diagnostic biomarkers for BC may be identified with the findings from the present study.

## Results

### Analysis of DElncRNAs, DEmiRNAs, and DEmRNAs

To better understand genes associated with tumorigenesis and progression of BC, the training set was firstly divided into three groups : one group included 19 non-tumor adjacent tissue samples, one comprising 65 relatively early stage(stage I/II, lesions restricted in bladder) BC samples, and one comprising 133 late stage (stage III/IV, lesions exceed bladder wall) BC samples. There was no significant difference in clinical features between the two datasets (all  $P > 0.05$ ). Table 1 lists the detailed features. After extracting the expression matrix from the training set, the differentially expressed RNAs were analyzed with "limma" package of R software.  $|\logFC| > 1$  and  $FDR < 0.01$  were used as cut-off criteria. The DElncRNAs, DEmiRNAs, and DEmRNAs were screened out by comparing the tumor groups with the normal group and visualized by volcano maps (Fig. 2A and 2B). The intersections of the two sets of DEGs consisted of 130 lncRNAs (51 downregulated, 79 upregulated), 159 miRNAs (68 downregulated, 91 upregulated ), and 2048 mRNAs (1387 downregulated, 661 upregulated,)(Fig. 2C. These DElncRNAs, DEmiRNAs and DEmRNAs were considered key genes in tumorigenesis and progression of BC.

Table 1  
Baseline clinical features of 397 bladder cancer cases involved in this study.

Characteristic	Training dataset	Validation dataset	P Value
	N = 198	N = 199	
<b>Age (years)</b>			0.7266
≥ 65	121(61.11%)	126(63.32%)	
< 65	77(38.89%)	73(36.68%)	
<b>Gender</b>			0.7096
Female	54(27.27%)	50(25.13%)	
Male	144(72.73%)	149(74.87%)	
<b>TNM stage</b>			1
I-II	65(32.83%)	65(32.66%)	
III-IV	133(67.17%)	134(67.34%)	
<b>Tumor stage</b>			1
T1-T2	99(50%)	100(50.25%)	
T3-T4	99(50%)	99(49.75%)	
<b>Lymph node metastasis</b>			0.3849
Nx	15(7.58%)	21(10.55%)	
no	116(58.59%)	121(60.80%)	
yes	67(33.84%)	57(28.64%)	
<b>Distant metastasis</b>			0.0757
Mx	102(51.52%)	88(44.22%)	
no	93(46.97%)	103(51.76%)	
yes	3(1.52%)	8(4.02%)	
Notes: Age, the age of patient at diagnosis.			

## Construction of a lncRNA-miRNA-mRNA regulatory network

To investigate the regulatory interaction of the intersecting 130 specific DElncRNAs, 159 DEmiRNAs and 2,048 DEmRNAs above, lncRNA-miRNA interactions were predicted with miRcode and miRNA-mRNA pairs were acquired by all three databases(TargetScan, Starbase and miRTarBase). However, based on the expression of miRNAs in tumor reverse to that of target lncRNAs and mRNAs, we only select the lncRNA-

miRNA-mRNA interaction as its expression pattern is up-down-up-regulation or down-up-down-regulation in BC tissues. Finally, the lncRNA-miRNA-mRNA ceRNA network was constructed. And it was visualized with Cytoscape. The final BC ceRNA regulatory network included 70 DElncRNAs, 30 DEmiRNAs, and 62 DE mRNAs, consisting of 162 nodes and 421 interactions (Fig. 3A and Table S1).

## Functional enrichment analysis

To explore the functions of differentially expressed genes in BC occurrence and progression, 62 DE mRNAs in the newly formed ceRNA network were analyzed using the GO terms and the KEGG pathways. The top GO results (including 10 biological processes (BPs), 4 cellular components (CCs) and 10 molecular functions (MFs)) and 10 KEGG pathways are shown in Fig. 3B and 3D. GO analysis demonstrated that these mRNAs were highly related to binding multiple biological factors in molecular function, including activating transcription factor binding, beta-catenin binding, SMAD binding, transmembrane receptor protein kinase activity and cytokine receptor binding. For the cellular component, the mRNAs were highly associated to apical dendrite and transcription-related complex that located in cell nucleus. Meanwhile, these mRNAs were found mainly enriched in cell cycle G1/S phase transition and epithelial cell proliferation in biological processes.  $P < 0.05$  and Benjamini corrected  $P < 0.05$  were set as filter criteria for the GO enrichment networks. The results of KEGG pathway enrichment analysis suggested that cancer-related pathways were the most important pathways for these selected genes (microRNA in cancer, colorectal cancer, prostate cancer, breast cancer, gastric cancer and melanoma). The other pathways also involved many screened genes including hepatitis B, signaling pathways regulating pluripotency of stem cell and cellular senescence.

## Survival analysis of genes in the ceRNA network

The training set was firstly analyzed to identify the potential lncRNA, miRNA and mRNA biomarkers which were strongly associated with the prognosis of BC patients. We conducted K-M survival analyses for lncRNAs, miRNAs, and mRNAs in the ceRNA network to analyze the impacts of each gene. At last, we screened out 7 lncRNAs, 3 miRNAs, and 11 mRNAs which were associated with OS of BC patients (Table S2). According to rank of the relationship between the expression level and prognosis of BC patients, K-M survival curves of top three lncRNAs, miRNAs, and mRNAs were shown in Fig. 4A, 4B, and 4C, respectively.

## Establishment of BC prognostic signatures

In the ceRNA regulatory network, lncRNAs play crucial roles as initial effectors to influence miRNAs and mRNAs downstream. Therefore, lncRNAs may be ideal diagnostic or prognostic biomarkers due to their status and highly specific expression. After extracting the prognostic information of the training set, univariate Cox regression analysis were performed for the 70 lncRNAs in the ceRNA network to identify lncRNAs which were associated with OS of BC patients. With cut-off for relevance set at  $P < 0.05$ , 12 lncRNAs were considered significantly correlated with overall survival (Table S3). Lasso-penalized Cox regression for these 12 lncRNAs were then performed to determine the independent predictors for

prognosis of BC patients (Fig. 5A and 5B). Ultimately, a 7-lncRNA signature prognostic model with OS was constructed:  $AC004803.1 + AC009690.2 + AC055713.1 + AC105942.1 + BX890604.1 + MIR200CHG + RNF139-AS1$ , whose contributions were weighted by their relative coefficients. The risk assessment score for predicting OS was calculated as follows:  $Risk\ score = (0.2203 \times Expression_{AC004803.1}) + (-0.1537 \times Expression_{AC009690.2}) + (-0.1201 \times Expression_{AC055713.1}) + (0.1202 \times Expression_{AC105942.1}) + (-0.0886 \times Expression_{BX890604.1}) + (-0.1562 \times Expression_{MIR200CHG}) + (-0.2370 \times Expression_{RNF139-AS1})$ . The distribution of risk scores for each BC patient were shown in Fig. 5C. The maximally selected rank statistics were estimated. Based on the cut-off for risk score set at “-3.2118”, the patients were classified into low-risk group (n = 99) and high-risk group (n = 99) (Fig. 5D). According to K-M survival and time-dependent ROC curve analyses shown in Fig. 5E, these two groups were well distinguished regarding overall survival. Compared with the low-risk group, the OS survival rate of the high-risk group tended to become lower ( $P < 0.001$ ). We conducted ROC curve analysis to analyze the accuracy of the 7-lncRNA-based risk score model. The value of area under curve (AUC) at 1, 3, 5 years was 0.665, 0.675 and 0.661 (Fig. 5F). Figure 6A reveals a scatter plot of the risk scores, the corresponding OS of 198 BC patients and the expression heat map of the 7 lncRNAs.

## Verification for survival prediction of the risk score system

The validation set (n = 199) was then used to test the ability of the 7-lncRNA signature for survival prediction. After applying the cut-off value of the 7-lncRNA signature to the validation set, the result of analysis was consistent with the findings of training set. In the validation set, K-M survival analysis suggested that the prognosis of BC patients was obviously worse in the high-risk group (n = 100) than in the low-risk group (n = 99) ( $P = 0.037$ ) (Fig. 5G). The AUC of the risk score signature at 1, 3, 5 years was 0.619, 0.646 and 0.646 in the validation set (Fig. 5H). Figure 6B reveals a scatter plot of the risk scores, the corresponding OS of 198 BC patients and the expression heatmap of the 7 lncRNAs of validation set. Thus, the desired accuracy prediction ability of this prognostic model was shown in both the training set and the validation set.

## Prognostic value of the risk score system

Clinical information of all BC patients involved in this study was obtained through TCGA. Univariate and multivariate Cox regression analyses were subsequently conducted to estimate the ability of prognostic prediction of the 7-lncRNA signature in training set, validation set and entire set respectively. The risk score, N stage, and pathological stage were significantly associated with OS in BC patients ( $P < 0.05$ ) all for the three datasets according to the result of univariate analysis. However, only the risk score and N stage were considered independent prognostic factors of OS by multivariate analysis ( $P < 0.05$ , Table 2).

## Relationship of lncRNAs and mRNAs

Based on the ceRNA theory, the expression of mRNAs was positively regulated by lncRNAs, which act as sponges to interact with miRNAs. To further analyze the relationship of lncRNAs and mRNAs in BC, we

conducted regression analysis of the 7 lncRNAs included in the risk score signature and 11 mRNAs significantly associated with overall survival. Then, 25 lncRNA-mRNA pairs were found positive correlations (Fig. 7A and Table S4). Subsequently, the shared miRNAs existing between the lncRNAs and mRNAs were investigated. Finally, five ceRNA pathways were digged out, including BX890604.1-miR-29a-3p-TRAF4, AC105942.1-miR-182-5p-RECK, AC105942.1-miR-31-5p-STK40, RNF139-AS1-miR-29a-3p-TRAF4, AC004803.1-miR-503-5p-RECK (Fig. 7B).

## Discussion

As one of the most common tumors of urinary system, bladder cancer carried a poor prognosis with hundreds of thousands of new cases and deaths every year [2]. Currently, clinician usually take the TNM staging system as the most important standard for prognostic predictions for BC patients, which was developed by the American Joint Committee on Cancer [18]. However, the TNM system distinguishes different cancer stages by anatomical disease progression. It is difficult to explain why patients with same TNM stage exhibit variable responses to the similar therapy. This phenomenon prompts us to pay attention on the genomic heterogeneity of malignant lesions. Therefore, we aimed to screen several effective biomarkers for earlier BC diagnosis and prognosis prediction.

According to the accumulating evidence, lncRNAs are involved in tumorigenesis and cancer progression. lncRNAs participate in cell proliferation, apoptosis, invasion, metastasis and immune response [19–21]. Without function of protein coding, lncRNAs are attracting as miRNA sponges or ceRNAs to regulate post-transcriptional processes or gene expression indirectly [22]. In recent years, various ceRNA networks have been reported in which lncRNAs regulate mRNA expression by interacting with miRNAs. As a miRNA sponge distributed mostly in the cytoplasm, the up-regulated expression of lncRNA DANCR is detected in BC. It positively regulates the expression of MSI2 by adsorbing miR-149, and thus promotes BC cell proliferation, migration, invasion and tumorigenicity [23]. HOTAIR is a well-known lncRNA involved in various tumors. In breast cancer, the high expression of HOTAIR inhibits cell apoptosis, facilitates cell growth and metastasis through miR-20a-5p/HMGA2 axis [24]. What's more, due to the initial status in ceRNA network, lncRNAs can be taken as potential diagnostic or prognostic biomarkers for cancer patients. In the hepatocellular carcinoma (HCC), a ceRNA network with 37 lncRNAs, 10 miRNAs, and 26 mRNAs included was constructed based on TCGA data. And then by means of multivariate Cox regression and lasso analysis, a risk score system consisting of 13 lncRNAs was established for prognostic prediction of patients with HCC [25]. Nevertheless, the ceRNA network of BC progression and the prognostic values of lncRNA signature in the ceRNA network have been rarely investigated. Therefore, the study of RNA interactions in BC progression will help us to a better understand the biological processes in tumor.

In the present study, RNA sequencing data of BC was firstly extracted from TCGA database. We randomly divided the BC patients into two group, training set and validation set. Differentially expressed lncRNAs, miRNAs, and mRNAs were screened by comparing early tumor stages (I/II) tissues with normal tissues, late tumor stages (III/IV) with normal tissues respectively according to data of training set. Then the

overlapping 130 lncRNAs, 159 miRNAs, and 2,048 mRNAs in were identified. We screened the lncRNA-miRNA-mRNA chain as its expression pattern is up-down-up-regulation or down-up-down-regulation in BC tissues after predicting the interactions between lncRNA and miRNA, miRNA and mRNA. Then a BC-associated ceRNA regulatory network including 70 DElncRNAs, 30 DEmiRNAs, and 62 DEmRNAs was constructed. Next, the functions and signaling pathways in which these dysregulated mRNAs were involved were determined by means of GO and KEGG analyses. We performed K-M survival analysis to dig out the RNAs in this network which were significantly associated with OS. Subsequently, we established a 7-lncRNA-based risk prediction model (AC004803.1, AC009690.2, AC055713.1, AC105942.1, BX890604.1, MIR200CHG, RNF139-AS1) via univariate Cox regression and LASSO regression analysis. According to the risk score, the patients were divided into low-risk and high-risk groups in the K-M survival analysis. For the time-dependent ROC curve, the AUC values of one year, three years and five years were 0.665, 0.675 and 0.661 respectively, showing moderate efficacy in prognostic prediction. Then we verified the prognostic value of the multiple lncRNAs signature in the validation set. We also conducted univariate and multivariate Cox regression analyses to combine the risk score system and clinical characteristics, and the result suggested a good predictive power of the risk score, which can be an independent factor for prognosis. Finally, based on ceRNA network theory, the correlation between lncRNAs and mRNAs was explored.

Through GO and pathway analyses, we can explore the functions of dysregulated mRNAs in the ceRNA network. Several GO terms of the differentially expressed mRNAs belonged predominantly to the following categories: "epithelial cell proliferation", "mesenchymal cell proliferation", "cell cycle G1/S phase transition", "regulation of cell cycle G1/S phase transition" and "lymphocyte differentiation", suggesting that the alteration of cell cycle, cell proliferation and immune response may play crucial roles in BC progression. The result of KEGG pathway analysis showed that mRNAs were most enriched in "microRNA in cancer", "breast cancer", "gastric cancer", "hepatitis B", "proteoglycans in cancer", "signaling pathways regulating pluripotency of stem cells", "cell senescence", "melanoma", "colorectal cancer" and "prostate cancer". This indicated that different cancer types may share the common abnormal signaling pathways.

Among the 7 lncRNAs involved in the signature model, limited studies were performed on the role of MIR200CHG. Previously, an up-regulated expression of MIR200CHG was demonstrated in colon adenocarcinoma. MIR200CHG was identified as an immune-related lncRNA and independent biomarker for prognostic prediction of patients with colon adenocarcinoma[26]. MIR200CHG was also included in a prognostic four-lncRNA based Risk Score evaluation model for urothelial bladder carcinoma. The up-regulated expression of MIR200CHG was observed in tumor than normal tissues in bladder cancer. However, the Risk Score system showed that MIR200CHG acts as a protective factor in BC prognosis [27]. The result may be contributed to the various functional mechanisms of lncRNAs, which regulated expression of multiple mRNAs by competitively bind miRNAs. In our study, the upregulated expression of MIR200CHG and its protective action in BC progression was consistent with the study above. This indicates that different regulatory function of MIR200CHG may be involved in BC tumorigenesis and progression.

In another study, Hang et al.[28] selected AC105942.1 as an EMT-related lncRNA of which the expression is negatively related to BC prognosis. Consistently, AC105942.1 is a risky factor in BC prognosis in our study. According to the correlation test, AC105942.1 showed high positive correlations with the RECK mRNA via miR-182-5p, and STK40 mRNA via miR-31-5p. No reports that address the roles of AC004803.1, AC009690.2, AC055713.1, BX890604.1 and RNF139-AS1, but these lncRNAs were speculated to be involved in BC oncogenesis and progression.

Several limitations of our study remain to be solved. First of all, the sequencing data of a small amount of tumor-free normal bladder tissues from TCGA may cause some error for DEGs identification. More samples of control group should be included in the study to pledge the accuracy of analysis. Second, external validation was not performed. The valuable of the risk model could be proved with more data available in other databases. Third, the functions of most lncRNAs in the risk model remain unclear because of the lack of relevant literature. And thus, the functions of unreported lncRNAs should be evaluated via a series of exploratory experiments.

## Conclusions

In conclusion, we identified dysregulated lncRNAs, miRNAs and mRNAs related to the occurrence and development of BC, and constructed a BC-associated ceRNA regulatory network. By univariate Cox regression and LASSO regression analysis, we constructed a seven lncRNA prognostic model for prognostic prediction of BC patients. According to further analysis, the 7-lncRNA signature may be an independent factor for prognosis BC patients. The prognostic model presented satisfying accuracy in OS prediction. The practicability of the prognostic model are necessary to be verified by more extensive investigations.

## Methods

### RNA sequence data acquisition

The flow chart of our study was showed in Fig. 1. RNA sequence data (lncRNA and mRNA), miRNA expression profiles of 414 bladder cancer samples and 19 non-tumor adjacent tissue samples and corresponding clinical information were downloaded from the TCGA database (<https://portal.gdc.cancer.gov/>). The exclusion criteria were (1) a pathologic diagnosis (Ta and Tis) or histological diagnosis denying BC, (2) lack of complete prognostic data, (3) existence of other malignancies. Ultimately, we obtained 19 adjacent non-tumor bladder tissue samples and 397 tumor tissue samples for further analysis. Then, we divided 397 patients into training and validation groups randomly based on their batch numbers. No ethical approval was necessary because our study was based on open source data.

### Data processing and identification of DEGs

After annotating the genes for lncRNA and mRNA, we standardized the expression profiles among samples. Then, the “limma” package of R software was applied for differentially expressed lncRNAs,

miRNAs, and mRNAs analyses with two comparisons: early stage (stage I/II) BC samples versus adjacent non-tumor samples and late stage (stage III/IV) BC samples versus adjacent non-tumor samples. The cut-off criteria were set as  $|\log FC| > 1$  and  $FDR \leq 0.01$ . The “ggplot2” package was used to visualize the results with volcano plots. The intersection of DEGs between two comparisons was visualized by Venn diagrams for further analysis.

## Construction of the ceRNA network

The dysregulated lncRNA-miRNA-mRNA ceRNA network was constructed according to the hypothesis that lncRNAs act as miRNA sponges to regulate mRNA[14]. The lncRNA-miRNA interactions were predicted by means of miRcode(<http://www.mircode.org/>; ). The target mRNAs of miRNAs were collectively predicted using TargetScan (<http://www.targetscan.org/>), miRTarBase (<http://mirtarbase.mbc.nctu.edu.tw/>) and StarBase (<http://starbase.sysu.edu.cn/>) databases. Since expression of miRNAs in tumor is reverse to that of their ceRNAs and target mRNAs, we only select the lncRNA-miRNA-mRNA chain as its expressed pattern is upregulation-downregulation-upregulation or downregulation-upregulation-downregulation in BC tissues. Then the ceRNA network was visualized with Cytoscape software.

## Function and pathway enrichment of DEmRNAs

The “clusterProfiler” package of R was applied to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) to analyze the DEmRNAs. A GO analysis covered three domains: molecular functions (MFs), cellular components (CCs) and biological processes (BPs). KEGG was used to predict possible pathways. The significant GO terms were identified by Fisher’s test, and GO categories with  $P < 0.05$  were selected to be the enriched functions. The possible signaling pathways at the significance level set ( $P < 0.05$ ) were searched for with KEGG.

## Kaplan-Meier analysis

The “survival” package of R was used to perform Kaplan–Meier (K–M) survival analyses to explore the association between the DElncRNA, DEmiRNA, DEmRNA and overall survival(OS) of patients individually. We divided the patients into 2 groups based on median expression of the DEGs for the OS analysis using log-rank test with  $P < 0.05$  considered statistically significant.

## Construction of the risk score system

To verify the accuracy of risk score system, the patients were randomly divided into two groups according to the ratio of 1:1 (the training set = 198 and validation set = 199) by hierarchical grouping. Chi-square test was conducted to analyse the differences of clinical features between two groups. Then, we performed Lasso-penalized Cox regression to identified the DElncRNAs related to survival of BC patients to simplify the model. Firstly, the penalized maximum likelihood method was applied to generate a Cox model. Subsequently, we used ten-fold cross-validation to screen out the best lambda which was used to assure the mean cross-validated error minimization. And it was also applied in calculating the regression coefficients ( $\beta$ ) of the multivariate Cox regression model. Ultimately, a multiple lncRNAs prognostic risk score system was constructed. The risk score formula for predicting OS was constructed as follows:

$$\text{Risk Score} = \sum_{i=1}^n (C_i \times \text{Exp}_i)$$

where  $n$  is the number of lncRNAs involved,  $\text{Exp}_i$  is the expression value of each lncRNA and  $C_i$  is the corresponding estimated regression coefficient from the multivariate Cox regression analysis. According to the median risk score of the training set, the patients were divided into low-risk and high-risk group both in the training set and the validation set. The abilities of OS prediction and risk discrimination of the risk score system were evaluated by means of and K-M survival analysis and time-dependent receiver operating characteristic (ROC) curves.

## **Regression analysis of risk score and clinical characteristics**

In order to test whether the clinical characteristics and risk score level of BC patients were significantly related to OS, we conducted univariate and multivariate Cox regression analysis in training set and validation set.  $P < 0.05$  was considered statistically significant. Calculation of the hazard ratio and 95% confidence intervals for every factor was performed as well.

## **Correlation analysis between DElncRNAs and DEmRNAs**

To estimate the expressive correlation among DElncRNAs and DEmRNAs, we conducted regression analysis of DElncRNAs and DEmRNAs. The “ggpubr”, “tidyverse”, “corrplot” packages of R were used to visualize the results.  $P < 0.05$  and  $r > 0.25$  were considered a significant relation between DElncRNA and DEmRNA expression.

## **Declarations**

## **ETHICS APPROVAL AND CONSENT TO PARTICIPATE**

Our study is based on open source data, so there are no ethical issues and other conflicts of interest.

## **CONSENT FOR PUBLICATION**

Not applicable.

## **AVAILABILITY OF DATA AND MATERIALS**

The datasets analyzed for this study can be found in the TCGA database ([cancergenome.nih.gov](http://cancergenome.nih.gov)).

## **COMPETING INTERESTS**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## AUTHOR CONTRIBUTIONS

Dehua Ou and Ming-en Lin had the idea for this study. Zesong Wu supervised the acquisition of the data. Dehua Ou and Peilin Shen undertook the statistical analysis. Dehua Ou wrote the first draft of the manuscript. Yichuan Huang, Yingkai Hong and Si Chen commented on and critically revised the manuscript. All authors approved the final version of the manuscript.

## FUNDING

This work was supported by the Guangdong Medical Research Foundation (A2020223) and Supporting Program of the First Affiliated Hospital of Shantou University Medical College (202003-3).

## ACKNOWLEDGEMENTS

We acknowledge TCGA database for providing their platforms and contributors for uploading their meaningful datasets.

## References

1. Antoni S, Ferlay J, Soerjomataram I, Znaor A, Jemal A, Bray F. Bladder Cancer Incidence and Mortality: A Global Overview and Recent Trends. *Eur Urol*. 2017;71:96–108.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394–424.
3. Flaig TW, Spiess PE, Agarwal N, Bangs R, Boorjian SA, Buyyounouski MK, et al. NCCN Guidelines Insights: Bladder Cancer, Version 5.2018. *J Natl Compr Canc Netw*. 2018;16:1041–53.
4. DeGeorge KC, Holt HR, Hodges SC. Bladder Cancer: Diagnosis and Treatment. *Am Fam Physician*. 2017;96:507–14.
5. Gui Y, Guo G, Huang Y, Hu X, Tang A, Gao S, et al. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat Genet*. 2011;43:875–8.
6. Rentsch CA, Muller DC, Ruiz C, Bubendorf L. Comprehensive Molecular Characterization of Urothelial Bladder Carcinoma: A Step Closer to Clinical Translation? *Eur Urol*. 2017;72:960–1.
7. Porten SP. Epigenetic Alterations in Bladder Cancer. *Curr Urol Rep*. 2018;19:102.
8. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet*. 2016;17:47–62.

9. Fatima R, Akhade VS, Pal D, Rao SM. Long noncoding RNAs in development and cancer: potential biomarkers and therapeutic targets. *Mol Cell Ther.* 2015;3:5.
10. Audas TE, Lee S. Stressing out over long noncoding RNA. *Biochim Biophys Acta.* 2016;1859:184–91.
11. Fan Y, Shen B, Tan M, Mu X, Qin Y, Zhang F, et al. TGF-beta-induced upregulation of malat1 promotes bladder cancer metastasis by associating with suz12. *Clin Cancer Res.* 2014;20:1531–41.
12. Liu Q, Liu H, Cheng H, Li Y, Li X, Zhu C. Downregulation of long noncoding RNA TUG1 inhibits proliferation and induces apoptosis through the TUG1/miR-142/ZEB2 axis in bladder cancer cells. *Onco Targets Ther.* 2017;10:2461–71.
13. Zhai X, Xu W. Long Noncoding RNA ATB. Promotes Proliferation, Migration, and Invasion in Bladder Cancer by Suppressing MicroRNA-126. *Oncol Res.* 2018;26:1063–72.
14. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell.* 2011;146:353–8.
15. Zhang P, Wu W, Chen Q, Chen M. Non-Coding RNAs and their Integrated Networks. *J Integr Bioinform.* 2019;16.
16. Karreth FA, Pandolfi PP. ceRNA cross-talk in cancer: when ce-bling rivalries go awry. *Cancer Discov.* 2013;3:1113–21.
17. Hutter C, Zenklusen JC. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. *Cell.* 2018;173:283–5.
18. Roupert M, Babjuk M, Burger M, Capoun O, Cohen D, Comperat EM, et al. European Association of Urology Guidelines on Upper Urinary Tract Urothelial Carcinoma: 2020 Update. *Eur Urol.* 2020.
19. Chen C, Luo Y, He W, Zhao Y, Kong Y, Liu H, et al. Exosomal long noncoding RNA LNMAT2 promotes lymphatic metastasis in bladder cancer. *J Clin Invest.* 2020;130:404–21.
20. Chen ZZ, Huang L, Wu YH, Zhai WJ, Zhu PP, Gao YF. LncSox4 promotes the self-renewal of liver tumour-initiating cells through Stat3-mediated Sox4 expression. *Nat Commun.* 2016;7:12598.
21. Liu G, Ye Z, Zhao X, Ji Z. SP1-induced up-regulation of lncRNA SNHG14 as a ceRNA promotes migration and invasion of clear cell renal cell carcinoma by regulating N-WASP. *Am J Cancer Res.* 2017;7:2515–25.
22. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. *Nature.* 2014;505:344–52.
23. Zhan Y, Chen Z, Li Y, He A, He S, Gong Y, et al. Long non-coding RNA DANCR promotes malignant phenotypes of bladder cancer cells by modulating the miR-149/MSI2 axis as a ceRNA. *J Exp Clin Cancer Res.* 2018;37:273.
24. Zhao W, Geng D, Li S, Chen Z, Sun M. LncRNA HOTAIR influences cell growth, migration, invasion, and apoptosis via the miR-20a-5p/HMGA2 axis in breast cancer. *Cancer Med.* 2018;7:842–55.
25. Bai Y, Long J, Liu Z, Lin J, Huang H, Wang D, et al. Comprehensive analysis of a ceRNA network reveals potential prognostic cytoplasmic lncRNAs involved in HCC progression. *J Cell Physiol.* 2019;234:18837–48.

26. Li Z, Wang D, Yin H. A seven immune-related lncRNA signature predicts the survival of patients with colon adenocarcinoma. *Am J Transl Res.* 2020;12:7060–78.
27. He RQ, Huang ZG, Li TY, Wei YP, Chen G, Lin XG, et al. RNA-Sequencing Data Reveal a Prognostic Four-lncRNA-Based Risk Score for Bladder Urothelial Carcinoma: An in Silico Update. *Cell Physiol Biochem.* 2018;50:1474–95.
28. Tong H, Li T, Gao S, Yin H, Cao H, He W. An epithelial-mesenchymal transition-related long noncoding RNA signature correlates with the prognosis and progression in patients with bladder cancer. *Biosci Rep.* 2021;41.

## Tables

Due to technical limitations, table 2 is only available as a download in the Supplemental Files section.

## Figures

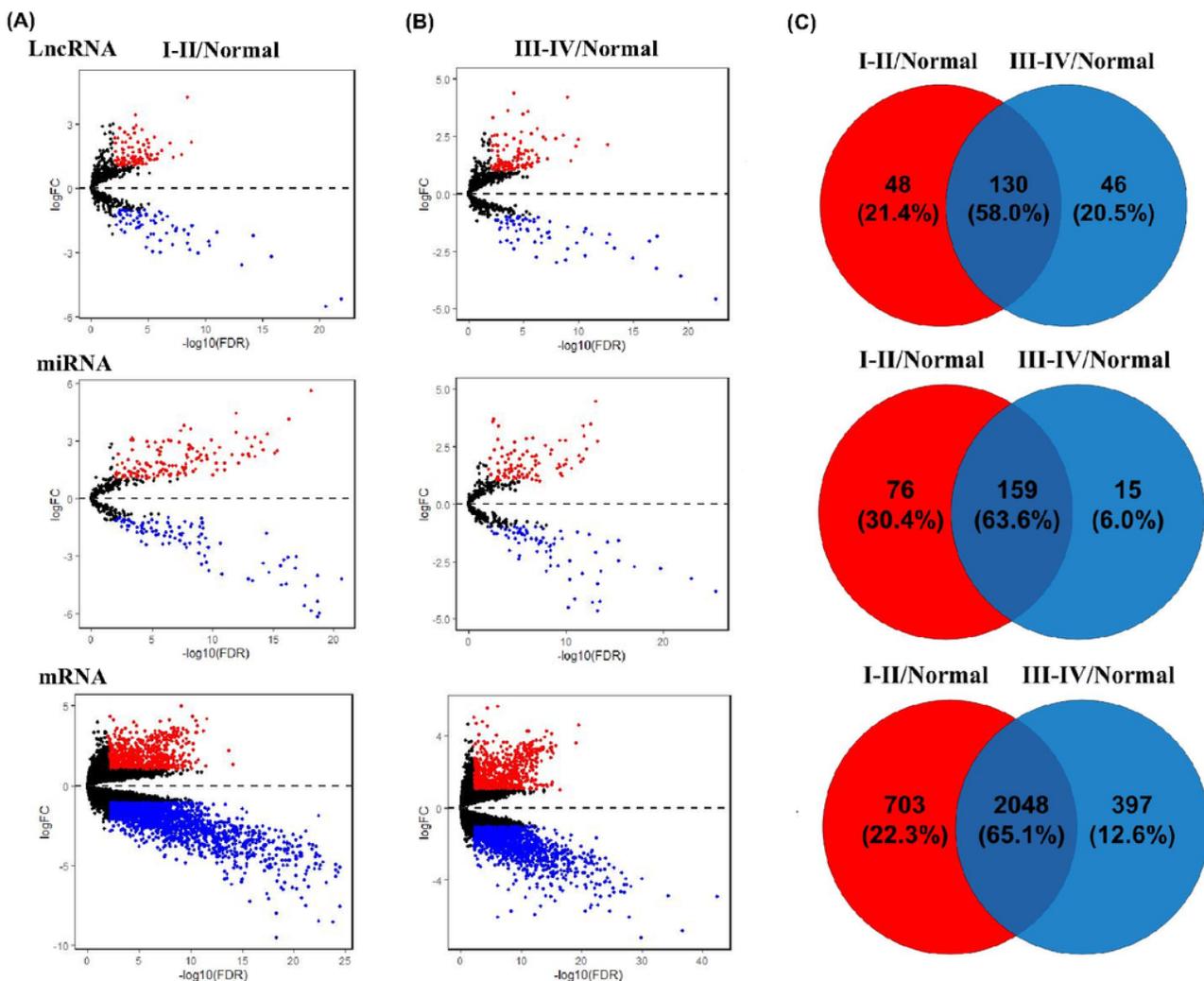
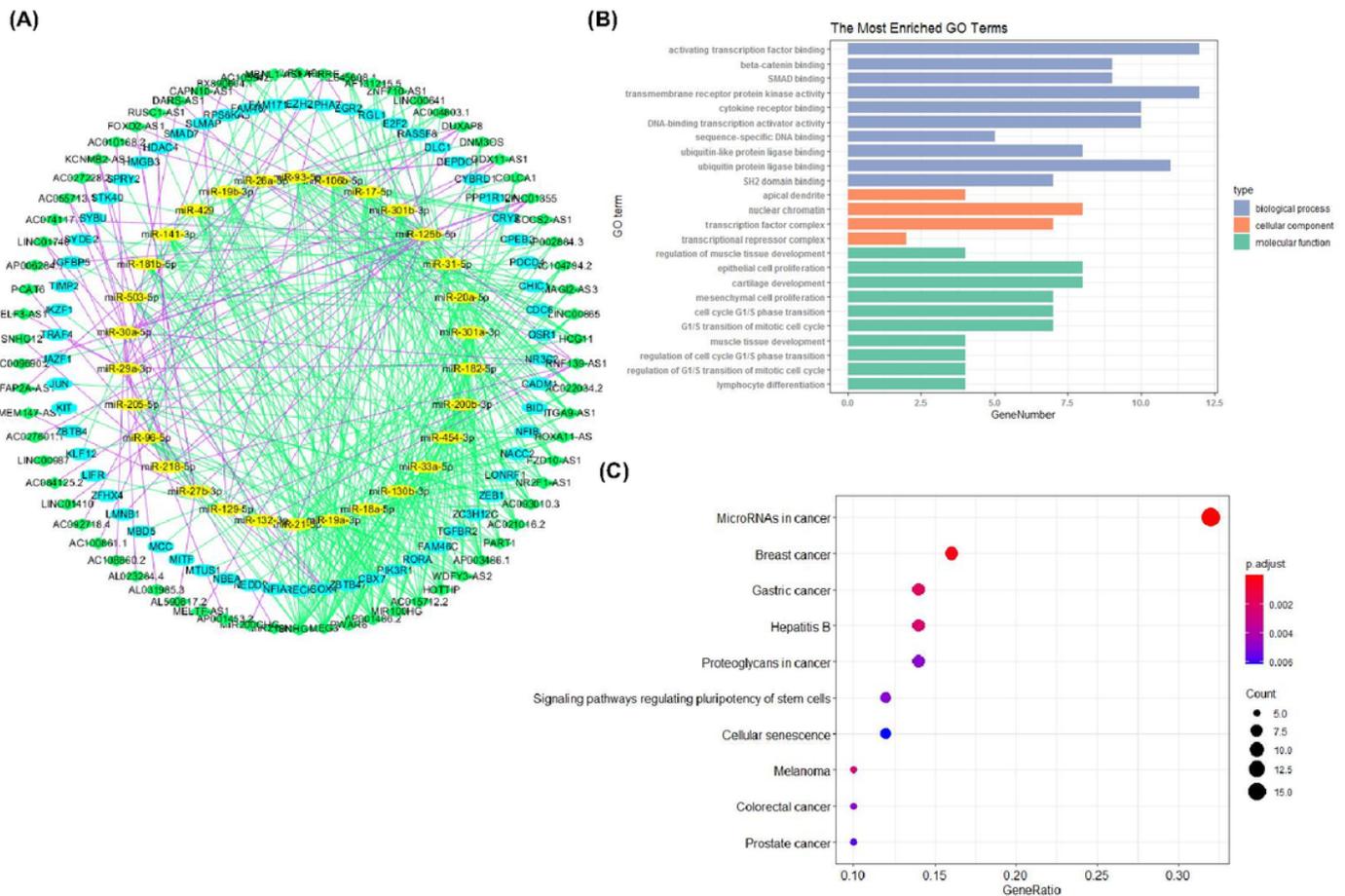


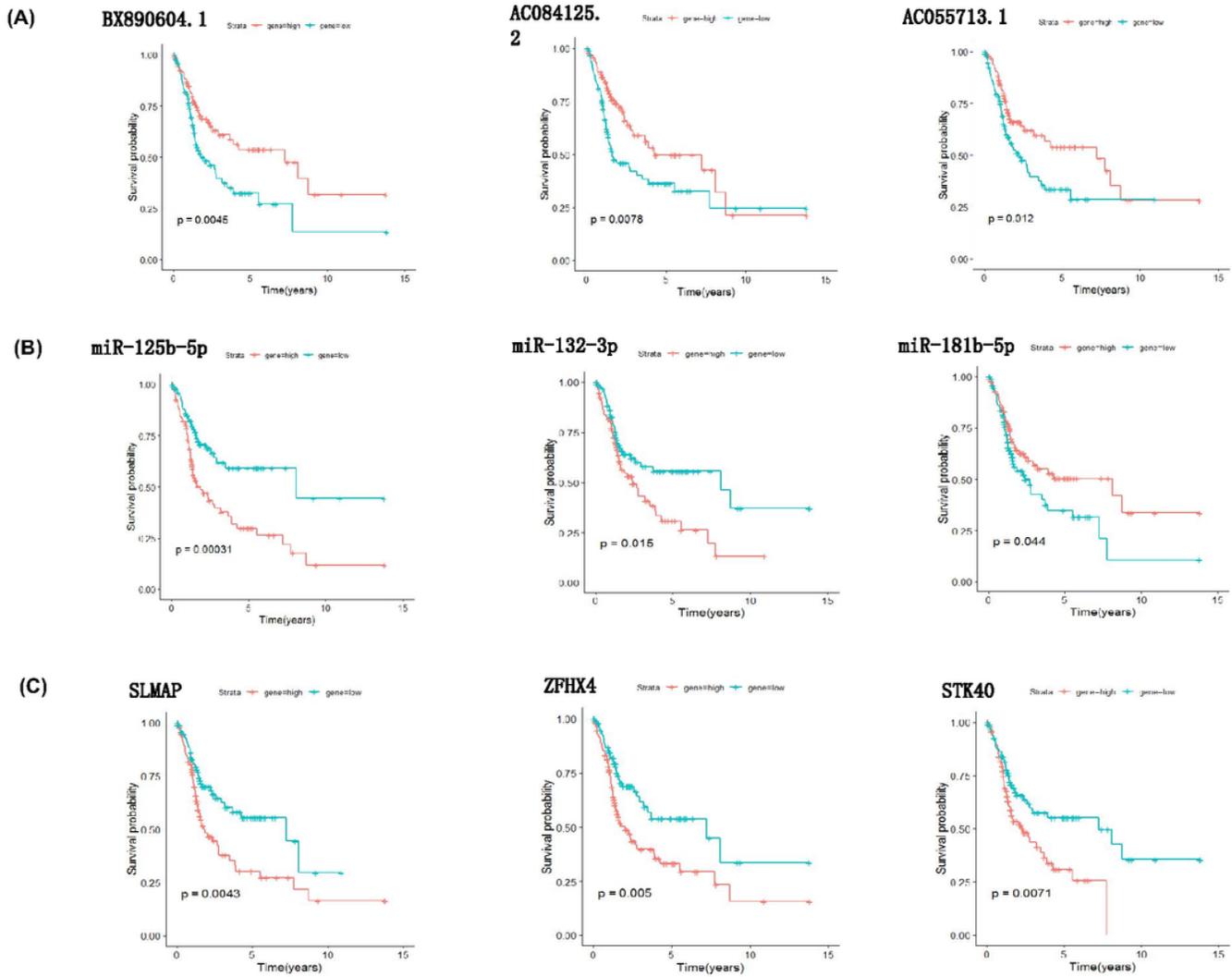
Figure 1

Flow chart of the analysis in this study. N: adjacent nontumor tissues; I/II: stage I or II; III/IV: stage III or IV; ceRNA: competitive endogenous RNA; lncRNA: long noncoding RNA; miRNA: microRNA; mRNA:messenger RNA; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes.



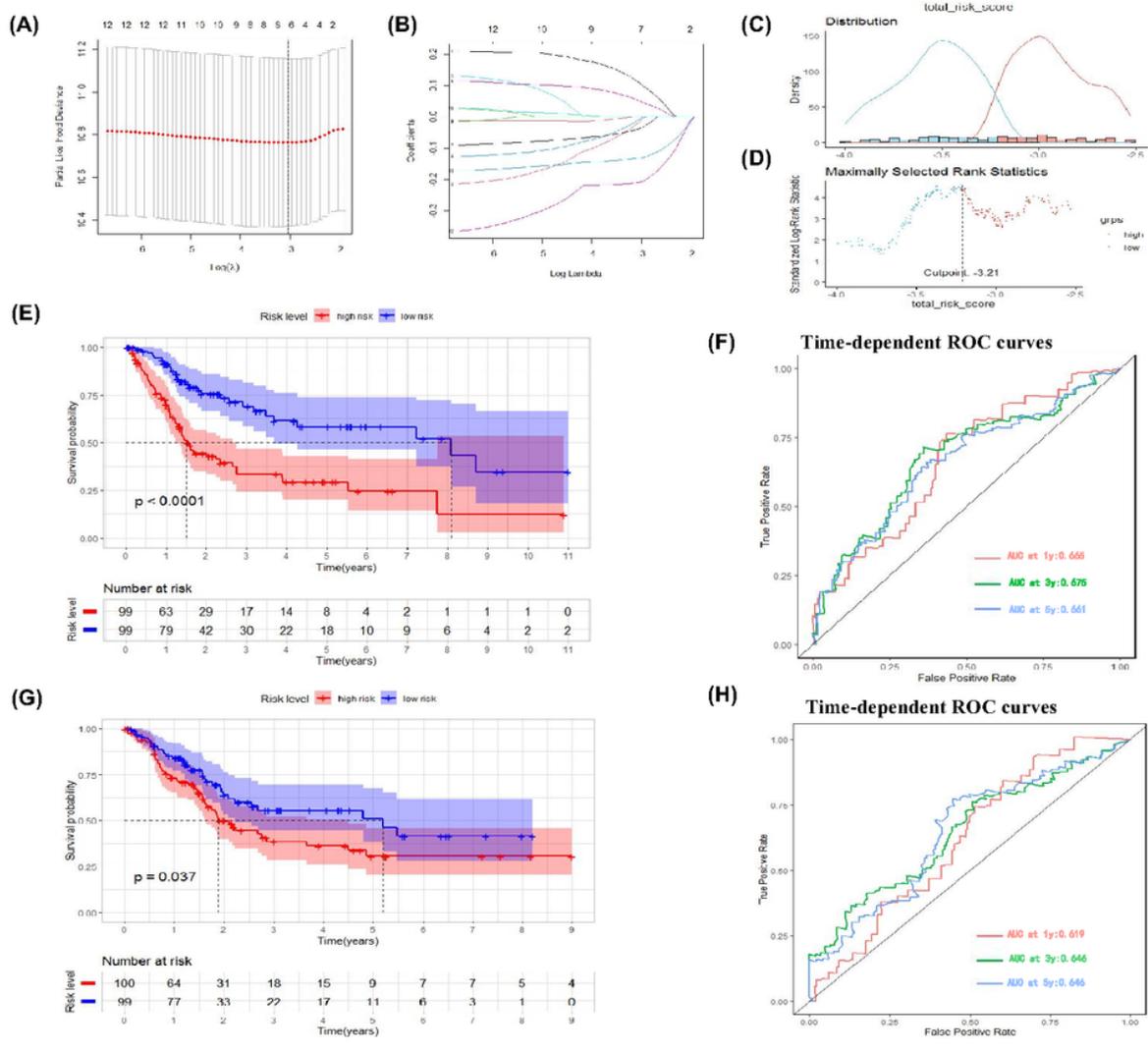
**Figure 2**

Identification of differentially expressed genes. Volcano maps of aberrantly expressed lncRNAs (above graph), miRNAs (medium graph), and mRNAs (below graph) between two groups: early stage(stage I or II) BC samples versus normal samples (A); late stage (stage III or IV) BC samples versus normal samples(B). Green dots are shown as downregulated genes, and red dots are shown as upregulated. Venn diagrams represent the overlapping differentially expressed genes (C). Red areas are derived from left volcano maps, blue areas are derived from right volcano maps. FC: fold change; lncRNAs: long noncoding RNAs; miRNA: microRNA; mRNA: messenger RNA.



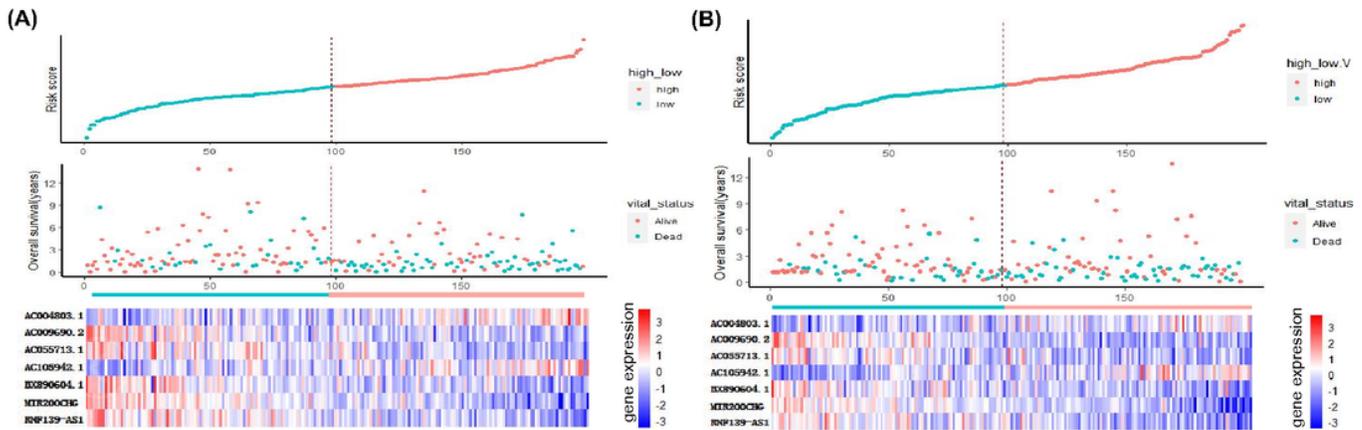
**Figure 3**

Integrated analysis of ceRNA network. The ceRNA network derived from DEGs (A). Green rhombi represent 37 lncRNAs; yellow ovals represent 10 miRNAs; cyan hexagons represent 26 mRNAs. Purple lines represent the expressed pattern of lncRNA-miRNA-mRNA as upregulation-downregulation-upregulation in BC samples; Green lines represent the expressed pattern of lncRNA-miRNA-mRNA or downregulation-upregulation-downregulation in BC samples. Top GO terms of the DEmRNAs in the ceRNA network(B). Top KEGG pathways of the DEmRNAs in the ceRNA network(C). GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes.



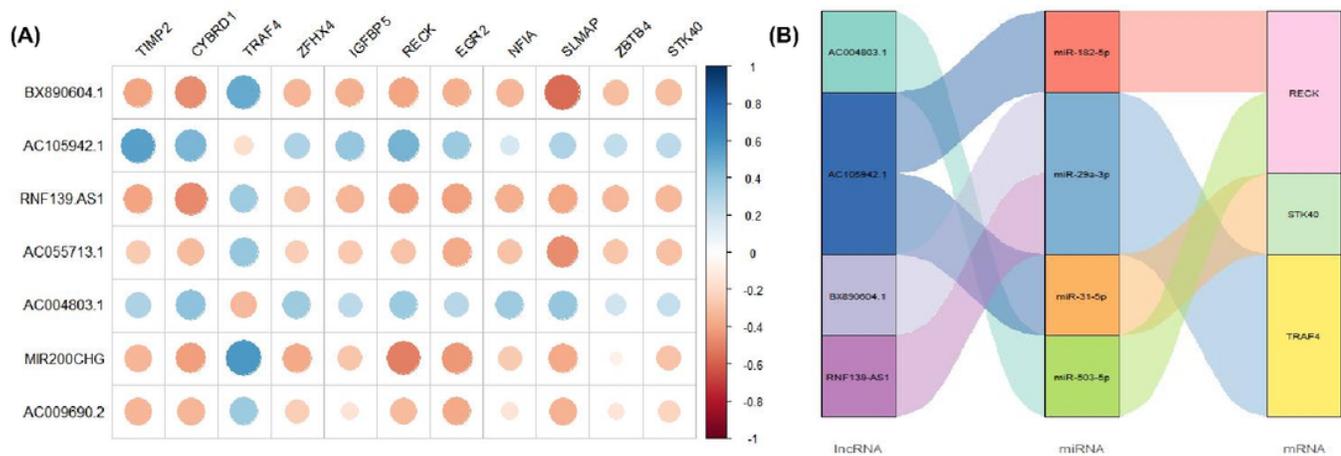
**Figure 4**

Kaplan-Meier survival analysis of DEGs in BC patients. Top three lncRNAs(A), miRNAs (B), and mRNAs(C) rank of the correlation between the expression level and prognosis of BC patients. DEGs: differentially expressed genes; lncRNA: long noncoding RNA; miRNA: microRNA; mRNA: messenger RNA.



**Figure 5**

Construction and validation of risk score system. Best lambda was calculate to assure a minimum mean cross-validated error by ten-fold cross-validation (A). Red dots represent partial likelihood deviance; solid vertical lines indicate their corresponding 95% CI; the left dotted vertical line is the value of lambda that gives minimum cvm, named lambda.min; the right dotted vertical line is the largest value such that error is within 1 standard error of the minimum, named lambda. 1se. The selection of the optimal cutoff and survival curve based on risk score. The coefficient values at varying levels of penalty (B). Each curve represents an lncRNA. Distribution of densities for low- and high-risk score BC patients(C). Risk score-related standardized log-rank statistics(D). Maximally statistic was defined as the optimal cutoff value. Kaplan-Meier survival curve of two groups in training set(E). Time-dependent ROC curves based on risk score level in training set (F). Kaplan-Meier survival curve of two groups in validation set(G). Time-dependent ROC curves based on risk score level in validation set (H).ROC: receiver operating characteristic.



**Figure 6**

Risk score analysis of 7 DElncRNAs. Training set(A).Validation set(B).The above scatter plot displays the risk score of 7 DElncRNAs, medium scatter plot displays the survival state of BC patients, and the below heatmap exhibits the DElncRNA expression profiles in each BC patients with survival data. Red is defined as high expression, and blue is defined as low expression.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table2.xls](#)
- [TableS1.pdf](#)
- [TableS2.pdf](#)
- [TableS3.pdf](#)

- [TableS4.pdf](#)