

Construction of a lncRNA-Associated ceRNA Network and Identification of A Potential Six-lncRNA Prognostic Signature for Esophageal Cancer

Jun Hu

The second affiliated hospital, university of south china

Fang Wang

The second affiliated hospital, university of south china

Logen Liu

The second affiliated hospital, university of south china

Wenfeng Ning (✉ ningwf2003@gmail.com)

The second affiliated hospital, university of south china <https://orcid.org/0000-0003-1360-4498>

Primary research

Keywords: competing endogenous RNA, long noncoding RNAs, esophageal

Posted Date: June 1st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-31505/v1>

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

[Read Full License](#)

Abstract

BACKGROUND: Mounting evidence has shown that long noncoding RNAs (lncRNAs) can function as competing endogenous RNAs (ceRNAs) which participate in the initiation and progression of cancers. In the ceRNA network, lncRNAs, microRNAs (miRNAs) and mRNAs, communicate with and co-regulate each other. Rarely there is a systematic lncRNA-mediated ceRNA network and potential specific ceRNA pairs or triples of esophageal cancer (EC). In this study, we investigate the lncRNA-mediated ceRNA network in EC and screen the potential prognostic lncRNA biomarkers.

METHODS: We obtained mRNA, miRNA, and lncRNA expression data and relevant clinical features on patients with EC from The Cancer Genome Atlas (TCGA), and used the edgeR package to identify differentially expressed mRNAs, lncRNAs and miRNAs between EC samples and normal samples. The EC ceRNA network was constructed based on miRNA target prediction through the databases of miRcode, miRDB, miRTarBase and TargetScan. And then Pearson's correlation analysis was adopted to identify co-expression mRNA-lncRNA pairs. Finally, the robust likelihood-based survival analysis and Cox regression models were used to identify prognosis-related lncRNAs, which was evaluated by Kaplan-Meier and receiver operating characteristic (ROC) curve analysis.

RESULTS: A total of 3,200 mRNAs, 131 miRNAs and 1,338 lncRNAs were identified as significantly differentially expressed in EC, of which, 30 mRNAs, 15 lncRNAs, and 8 miRNAs were incorporated in the ceRNA network. According to the ceRNA network node degrees, lncRNA MAGI2-AS3, hsa-mir-93 and TGFBR2 were the key genes. Also, the ceRNA network revealed some important ceRNA pairs and triples, such as SNX29P2-TGFBR2 and MAGI2-AS-hsa-mir-143-COL1A1. Finally, we developed a six-lncRNA signature (ZNF341-AS1, AC130324.2, AC027271.1, AL591212.1, AL732314.4 and LOC105372352), with improved diagnostic potential for EC with the area under the ROC curve of 0.93.

CONCLUSIONS: our present work sheds new light on the tumorigenesis roles of lncRNA-mediated ceRNA network in EC and identifies a six-lncRNA model that could be used as candidate prognostic signature.

Introduction

Esophageal cancer (EC) was the eighth most commonly occurring cancer overall

and the sixth most common cause of cancer-related deaths with a mortality of 400,000 in 2012. The major histologic types of EC are esophageal squamous cell carcinoma and esophageal adenocarcinoma [1]. Thus, it is urgent to explore the molecular mechanisms underlying EC for developing effective diagnostic and therapeutic targets.

Long noncoding RNA (lncRNA), with transcripts ≥ 200 bp in length, can modulate gene expression at the transcriptional, posttranscriptional and epigenetic levels [2]. In 2011, Salmena et al. proposed a ceRNA hypothesis in which lncRNAs, mRNAs and other RNAs act as miRNA sponges via shared miRNA response elements (MREs) to suppress miRNA functions. While competing for binding to shared MREs, these RNA

transcripts interact with and coregulate one another [3]. lncRNAs can be viewed as master regulators that modulate the expression of target genes.

Accumulating evidence shows that lncRNA-miRNA-mRNA networks are involved in the pathogenesis and progression of tumors [4–7]. However, our understanding of RNA transcript interactions in EC remains obscure.

Advancements in high-throughput sequencing technology and bioinformatics provide the opportunity to explore tumorigenic mechanisms [8]. The Cancer Genome Atlas (TCGA) database has profiles from more than 11,000 patient samples, obtained through large-scale genome sequencing, together with clinicopathological information derived from over 30 human cancer types [9]. In this study, a lncRNA-related ceRNA network for EC was constructed, and the prognostic value of lncRNAs was subsequently evaluated.

Methods

Data

EC profiles including miRNA-seq and RNA-Seq data, together with patient clinical information, were downloaded from TCGA database in November 2018. All of the RNA-Sequencing (RNA-Seq) data in the present work were level 3 in TCGA, which has been normalized and processed [10].

Differentially expressed lncRNAs, miRNAs and mRNAs

The significantly differentially expressed mRNAs (DEmRNAs), miRNAs (DEmiRNAs) and lncRNAs (DElncRNAs) were identified in EC samples in comparison with adjacent nontumorous tissues using the edgeR package [11], and mRNAs and lncRNAs were defined and annotated by The Genome Research Project of ENCyclopedia of DNA Elements (GENCODE) (GRCh38) (v25) catalog [12].

Identification of miRNA-target interactions

The prediction of miRNA-lncRNA interactions was based on the miRcode database (<http://www.mircode.org/>)[13]. MiRNA targeted mRNAs were cooperatively predicted using 3 programs, and the most targeted genes were included in all 3 datasets, including miRTarBase (<http://mirtarbase.mbc.nctu.edu.tw/>), TargetScan (<http://www.TargetScan.org/>), and miRDB (<http://www.mirdb.org/>)[14].

Identification of competing lncRNA-miRNA pairs in the ceRNA network based on gene expression

According to the ceRNA hypothesis, lncRNAs may positively regulate mRNAs by competing for their target miRNAs. To identify co-expressed lncRNA-mRNA pairs, we evaluated expression correlations among the

above DE mRNAs and DE lncRNAs using Pearson's correlation coefficients. A coefficient parameter >0.50 with a p -value < 0.05 was the criterion for retaining the lncRNA-mRNA pairs for further analysis [15].

Construction of the lncRNA-mediated ceRNA network and topological analysis

The construction of the lncRNA-miRNA-mRNA ceRNA network was based on potential lncRNA-miRNA and miRNA-mRNA interactions and co-expression lncRNA-mRNA pairs.

The ceRNA network was visualized using Cytoscape software [16], and the NetWorkAnalyzer toolkit [17] was used to analyze the topology of the ceRNA network. NetWorkAnalyzer is usually used to analyze the clustering coefficient, network node degree and average clustering coefficient. In this study, the network node degree (ND) was calculated. The number of edges connected to the node was defined as the node degree, displaying the local centrality of the node in the network [18]. A higher node degree indicated a greater importance of the node in the network.

Analyzing the relationship between the expression of DE lncRNAs in the ceRNA network and EC pathological stage

The relationship between DE lncRNA expression and pathological tumor stage in EC was estimated using the Kruskal-Wallis test in R [19].

Functional enrichment analyses

To explore the function of lncRNAs in the ceRNA network, we performed enrichment analyses of the DE mRNAs involved in the network to reveal the potential biological functions and pathways (p -value < 0.05) [20] in Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG), and then visualized the results using the clusterProfiler package in R [21, 22].

Survival analysis

To develop a lncRNA-based signature to improve prognostic prediction in EC, we randomly divided a total of 159 EC samples in TCGA into training and test cohorts. When the lncRNA expression level was above 0 and appeared in 50% of the total samples, it was defined as an abundance of expression. Univariate Cox regression survival analysis, followed by robust likelihood-based survival analysis, was performed to screen survival-related lncRNAs in the training cohort using the survival and rbsurv packages in R, respectively [23]. The procedure was in accordance with the report by Zhiqiang Wang [24]. Finally, the lncRNAs significantly associated with overall survival (OS) in the multivariate Cox survival analysis (p -value < 0.05) were selected to generate a risk-score formula.

The risk score for each patient in the training cohort was calculated using the formula. A time-dependent receiver operating characteristic (ROC) curve plotted using survivalROC package in R was adopted to

assess the predictive capacity of the risk-score model, and the area under the curve (AUC) was calculated. An optimal cut-off point for the risk score was determined with maximal specificity and sensitivity. Based on the optimal cut-off point, survival differences in patients between the high-risk and compared by low-risk subgroups were assessed by the Kaplan-Meier method and log-rank test using the survival package in R [25].

Results

Patient characteristics

EC miRNA-seq and RNA-Seq data derived from 170 samples, including 159 EC samples and 11 adjacent nontumorous samples, were downloaded from TCGA. A total of 159 EC cases were pathologically confirmed. The median age was 62 years old for the 159 EC patients. The detailed clinical characteristics and pathological characteristics of the 159 EC patients are shown in Table 1.

Table 1
Clinicopathological characteristics of 159 EC patients

Parameter	Subtype	Patients
Histological Type	Esophagus Adenocarcinoma	79
	Esophagus Squamous Cell Carcinoma	80
Age (years)	>62	68
	≤62	91
Gender	Male	137
	Female	22
Race	White	99
	Asian	38
	Black or African American	5
	Unknow	17
Pathologic stage	Stage I	16
	Stage II	68
	Stage III	48
	Stage IV	8
	Unknow	19
Pathologic T	T0	1
	T1	26
	T2	37
	T3	76
	T4	4
	TX [†]	15
Pathologic N	N0	65
	N1	62
	N2	9
	N3	6
	NX [†]	17
Pathologic M	M0	120

M1	7
MX †	32

†stand for status unknown

Identifying DEmRNAs, DEmiRNAs, and DElncRNAs

A total of 3,200 DEmRNAs, 131 DEmiRNAs and 1,338 DElncRNAs (Table S1) were identified by the “DESeq” package in R (Figure 1) with thresholds of $|\log_2FC| > 1.0$ and false discovery rate (FDR) < 0.05 . Of these, 1,413 mRNAs, 758 lncRNAs and 86 miRNAs were overexpressed, and 1,787 mRNAs, 580 lncRNAs and 45 miRNAs were under expressed in EC patient samples compared with those in adjacent nontumorous samples.

Construction of the ceRNA network and analysis of topology

The 16 specific DEmiRNAs were identified to interact with 113 DElncRNAs, predicted by miRcode based on MREs. The candidate mRNA targets of the 16 DEmiRNAs were predicted using a combination of MiRanda, miRTarBase and TargetScan databases, excluding the candidate mRNAs not involved with the 3,200 DEmRNAs, and 103 DEmRNAs were identified.

Co-expression among the above 103 DEmRNAs and 113 DElncRNAs was evaluated by Pearson’s correlation coefficients, and a total of 101 lncRNA-mRNA pairs were identified with an absolute value of the coefficients, whereas 37 DEmRNAs and 24 DElncRNAs remained. Lastly, only the co-expressed lncRNA-mRNA pairs sharing common DEmiRNAs were retained (Figure 2). The ceRNA network was visualized using Cytoscape software and contained 53 nodes (30 mRNAs, 15 lncRNAs and 8 miRNAs) and 63 edges (Figure 3). In addition, we calculated the node degree of each gene by topology. hsa-mir-93 (ND = 21), MAGI2-AS3 (ND = 5) and TGFBR2 (ND = 4) were considered important genes in the ceRNA network among the selected miRNAs, lncRNAs, and mRNAs, respectively (Figure 4).

Association between DElncRNA expression and pathological stage

The Kruskal-Wallis test was used to assess the difference in DElncRNAs in the ceRNA network according to EC pathological stages. The number of patients with stage I, stage II, stage III and stage IV disease was 16, 68, 48 and 8, respectively. There were 6 DElncRNAs with statistically significant differences, with only MAGI2-AS3, AC009093.1 and TTTY14 expression displaying an increasing trend with an increase in

histological grade (Figure 5). Higher expression levels of these lncRNAs were associated with histological grade. Our results implied that these DElncRNAs promote EC development and metastasis.

Functional enrichment analysis of DEmRNAs in the ceRNA network

The function of a ceRNA network is inferred from the functions of the constituent mRNAs. Functional analysis revealed that the 30 DEmRNAs in the ceRNA network were enriched in 4 GO biological process categories and 1 KEGG pathway (p -value < 0.05). Significant GO biological processes included the BMP signaling pathway (GO:0030509), negative regulation of cell migration (GO:0030336), glucose homeostasis (GO:0042593) and extracellular matrix organization (GO:0030198) (Figure 6A). The enriched KEGG pathway was platelet activation (hsa04611) (Figure 6B). The genes in the “platelet activation” pathway were COL1A1, COL5A2 and ITPR1. These results implied that the function of the ceRNA network was primarily associated with tumor progression and metastasis.

Construction of a DElncRNA-based prognostic signature

A total of 7,309 altered lncRNAs were recognized among the 14,386 lncRNAs in the training cohort using the screening criteria. Univariate survival analysis identified 483 lncRNAs significantly associated with the OS of EC patients with a p -value < 0.05 . Then, using robust likelihood-based survival modeling with 100 iterations for 483 lncRNAs, 12 featured lncRNAs were screened (Table S2). Based on multivariate Cox regression analysis with those featured lncRNAs, six prognostic lncRNAs with statistical significance (p -value < 0.05) were further selected, and the risk coefficients were also calculated. The risk scoring system for each patient was calculated as follows: risk score = $(0.66) \times \text{ZNF341-AS1 value} + (-0.3956) \times \text{AC130324.2 value} + (0.9481) \times \text{AC027271.1 value} + (1.0792) \times \text{AL591212.1 value} + (1.0369) \times \text{AL732314.4 value} + (-0.5935) \times \text{LOC105372352 value}$.

ROC analysis was used to estimate the sensitivity and specificity of the six-lncRNA signature for survival prediction. An optimal cut-off point of 0.589 was selected, which yielded the maximal specificity and sensitivity as evidenced by the AUC 0.93 (Figure 7A). Additionally, the risk scores for each patient in the training cohort were calculated. Using the optimal cut-off point, patients were classified into a high-risk group ($n = 46$) and a low-risk group ($n = 34$). Figure 7B shows the distribution of patient survival status and survival time. Survival differences, estimated with the Kaplan–Meier method and compared by log-rank test, demonstrated that patients with a high-risk score had a shorter survival time (p -value < 0.0001) (Figure 7B). Subsequently, this risk score formula was verified in the test cohort and the total cohort (Figure 7C, D).

Discussion

In recent years, researchers have observed that noncoding RNAs have a striking variety of biological functions[22].

Understanding the crosstalk among lncRNAs, miRNAs and mRNAs on the basis of the ceRNA hypothesis will lead to significant insights into gene regulatory networks with implications on tumorigenesis. However, there have been relatively few studies on the roles of lncRNAs, miRNAs and mRNAs as ceRNAs in EC.

In the present study, through bioinformatics miRNA target prediction and co-expression analysis of lncRNA-mRNA pairs, we constructed a ceRNA network with 15 DElncRNAs, 30 DEmRNAs and 8 DEmiRNAs in EC. In our ceRNA network, lncRNAs and mRNAs interacted with each other *via* miRNAs, forming a complex regulatory network.

In the EC ceRNA network, lncRNA MAGI2-AS3 was the most prominent lncRNA node that competed with some other lncRNAs (AC009093.1, TTTY14, SNX29P2, etc.) and sponged hsa-mir-143, hsa-mir-93, hsa-mir-204 and hsa-mir-372. Additionally, by interacting with these miRNAs, lncRNA MAGI2-AS3 regulated the expression of some mRNAs (RGMB, FSCN1, TMEM100, SLC2A4, etc.). We also found that the expression level of MAGI2-AS3 was associated with EC pathological stage. This implied that lncRNA MAGI2-AS3 might play a major role in EC pathogenesis. In recent studies, lncRNA MAGI2-AS3 has been shown to suppress bladder cancer progression by sponging miR-15b [26]. However, reports on lncRNA MAGI2-AS3 sponging miRNAs and interacting with mRNAs in EC are limited.

ceRNAs can affect the abundance of miRNAs and their impact on targets and impose an additional level of posttranscriptional regulation. MiRNAs act as key regulators in ceRNA networks[27]. In this study, hsa-mir-93 represented the node with the highest degree in the EC network and targeted 7 lncRNAs and 14 mRNAs (TGFB2, OSR1, SLC2A4, etc.). These results were not consistent with the EC ceRNA network proposed by Li-Ping Chen et al., as hsa-mir-93 was not in their ceRNA network[28]. In previous studies, hsa-mir-93 was shown to be an important oncogene in prostate cancer[29] and to function as biomarkers for the diagnosis and prognosis of esophageal cancer.[30] Additionally, many studies have confirmed that hsa-mir-93 shares several lncRNA-binding sites [31–33]. Our results may shed new light on the interactions of hsa-mir-93 with lncRNAs and mRNAs in EC.

In our ceRNA network, SNX29P2-TGFB2 pair was the common target of several miRNAs, such as hsa-mir-93, hsa-mir-372, hsa-mir-17 and hsa-mir-145. TGFB2 was also the mRNA node with the highest degree in the EC network. TGFB2 is a member of the TGFβ receptor subfamily. In previous research, TGF-β signaling was shown to play an important role in carcinogenesis[34], and hsa-mir-17 can impede migration and invasion via TGF-β pathway in esophageal squamous cell carcinoma[35]. Our data implied that lncRNA SNX29P2 might function as a ceRNA to impede the TGF-beta pathway in EC.

In the EC ceRNA network, sponging of overexpressed lncRNA AC009093.1 and under expressed lncRNA MAGI2-AS3, relieved their targets COL1A1 and COL5A2 from being suppressed by hsa-mir-143. Previous studies have suggested that hsa-mir-143 functions as a tumor suppressor in several cancer types[36, 37], and COL1A1 has been shown to act as an oncogene to regulate tumors [38–40]. Our findings implied that hsa-mir-143 may function as a tumor suppressor by targeting COL1A1 and COL5A2 in EC and competing with lncRNA MAGI2-AS3/AC009093.1.

Most of the ceRNA triples included downregulated lncRNA-mRNA pairs and upregulated miRNAs, whereas in this study, downregulated hsa-mir-204 targeted many downregulated lncRNA-mRNAs. Many studies have demonstrated downregulation of hsa-mir-204 in the progression of various types of malignant tumors, including EC and have shown that hsa-mir-204 functions as a tumor suppressor[41]. This finding indicated that miRNA expression may not always be correlated with target ceRNAs because of the complexity of crosstalk among RNA transcripts.

Our study has provided a global view of regulatory ceRNA interaction networks in EC and laid a foundation for further functional research in EC. Because of their functional roles, lncRNAs have attracted much attention in the discovery of prognostic biomarkers [42]. Some studies have reported that lncRNAs may serve as biomarkers for EC, however, there are no commonly accepted prognostic biomarkers.

In survival analysis, the lncRNA expression data may have led to problems in predictive modeling because these data indicate an imbalance between a large number of covariates involving lncRNAs and a small number of samples, denoted as an issue with dimensionality [43].

In this study, to overcome the limitation involving high-dimensionality of expression data for 514 lncRNAs, we used a robust likelihood-based survival model utilizing a cross-validation technique to reduce dimensionality [44, 45] and finally developed a six-lncRNA signature to predict OS for EC patients. There have been very few studies on ZNF341-AS1, AC130324.2, AC027271.1, AL591212.1, AL732314.4 and LOC105372352. The six-gene signature was identified as an indicator for EC with high prognostic significance and demonstrated that lncRNAs have significant prognostic value for EC patients [46].

Our results were not consistent with the findings of Guo-Wei Huang et al. [47], in which a three-lncRNA signature including RP11-366H4.1.1, LINC00460 and AC093850.2 was identified as a potential prognostic indicator for esophageal squamous cell carcinoma. The discrepancy may be mainly ascribed to data and prediction method.

Conclusions

Taken together, through construction the ceRNA network of EC with 15 lncRNAs competed with 30 mRNAs for 8 miRNAs, we revealed important genes and the specific lncRNA-miRNA-mRNA interactions, shedding new light on the functional roles of lncRNA-mediated ceRNA network in esophageal tumorigenesis. Moreover, we identified a six-lncRNA model that could be used as candidate prognostic signature.

However, some limitations should be taken into consideration. ceRNA network construction was performed using bioinformatics predictions and lacked functional validation of ceRNA pairs (lncRNA-mRNA) and triples (lncRNA-miRNA-mRNA). Additionally, the EC sample size was not large, and validation of this six-lncRNA signature in larger cohorts of EC patients and clinical trials is essential for further investigation. In conclusion, we identified a six-lncRNA signature that might predict the prognosis of EC

patients. Further study is required to explore the functional roles of these lncRNAs in esophageal carcinogenesis.

Abbreviations

CeRNA Competing endogenous RNA

LncRNAs Long noncoding RNAs

MiRNAs MicroRNAs

EC Esophageal Cancer

TCGA The Cancer Genome Atlas

MREs MiRNA Response Elements

GENCODE Genome Research Project of ENCyclopedia of DNA Elements

ND Node Degree

GO Gene Ontology

KEGG Kyoto Encyclopedia of Genes and Genomes

OS Overall Survival

ROC Receiver Operating Characteristic

AUC Area Under the Curve

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for Publication

Not applicable.

Availability of data and material

The data that support the findings of this study are openly available in TCGA at <https://portal.gdc.cancer.gov>. Detailed information is available in main text.

Competing interests

The authors declare no conflict of interest.

Funding

No funding.

Authors' contributions

Jun Hu and Fang Wang wrote manuscript and drew diagrams. Logen Liu and Wenfeng Ning wrote codes. Wenfeng Ning revised the manuscript, designed and guided the project. All authors read and approved the final manuscript.

Acknowledgements

The manuscript was edited for English language at the American Journal Experts (Certificate key: 07E9-E334-53B3-4BF7-09C3).

References

1. Sardana RK, Chhikara N, Tanwar B, Panghal A. Dietary impact on esophageal cancer in humans: a review. *Food Funct.* 2018;9(4):1967–77. doi:10.1039/c7fo01908d.
2. Sun W, Yang Y, Xu C, Guo J. Regulatory mechanisms of long noncoding RNAs on gene expression in cancers. *Cancer Genet.* 2017;216–217:105–10. doi:10.1016/j.cancergen.2017.06.003.
3. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell.* 2011;146(3):353–8. doi:10.1016/j.cell.2011.07.014.
4. Sui J, Li YH, Zhang YQ, Li CY, Shen X, Yao WZ et al. Integrated analysis of long non-coding RNA-associated ceRNA network reveals potential lncRNA biomarkers in human lung adenocarcinoma. *Int J Oncol.* 2016;49(5):2023–36. doi:10.3892/ijo.2016.3716.
5. Zhang Y, Zhang P, Wan X, Su X, Kong Z, Zhai Q et al. Downregulation of long non-coding RNA HCG11 predicts a poor prognosis in prostate cancer. *Biomed Pharmacother.* 2016;83:936–41. doi:10.1016/j.biopha.2016.08.013.
6. Wu H, Wu R, Chen M, Li D, Dai J, Zhang Y et al. Comprehensive analysis of differentially expressed profiles of lncRNAs and construction of miR-133b mediated ceRNA network in colorectal cancer. *Oncotarget.* 2017;8(13):21095–105. doi:10.18632/oncotarget.15045.
7. Zhao L, Liu B. Identification of potential prognostic ceRNA module biomarkers in patients with pancreatic adenocarcinoma. *Oncotarget.* 2017;8(55):94493–504. doi:10.18632/oncotarget.21783.
8. Yohe S, Thyagarajan B. Review of Clinical Next-Generation Sequencing. *Arch Pathol Lab Med.* 2017;141(11):1544–57. doi:10.5858/arpa.2016-0501-RA.

9. Tomczak K, Czerwinska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)*. 2015;19(1A):A68–77. doi:10.5114/wo.2014.47136.
10. Zhang Z, Huang K, Gu C, Zhao L, Wang N, Wang X et al. Molecular Subtyping of Serous Ovarian Cancer Based on Multi-omics Data. *Sci Rep*. 2016;6:26001. doi:10.1038/srep26001.
11. Pan WK, Zhang YF, Yu H, Gao Y, Zheng BJ, Li P et al. Identifying key genes associated with Hirschsprung's disease based on bioinformatics analysis of RNA-sequencing data. *World J Pediatr*. 2017;13(3):267–73. doi:10.1007/s12519-017-0002-0.
12. Jalali S, Gandhi S, Scaria V. Navigating the dynamic landscape of long noncoding RNA and protein-coding gene annotations in GENCODE. *Hum Genomics*. 2016;10(1):35. doi:10.1186/s40246-016-0090-2.
13. Jeggari A, Marks DS, Larsson E. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics*. 2012;28(15):2062–3. doi:10.1093/bioinformatics/bts344.
14. Hou C, Wang F, Liu X, Chang G, Geng X. Comprehensive Analysis of Interaction Networks of Telomerase Reverse Transcriptase with Multiple Bioinformatic Approaches: Deep Mining the Potential Functions of Telomere and Telomerase. *Rejuvenation Res*. 2017;20(4):320–33. doi:10.1089/rej.2016.1909.
15. Wang P, Ning S, Zhang Y, Li R, Ye J, Zhao Z et al. Identification of lncRNA-associated competing triplets reveals global patterns and prognostic markers for cancer. *Nucleic Acids Res*. 2015;43(7):3478–89. doi:10.1093/nar/gkv233.
16. Jiang H, Ma R, Zou S, Wang Y, Li Z, Li W. Reconstruction and analysis of the lncRNA-miRNA-mRNA network based on competitive endogenous RNA reveal functional lncRNAs in rheumatoid arthritis. *Mol Biosyst*. 2017;13(6):1182–92. doi:10.1039/c7mb00094d.
17. Doncheva NT, Assenov Y, Domingues FS, Albrecht M. Topological analysis and interactive visualization of biological networks and protein structures. *Nat Protoc*. 2012;7(4):670–85. doi:10.1038/nprot.2012.004.
18. Ye C, Comin CH, Peron TK, Silva FN, Rodrigues FA, Costa Lda F et al. Thermodynamic characterization of networks using graph polynomials. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2015;92(3):032810. doi:10.1103/PhysRevE.92.032810.
19. Li Y, Diao F, Shi S, Li K, Zhu W, Wu S et al. Computed tomography and magnetic resonance imaging evaluation of pelvic lymph node metastasis in bladder cancer. *Chin J Cancer*. 2018;37(1):3. doi:10.1186/s40880-018-0269-0.
20. Gong XJ, Yu H, Yang CB, Li YF. Knowledge enrichment analysis for human tissue-specific genes uncover new biological insights. *J Integr Bioinform*. 2012;9(2):194. doi:10.2390/biecoll-jib-2012-194.
21. Miao R, Wang Y, Wan J, Leng D, Gong J, Li J et al. Microarray expression profile of circular RNAs in chronic thromboembolic pulmonary hypertension. *Medicine (Baltimore)*. 2017;96(27):e7354. doi:10.1097/MD.0000000000007354.

22. Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*. 2014;157(1):77–94. doi:10.1016/j.cell.2014.03.008.
23. Mao X, Qin X, Li L, Zhou J, Zhou M, Li X et al. A 15-long non-coding RNA signature to improve prognosis prediction of cervical squamous cell carcinoma. *Gynecol Oncol*. 2018;149(1):181–7. doi:10.1016/j.ygyno.2017.12.011.
24. Wang Z, Chen G, Wang Q, Lu W, Xu M. Identification and validation of a prognostic 9-genes expression signature for gastric cancer. *Oncotarget*. 2017;8(43):73826–36. doi:10.18632/oncotarget.17764.
25. Qu L, Wang ZL, Chen Q, Li YM, He HW, Hsieh JJ et al. Prognostic Value of a Long Non-coding RNA Signature in Localized Clear Cell Renal Cell Carcinoma. *Eur Urol*. 2018;74(6):756–63. doi:10.1016/j.eururo.2018.07.032.
26. Wang F, Zu Y, Zhu S, Yang Y, Huang W, Xie H et al. Long noncoding RNA MAGI2-AS3 regulates CCDC19 expression by sponging miR–15b–5p and suppresses bladder cancer progression. *Biochem Biophys Res Commun*. 2018;507(1–4):231–5. doi:10.1016/j.bbrc.2018.11.013.
27. Martirosyan A, De Martino A, Pagnani A, Marinari E. ceRNA crosstalk stabilizes protein expression and affects the correlation pattern of interacting proteins. *Sci Rep*. 2017;7:43673. doi:10.1038/srep43673.
28. Chen LP, Wang H, Zhang Y, Chen QX, Lin TS, Liu ZQ et al. Robust analysis of novel mRNA-lncRNA cross talk based on ceRNA hypothesis uncovers carcinogenic mechanism and promotes diagnostic accuracy in esophageal cancer. *Cancer Manag Res*. 2019;11:347–58. doi:10.2147/CMAR.S183310.
29. Yang Y, Jia B, Zhao X, Wang Y, Ye W. miR–93–5p may be an important oncogene in prostate cancer by bioinformatics analysis. *J Cell Biochem*. 2018. doi:10.1002/jcb.28332.
30. Liu MX, Liao J, Xie M, Gao ZK, Wang XH, Zhang Y et al. miR–93–5p Transferred by Exosomes Promotes the Proliferation of Esophageal Cancer Cells via Intercellular Communication by Targeting PTEN. *Biomed Environ Sci*. 2018;31(3):171–85. doi:10.3967/bes2018.023.
31. Li JP, Xiang Y, Fan LJ, Yao A, Li H, Liao XH. Long noncoding RNA H19 competitively binds miR–93–5p to regulate STAT3 expression in breast cancer. *J Cell Biochem*. 2019;120(3):3137–48. doi:10.1002/jcb.27578.
32. Chen C, Zheng Q, Kang W, Yu C. Long non-coding RNA LINC00472 suppresses hepatocellular carcinoma cell proliferation, migration and invasion through miR–93–5p/PDCD4 pathway. *Clin Res Hepatol Gastroenterol*. 2018. doi:10.1016/j.clinre.2018.11.008.
33. Wei H, Yang Z, Lin B. Overexpression of long non coding RNA CA3-AS1 suppresses proliferation, invasion and promotes apoptosis via miRNA–93/PTEN axis in colorectal cancer. *Gene*. 2019;687:9–15. doi:10.1016/j.gene.2018.11.008.
34. Liu S, Chen S, Zeng J. TGFbeta signaling: A complex role in tumorigenesis (Review). *Mol Med Rep*. 2018;17(1):699–704. doi:10.3892/mmr.2017.7970.
35. Jing C, Ma G, Li X, Wu X, Huang F, Liu K et al. MicroRNA–17/20a impedes migration and invasion via TGF-beta/ITGB6 pathway in esophageal squamous cell carcinoma. *Am J Cancer Res*.

- 2016;6(7):1549–62.
36. Shi H, Shen H, Xu J, Zhao S, Yao S, Jiang N. MiR–143–3p suppresses the progression of ovarian cancer. *Am J Transl Res*. 2018;10(3):866–74.
 37. Huang C, Huang J, Ma P, Yu G. microRNA–143 acts as a suppressor of hemangioma growth by targeting Bcl–2. *Gene*. 2017;628:211–7. doi:10.1016/j.gene.2017.07.046.
 38. Oleksiewicz U, Liloglou T, Tasopoulou KM, Daskoulidou N, Gosney JR, Field JK et al. COL1A1, PRPF40A, and UCP2 correlate with hypoxia markers in non-small cell lung cancer. *J Cancer Res Clin Oncol*. 2017;143(7):1133–41. doi:10.1007/s00432–017–2381-y.
 39. Li J, Ding Y, Li A. Identification of COL1A1 and COL1A2 as candidate prognostic factors in gastric cancer. *World J Surg Oncol*. 2016;14(1):297. doi:10.1186/s12957–016–1056–5.
 40. Zhang Z, Wang Y, Zhang J, Zhong J, Yang R. COL1A1 promotes metastasis in colorectal cancer by regulating the WNT/PCP pathway. *Mol Med Rep*. 2018;17(4):5037–42. doi:10.3892/mmr.2018.8533.
 41. Sun Y, Yu X, Bai Q. miR–204 inhibits invasion and epithelial-mesenchymal transition by targeting FOXM1 in esophageal cancer. *Int J Clin Exp Pathol*. 2015;8(10):12775–83.
 42. Evans JR, Feng FY, Chinnaiyan AM. The bright side of dark matter: lncRNAs in cancer. *J Clin Invest*. 2016;126(8):2775–82. doi:10.1172/JCI84421.
 43. Mirza B, Wang W, Wang J, Choi H, Chung NC, Ping P. Machine Learning and Integrative Analysis of Biomedical Big Data. *Genes (Basel)*. 2019;10(2). doi:10.3390/genes10020087.
 44. Wang Y, Ren F, Chen P, Liu S, Song Z, Ma X. Identification of a six-gene signature with prognostic value for patients with endometrial carcinoma. *Cancer Med*. 2018;7(11):5632–42. doi:10.1002/cam4.1806.
 45. Wang H, Xing F, Su H, Stromberg A, Yang L. Novel image markers for non-small cell lung cancer classification and survival prediction. *BMC Bioinformatics*. 2014;15:310. doi:10.1186/1471–2105–15–310.
 46. Song J, Xu Q, Zhang H, Yin X, Zhu C, Zhao K et al. Five key lncRNAs considered as prognostic targets for predicting pancreatic ductal adenocarcinoma. *J Cell Biochem*. 2018;119(6):4559–69. doi:10.1002/jcb.26598.
 47. Huang GW, Xue YJ, Wu ZY, Xu XE, Wu JY, Cao HH et al. A three-lncRNA signature predicts overall survival and disease-free survival in patients with esophageal squamous cell carcinoma. *BMC Cancer*. 2018;18(1):147. doi:10.1186/s12885–018–4058–6.

Figures

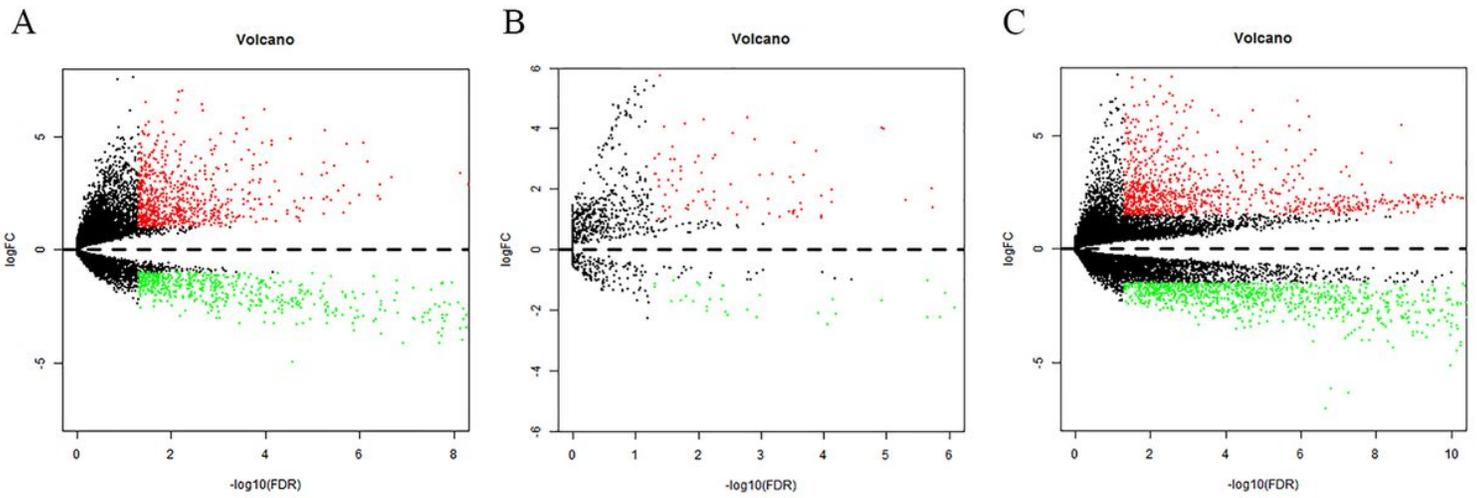


Figure 1

Volcano plots showing differentially expressed lncRNAs, mRNAs and miRNAs in A, B and C, respectively. The plots depict the number, significance, and reliability of differentially expressed lncRNAs, mRNAs and miRNAs. Red and green dots represent upregulated and downregulated genes, respectively.

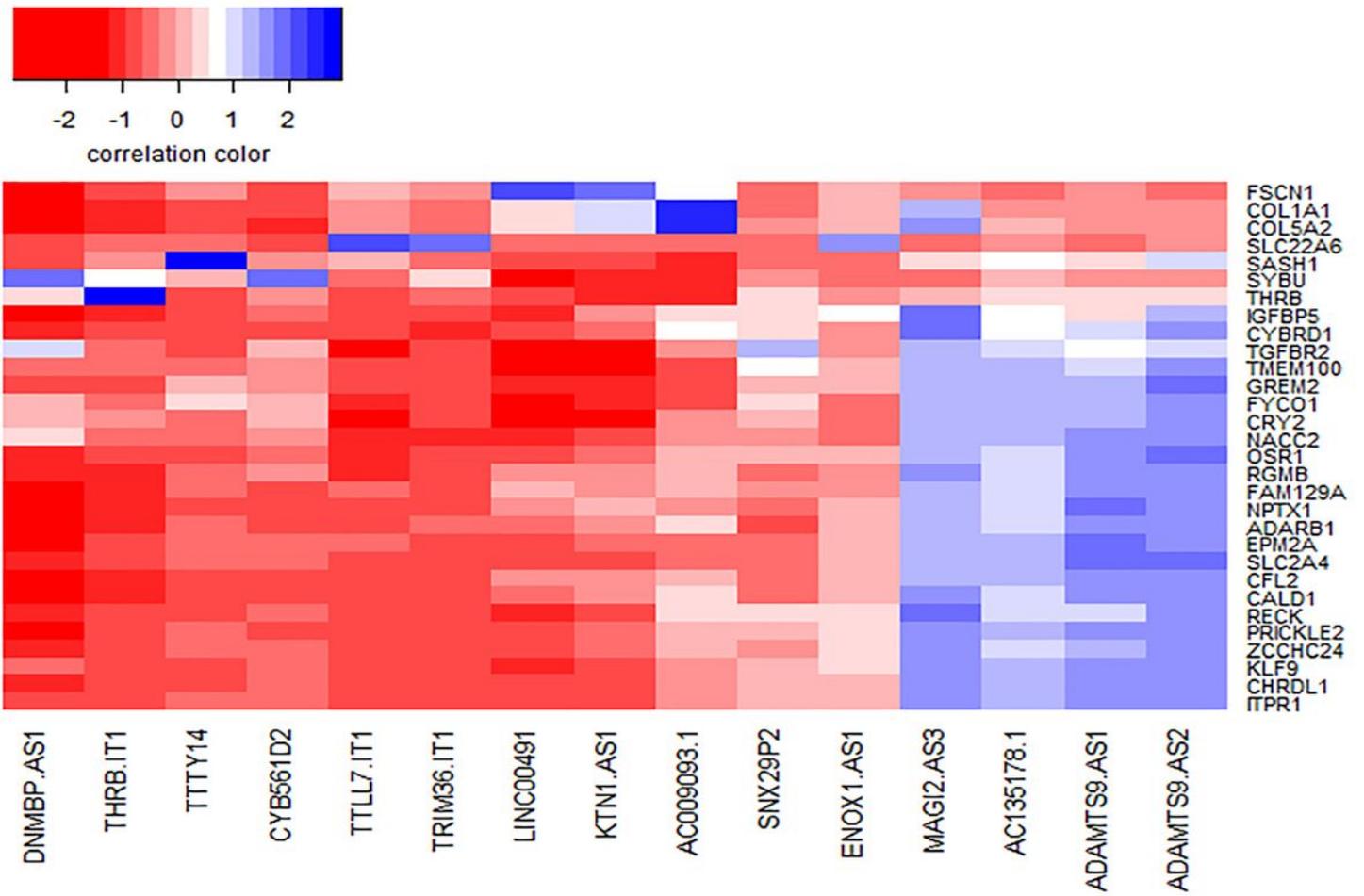


Figure 2

Correlation plot between DEmRNAs and DElncRNAs in the ceRNA network. The plot depicts the correlation matrix using colored cells to represent the extent of the correlation. The color scale ranges from dark red for strong negative correlations to white (no correlation) to dark blue for strong positive correlations. The X and Y axes represent DElncRNAs and DEmRNAs, respectively.

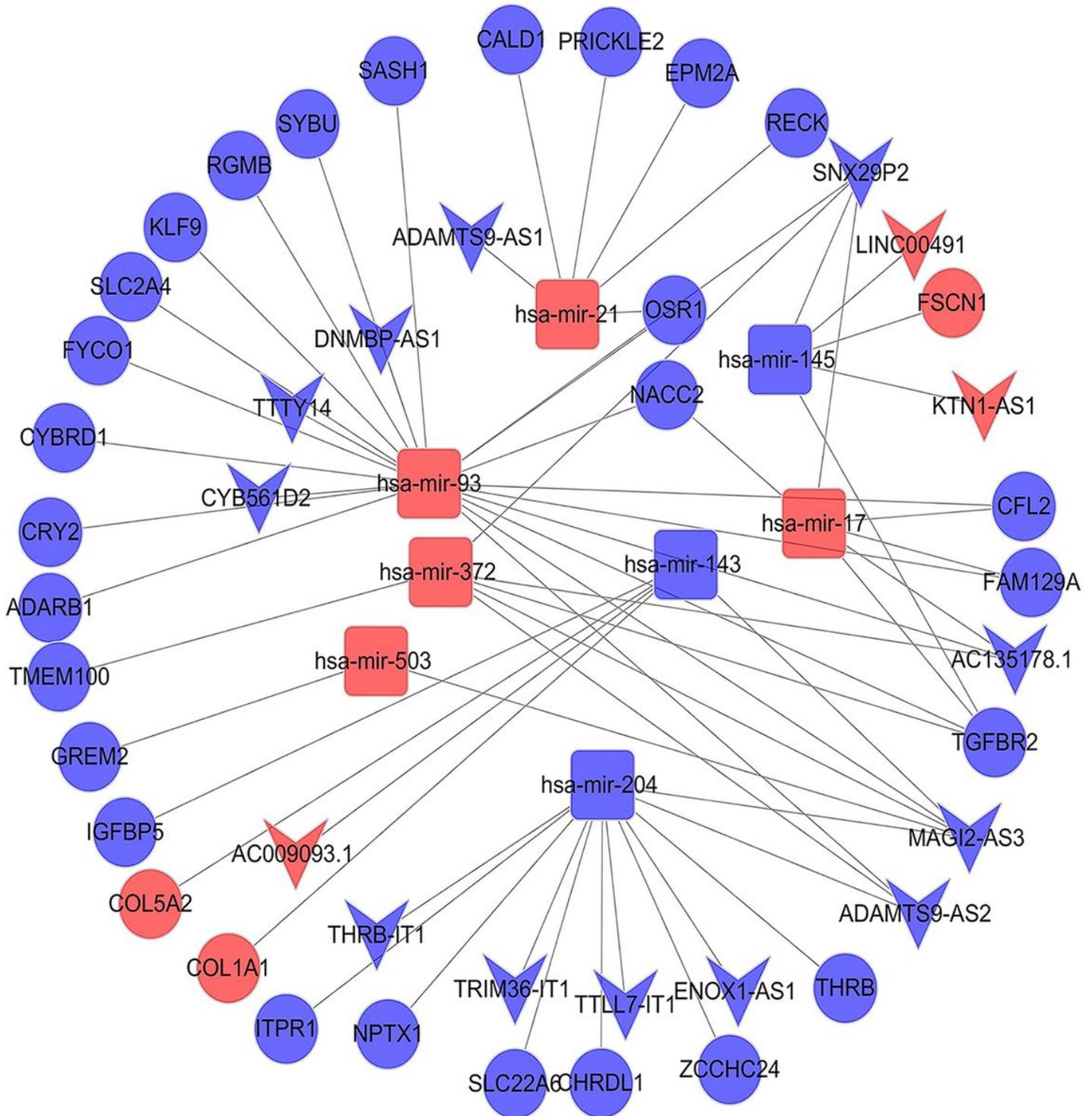


Figure 3

The ceRNA network for EC. DElncRNAs, DEmiRNAs, and DEmRNAs are indicated with Vs, rectangles and ellipses, respectively. Genes are denoted as nodes in the graph, and interactions among them are

presented as edges. The network involves 15 DElncRNAs, 8 DEmiRNAs, and 30 DEmRNAs. Red represents high expression and blue low expression

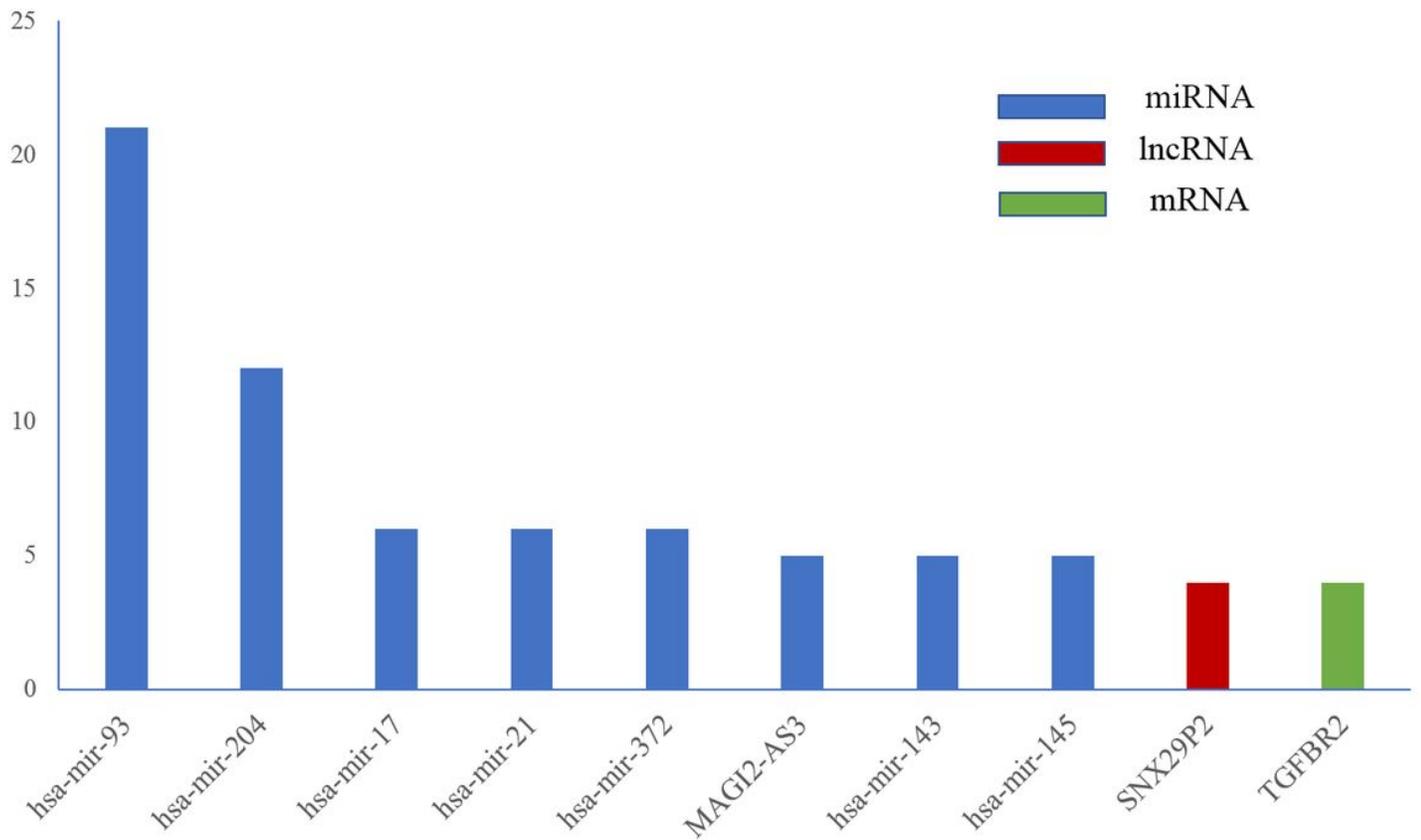


Figure 4

Top 10 genes with high node degree in the ceRNA network. X and Y axes represent genes and node degrees, respectively.

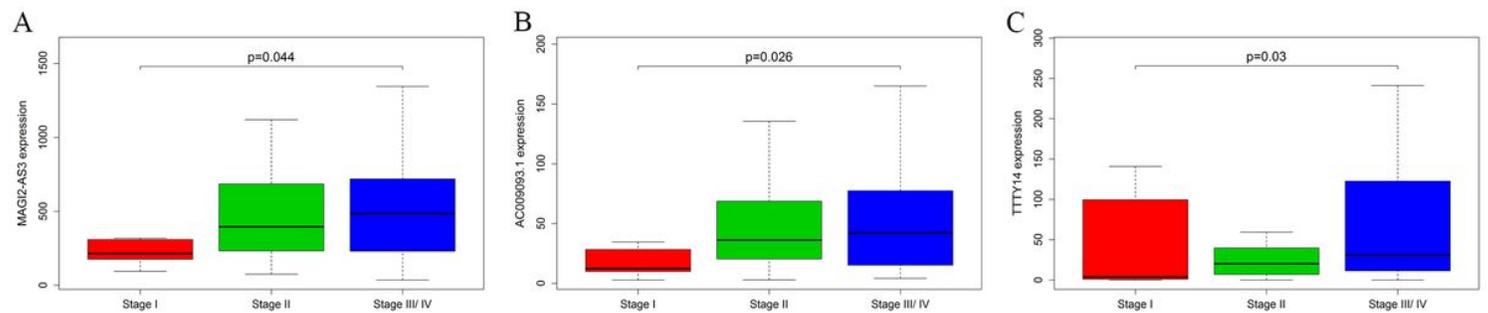


Figure 5

The difference between DElncRNA expression levels in different pathological stages. Stage III/IV represented stage III and stage IV; p-value < 0.05 was considered statistically significant.

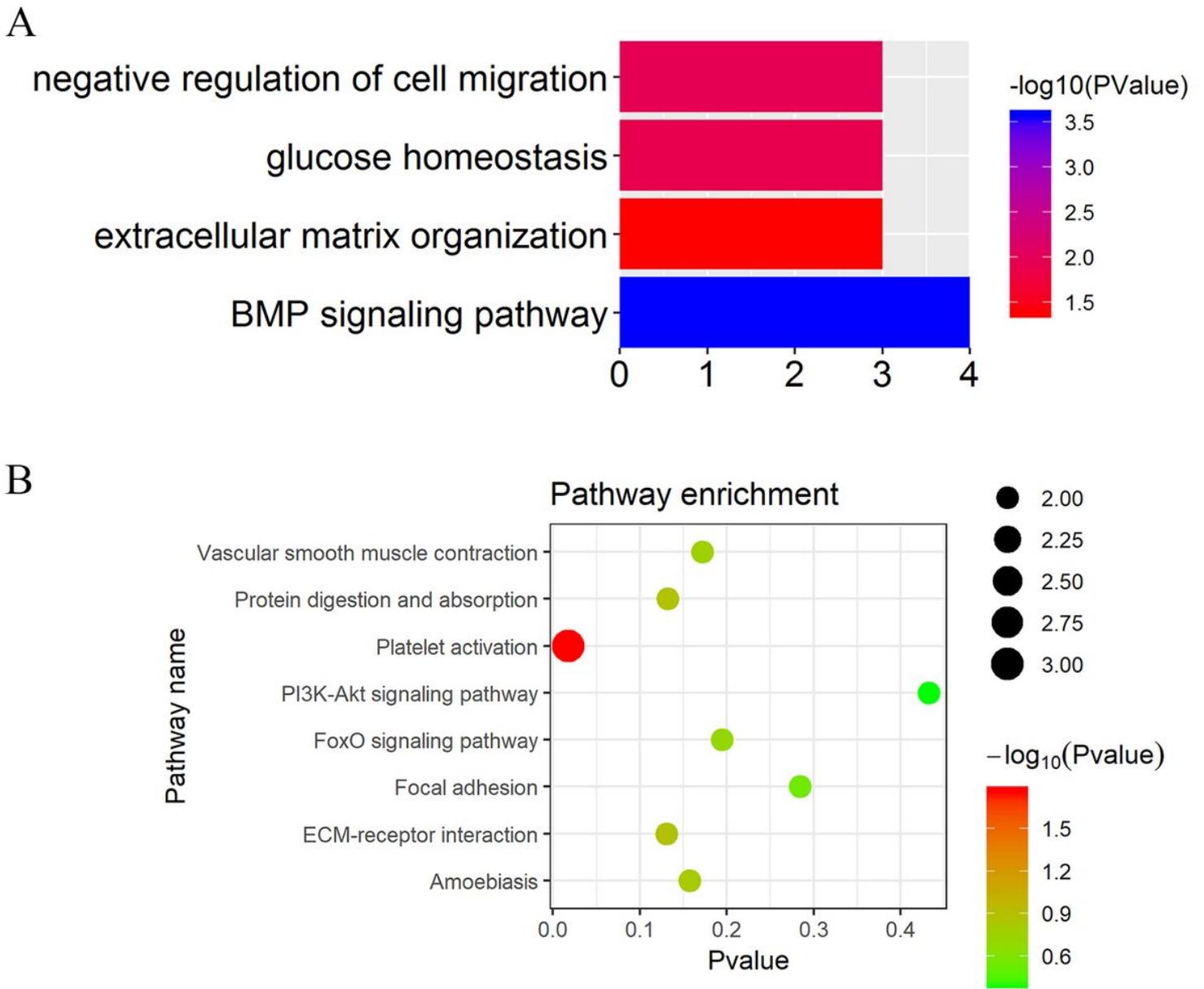


Figure 6

The plot of DEmRNAs in the ceRNA network enriched in GO and KEGG pathways. (A) The significantly enriched GO (biological process) terms for the DEmRNAs. The X axis represents the number of enriched genes. (B) The significantly enriched KEGG pathway for the DEmRNAs.

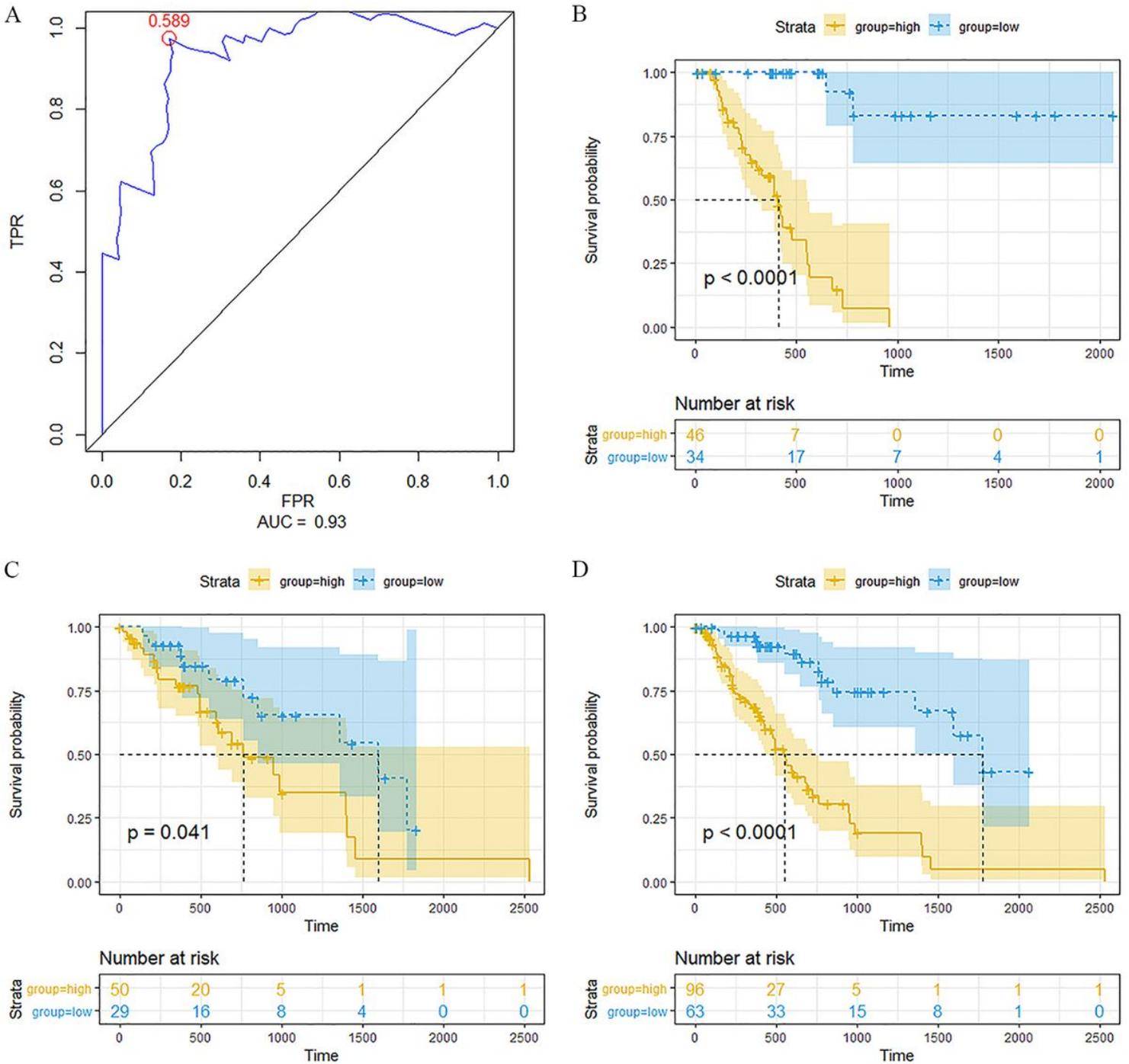


Figure 7

Survival ROC curve and Kaplan–Meier curve for the six-lncRNA signature in EC. (A) ROC curve showed sensitivity and specificity of the six-lncRNA signature for predicting OS. The red dot indicates the optimal cut-off value in the training set using ROC analysis. FPR: false positive rate; TPR: true positive rate. (B, C, D) The survival curve for the low-risk versus high-risk groups of patients using the optimal cut-off point in the three cohorts, (B) training cohort, (C) test cohort and (D) total cohort.

Supplementary Files

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

- [TableS2FeaturedlncRNAsbyrobustlikelihood.docx](#)
- [TableS1Differentiallyexpressedgenes.xlsx](#)