PORTLAND
PRESS

Research Article

Comprehensive analysis of prognostic biomarkers in lung adenocarcinoma based on aberrant lncRNA–miRNA–mRNA networks and Cox regression models

Yan Yao[1], Tingting Zhang[2], Lingyu Qi[2], Ruijuan Liu[3], Gongxi Liu[3], Jia Wang[4], Qi Song[1] and ⓘ Changgang Sun[5]

[1]Clinical Medical Colleges, Weifang Medical University. Postal Addresses: NO.7166, Baotong western street, WeiFang, ShanDong Province, P.R. China; [2]College of First Clinical Medicine, Shandong University of Traditional Chinese Medicine. Postal Addresses: NO.16369, Jingshi Road, Jinan, ShanDong Province, P.R. China; [3]Department of Oncology, Weifang Traditional Chinese Hospital. NO.1055, WeiZhouRoad, WeiFang, ShanDong Province, P.R.China; [4]College of Traditional Chinese Medicine, Shandong University of Traditional Chinese Medicine, Postal addresses: NO.16369, Jingshi Road, Jinan, ShanDong Province, P.R. China; [5]Innovative Institute of Chinese Medicine and Pharmacy, Shandong University of Traditional Chinese Medicine, Jinan 250014, Shandong, PR China

Correspondence: Changgang Sun (zhongliuyike@163.com)

OPEN ACCESS

Lung adenocarcinoma (LUAD) is the leading cause of cancer-related death worldwide, and its underlying mechanism remains unclear. Accumulating evidence has highlighted that long non-coding RNA (lncRNA) acts as competitive endogenous RNA (ceRNA) and plays an important role in the occurrence and development of LUAD. Here, we comprehensively analyzed and provided an overview of the lncRNAs, miRNAs, and mRNAs associated with LUAD from The Cancer Genome Atlas (TCGA) database. Then, differentially expressed lncR-NAs (DElncRNA), miRNAs (DEmiRNA), and mRNAs (DEmRNA) were used to construct a lncRNA–miRNA–mRNA regulatory network according to interaction information from miR-code, TargetScan, miRTarBase, and miRDB. Finally, the RNAs of the network were analyzed for survival and submitted for Cox regression analysis to construct prognostic indicators. A total of 1123 DElncRNAs, 95 DEmiRNAs, and 2296 DEmRNAs were identified ($|\log_2 \text{FoldChange}|$ (FC) > 2 and false discovery rate (FDR) or adjusted P value < 0.01). The ceRNA network was established based on this and included 102 lncRNAs, 19 miRNAs, and 33 mRNAs. The DEmRNAs in the ceRNA network were found to be enriched in various cancer-related biological processes and pathways. We detected 22 lncRNAs, 12 mRNAs, and 1 miRNA in the ceRNA network that were significantly associated with the overall survival of patients with LUAD ($P < 0.05$). We established three prognostic prediction models and calculated the area under the 1,3,5-year curve (AUC) values of lncRNA, mRNA, and miRNA, respectively. Among them, the prognostic index (PI) of lncRNA showed good predictive ability which was 0.737, 0.702 and 0.671 respectively, and eight lncRNAs can be used as candidate prognostic biomarkers for LUAD. In conclusion, our study provides a new perspective on the prognosis and diagnosis of LUAD on a genome-wide basis, and develops independent prognostic biomarkers for LUAD.

Introduction

Lung cancer is considered the leading cause of cancer-related deaths worldwide. Lung adenocarcinoma (LUAD) is the most prevalent pathological type of lung cancer; it accounts for approximately 40% of lung cancer cases and the average 5-year survival rate is only 18% [1–3]. Out of the 8.8 million cancer deaths in 2015, 1.69 million were due to lung cancer [4]. About 60–85% of all lung cancer patients are discovered at

advanced stages [5]. Individualized treatment guided by genetic testing results has become the main treatment for LUAD. However, the essential biomarkers and precise targets for the development and progression of LUAD remain unclear. Such cases underscore the critical need to find more rational, specific, and effective treatments and prognostic targets.

Increasing numbers of studies have indicated that next-generation sequencing techniques may provide new insights into the alteration of gene expression during tumorigenesis and help to better identify clinical biomarkers of cancer [6,7]. The application of high-throughput methods enables researchers to identify RNAs as biomarkers in cancer diagnosis and treatment. The discovery of long non-coding RNA (lncRNA) has dramatically altered our understanding of cancer. Generally, lncRNAs are >200 nucleotides (nt) long with no or limited protein-coding capacity [8]. The latest research shows that lncRNA expression is frequently dysregulated in cancer, and specific lncRNAs are correlated with cancer recurrence, metastasis, and poor prognosis in different kinds of cancer, including LUAD [9,10]. Because of the strong tissue specificity, lncRNAs may be early diagnostic biomarkers for various cancers [11]. However, much still remains unknown about the mechanics and significance of lncRNAs in many aspects of this disease, such as carcinogenesis, development, metastasis, response to anti-cancer treatment, and prognosis. Therefore, the identification of specific lncRNA biomarkers related to the prognosis and diagnosis of LUAC has important clinical significance.

Competing endogenous RNAs (ceRNAs) are proposed as tumor-specific regulatory pathways that affect protein levels [12]. The central concept is that messenger RNA (mRNA) and lncRNA share one or more microRNA (miRNA) response elements (MREs) and may act as natural miRNA sponges that lead to the down-regulation of the intracellular miRNA function [13]. Studies have demonstrated that various types of RNAs, such as mRNA, pseudogenes, lncRNA, circular RNA, and miRNA, can communicate with each other through ceRNA mechanisms, thereby regulating various tumor cells and their microenvironment, and affecting tumor proliferation and migration [14]. The introduction of ceRNA theory has revealed a new mechanism for interaction between RNAs, which has broadened our understanding of lncRNA function [15]. In LUAD, the potential of lncRNA to act as a biomarker for cancer prognosis has been widely reported [16–18]. Nevertheless, much still remains unknown about the ceRNA in LUAD and targets for tumor therapy.

In the present study, we comprehensively analyzed and provided an overview of the lncRNAs, miRNAs, and mRNAs in LUAD from The Cancer Genome Atlas (TCGA) database. Three differentially expressed RNAs were used to construct a lncRNA–miRNA–mRNA regulated ceRNA network. Functional enrichment analyses were performed to reveal and determine the functional roles and underlying mechanisms of the ceRNA network in LUAD. Next, we performed survival analysis, univariate, and multivariate Cox proportional hazard regression analysis of three RNAs in the ceRNA network, and established three new types of RNA models to predict the prognosis of LUAD patients. Through the construction of a ceRNA network and Cox regression models, we aimed to elucidate the regulatory mechanisms of lncRNA in the prognosis of LUAD to identify novel biomarkers that can effectively predict clinical outcomes.

Materials and methods
Data resources and pretreatment

Individual lung adenocarcinoma RNA sequencing (RNA-seq) data (level 3) and the corresponding clinical data were obtained from the TCGA database. The TCGA database is a public platform with more than 30 cancer types and clinical pathological information for at least 11,000 patients. It has been widely used by a large number of researchers to explore the genetic basis of tumors through high-throughput sequencing. The RNA-seq data were generated by Illumina HiSeq RNA-seq and Illumina HiSeq miRNA-Seq platforms. Exclusion criteria were set as follows: (1) histological diagnosis negating LUAD; (2) presence of a malignancy other than LUAD; (3) lack of complete clinical data. The three RNA expression profiles were integrated and extracted by using the R bioconductor package TCGABiolinks. Genes were annotated using the Ensembl online database. The present study was in compliance with the publication guidelines provided by TCGA, and the data obtained from TCGA did not require approval from an ethics committee.

Identification of differentially expressed lncRNAs, miRNAs, and mRNAs

To identify the differentially expressed lncRNAs (DElncRNA), miRNAs (DEmiRNA), and mRNAs (DEmRNA) between the LUAD and the normal samples, the downloaded lncRNA, miRNA, and mRNA data were standardized and differential expression analysis was performed using the edgeR package in R software. Filtering criteria for the differential expression of these three RNAs were as follows: (1) at least 25% of the samples have a gene expression level greater than 2; (2) the original RNA sequencing was normalized using the trimmed mean of M-values (TMM)
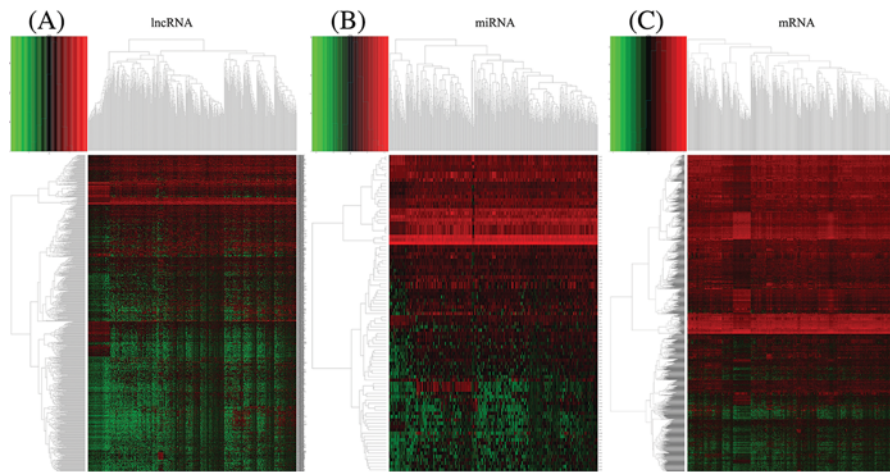
Figure 1. Clustered heat maps of the differentially expressed RNAs in LUAD

The rows represent RNAs and columns represent the samples. $|\log_2 FC| > 2$, FDR < 0.01. (**A–C**) Differentially expressed lncRNAs, miRNAs, and mRNAs in LUAD. FC: fold change, FDR: false discovery rate, LUAD: lung adenocarcinoma, lncRNAs: long noncoding RNAs, miRNAs: microRNAs, mRNAs: messenger RNAs.
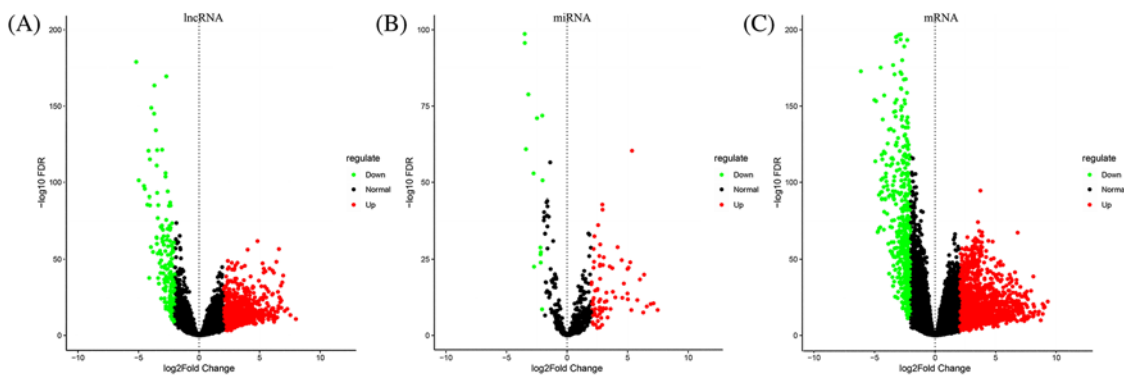

Figure 2. Volcano plot of differentially expressed RNAs in LUAD

$|\log_2 FC| > 2$, FDR < 0.01. (**A–C**) Differentially expressed lncRNAs, miRNAs, and mRNAs in LUAD. The red dot represents up-regulated RNAs and green dot represents down-regulated RNAs.

method; (3) differential expression thresholds of $|\log_2 \text{FoldChange}|$ (FC) > 2 and false discovery rate (FDR) or adjusted P value < 0.01. Heat map clustering and a volcano plot of the DElncRNAs, DEmiRNAs, and DEmRNAs were plotted using the R software gplots package.

Establishment of the lncRNA–miRNA–mRNA ceRNA network

Based on the hypothesis that lncRNA can act as a sponge for miRNA and prevent miRNA from binding to its target genes, and can directly interact with mRNA [19,20], the lncRNA–miRNA–mRNA ceRNA network was constructed by the following steps: (1) the miRcode online tool (http://www.mircode.org) was used to predict the potential miRNAs targeted by DElncRNAs and the lncRNA–miRNA interaction pairs; (2) the miRNA sequences were modified and the names of DEmiRNAs were transformed into human mature miRNA names by using the starBase v2.0 database (http://starbase.sysu.edu.cn) [21]; (3) miRDB (http://www.mirdb.org/) [22], miRTarBase (http://mirtarbase.mbc.nctu.edu.tw/) [23], and TargetScan (http://www.targetscan.org/) [24] were used to predict target genes of the DEmiRNAs and establish the miRNA–mRNA interaction pairs; (4) the online website Venny was used to compare the target genes with DEmRNAs and the target genes that overlapped with DEmRNAs in the present study were selected for further analysis. Finally, combined with the lncRNA–miRNA pairs and miRNA–mRNA pairs, we established the lncRNA–miRNA–mRNA ceRNA network. Cytoscape software version 3.6.1was used to visualize the resulting network.
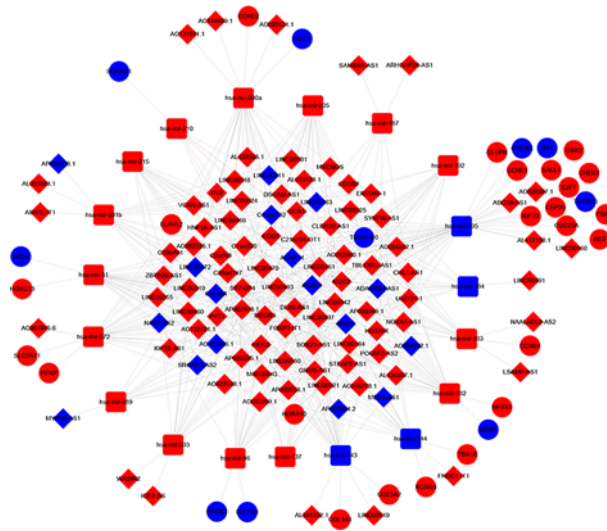
Figure 3. ceRNA networks of lung adenocarcinoma
Red represents up-regulation, and blue represents down-regulation. LncRNAs, miRNAs, and mRNAs in the networks are represented as diamonds, round rectangles, and circles, respectively.

Functional enrichment analysis of the ceRNA network

Based on the correlation between the three RNAs, we inferred the potential biological processes and pathways of abnormal expression of RNAs in ceRNA networks by analyzing mRNAs in networks. The Annotate, Visualize, and Integrate Discovery Database (DAVID6.8) (http://david.abcc.ncifcrf.gov/) online tool was used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. The results of GO analysis mainly consist of three parts: biological process (BP), cellular component (CC), and molecular function (MF). GO and KEGG terms with $P < 0.05$ were considered statistically significant, and the "Goplot" package in R was used to visualize the results as chord plots.

Overall survival analysis and construction of the LUAD-specific prognostic model

We utilized the survival package to generate Kaplan–Meier separate survival analysis and log-rank tests for the DElncRNAs, DEmiRNAs, and DEmRNAs in the ceRNA network to determine the overall survival (OS) relationship with LUAD patients in TCGA. A P value less than 0.05 was considered as statistically significant. The associations of three types of RNAs with OS were assessed by univariate Cox regression. Only those genes with a $P < 0.01$ were identified as candidate prognosis related biomarkers. Then, those candidate genes were analyzed by multivariate Cox regression to remove any genes that might not be independent factors in prognosis prediction. Finally, based on the multivariate Cox hazards regression model to determine the LUAD independent prognostic factors, the lncRNA, miRNA, and mRNA signature scores were constructed, and the prognostic index (PI) of OS was predicted as follows:

$$Prognostic\ index = Exp_{RNAI} * \beta_{RNAI} + Exp_{RNA2} * \beta_{RNA2} + \ \bullet \ \bullet \ \bullet \ Exp_{RNAn} * \beta_{RNAn}$$

(where "Exp" denotes the expression level of DERNAs, and "β" is the regression coefficient obtained from the multivariate Cox regression model) [25]. LUAD patients were stratified into high-risk and low-risk groups with a median risk score as a threshold. The survival package in R was used to plot the Kaplan–Meier (K-M) survival curve of the two groups. In addition, using the survival ROC package in R, we constructed time-dependent receiver operating characteristic (ROC) curves within 1,3 ,5-year and estimated the area under the ROC curve (AUC) for the three RNA models.

Results
Patient characteristics and differentially expressed RNAs

The LUAD transcriptome profiling data and the corresponding clinical information were obtained using the R bioconductor package TCGABiolinks. Gene expression quantification data for a total of 594 samples, which included 535
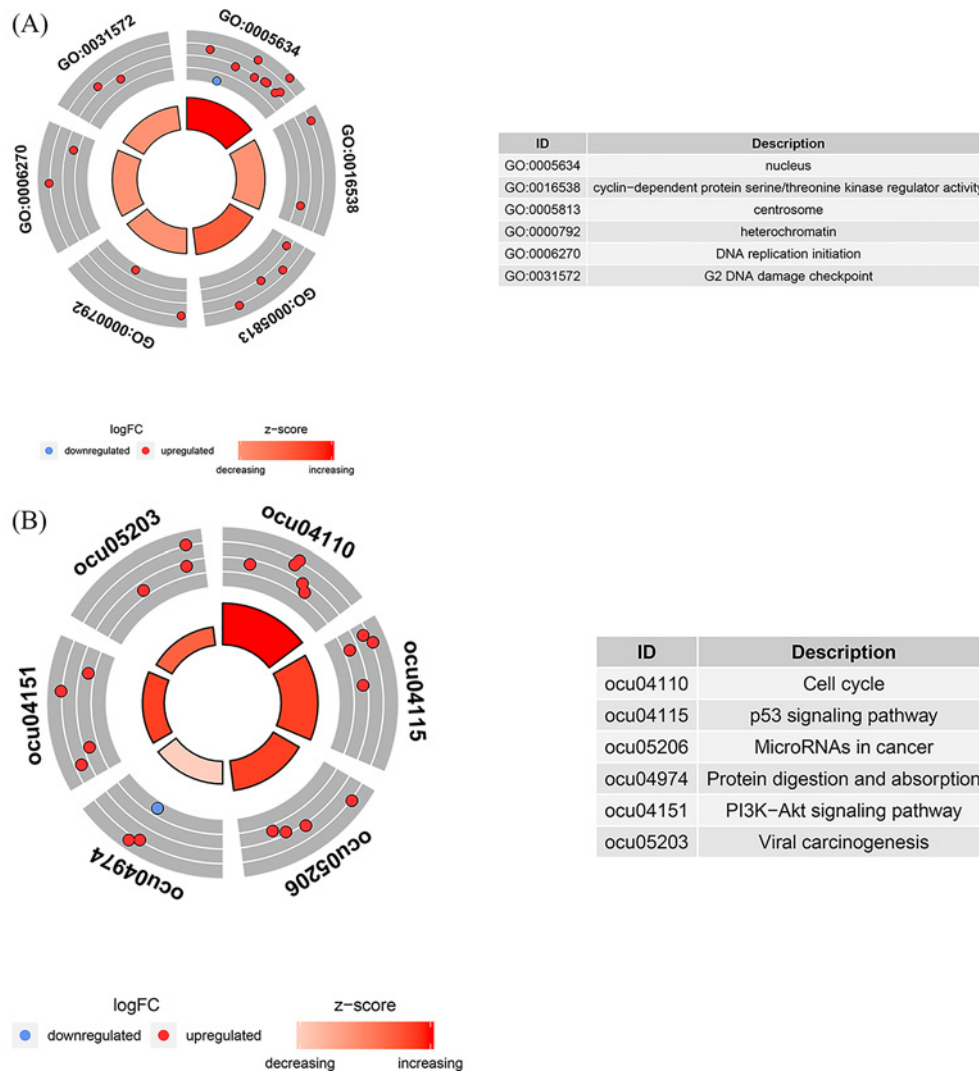
Figure 4. Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of genes included in the ceRNA network

(**A**) Chord plot display of the relationship between genes and GO analysis. (**B**) Chord plot display of the relationship between genes and KEGG pathways.

tumor samples and 59 normal samples, and miRNA expression quantification data for 567 samples, which included 521 tumor samples and 46 normal samples. Using the edgeR package in R with the cut-off criteria of $|\log_2 FC| > 2$ and FDR < 0.01, we identified 1123 DElncRNAs (940 up-regulated lncRNAs, 183 down-regulated lncRNAs), 95 DEmiR-NAs (81 up-regulated miRNAs, 14 down-regulated miRNAs), and 2296 DEmRNAs (1781 up-regulated mRNAs, and 515 down-regulated mRNAs). The heat map shown in Figure 1 depicted the variable gene expression in the LUAD and normal samples and the volcano plots visually demonstrating the distribution of DElncRNAs, DEmiRNAs, and DEmRNAs were shown in Figure 2.

Construction of the ceRNA network in lung adenocarcinoma

According to the relationship between DElncRNAs, DEmiRNAs, and DEmRNAs, we established a lncRNA–miRNA–mRNA ceRNA network in LUAD. (1) The online website miRcode was used to clarify the correlation between miRNAs targeting lncRNAs in the ceRNA cross-talk, the results showed that 102 DElncRNAs could interact with 86 miRNAs, but only 19 out of 86 miRNAs were utilized in relation in our study; (2) using the miRDB, miTarBase, and TargetScan online databases, 19 miRNAs were found to interact with 667 target genes. At

Table 1 Univariate Cox analysis of prognosis-associated RNAs in a ceRNA network of LUAD

RNA	HR	CI	*P*-value	Expression
lncRNA				
C20orf197	0.86	0.8–0.93	0.000160092	Up
C5orf64	0.81	0.7–0.95	0.007033541	Down
LINC00518	1.22	1.12–1.33	3.09E–06	Up
MUC2	1.08	1.03–1.13	0.003035376	Up
UCA1	1.09	1.03–1.14	0.001014659	Up
LINC00337	1.12	1.04–1.22	0.005081888	Up
MED4-AS1	0.82	0.71–0.94	0.004836856	Down
LINC00460	1.09	1.03–1.15	0.002507511	Up
SYNPR-AS1	0.90	0.85–0.97	0.00294044	Down
NAV2-AS2	0.84	0.75–0.94	0.003476934	Up
AP003064.2	0.87	0.78–0.96	0.007660574	Down
AP002478.1	1.11	1.04–1.18	0.00086024	Up
mRNA				
PFKP	1.28	1.13–1.45	0.000118643	Up
CLSPN	1.19	1.07–1.33	0.001424106	Up
RS1	0.87	0.79–0.96	0.00391744	Down
CCNE1	1.17	1.06–1.3	0.0018306	Up
CCNB1	1.32	1.16–1.49	1.28E–05	Up
KIF23	1.25	1.12–1.4	6.99E–05	Up
CEP55	1.24	1.11–1.38	0.000168944	Up
CHEK1	1.33	1.16–1.52	3.67E–05	Up
E2F7	1.25	1.13–1.39	1.14E–05	Up
miRNA				
hsa-mir-31	1.10	1.04–1.17	1.04–1.17	Up

Notes: ceRNA, competing endogenous RNA; LUAD, lung adenocarcinom; HR, hazard ratio; CI, confidence interval.

the intersection of these 667 target genes and 2296 DEmRNAs, 33 overlapping genes were obtained; (3) based on the DElncRNAs–DEmiRNAs pairs and DEmiRNAs–DEmRNAs pairs, the ceRNA network was constructed and visualized with the Cytoscape software (Figure 3), which comprised of 102 lncRNAs (17 down-regulated, 85 up-regulated), 19 miRNAs (4 down-regulated, 15 up-regulated), and 33 mRNAs (10 down-regulated, 23 up-regulated).

GO and KEGG enrichment analysis of DEmRNAs in ceRNA

To better understand the underlying mechanisms of ceRNA networks, we performed functional enrichment analysis of mRNAs in the network with the DAVID database. The GO enrichment analysis showed that six significant terms are involved in LUAD ($P < 0.05$) (Figure 4A). Among them, the BP enrichment process for GO analysis mainly included DNA replication initiation and G2 DNA damage checkpoint. MF enrichment revealed mainly cyclin-dependent protein serine/threonine kinase regulator activity, and CC enrichment mainly included nucleus, centrosome, and heterochromatin. KEGG pathway analysis focused on biological pathways, and a total of five pathways were found to be significantly enriched; cell cycle, p53 signaling pathway, microRNAs in cancer, protein digestion and absorption, and PI3K-Akt signaling pathway. All mentioned pathways are exhibited in Figure 4B.

Survival analysis and establishment of the three RNA type prognostic model

The relationship between DElncRNAs, DEmiRNAs, and DEmRNAs with OS in the ceRNA network was studied in detail. With $P < 0.05$ as the standard, a total of 22 lncRNAs, 12 mRNAs, and 1 miRNA were identified to be associated with prognosis. These significant survival-related lncRNAs, miRNAs, and mRNAs were further submitted to univariate and multivariate Cox regression analyses to determine independent prognostic factors for LUAD. Subsequently, 12 lncRNAs, 9 mRNAs, and 1 miRNA were obtained using $P < 0.01$ as the initial screening criteria for univariate Cox regression analysis. The results were shown in Table 1. Next, we constructed the multivariate regression models with the three RNAs based on the univariate results. The three prognostic signatures contained 8 lncRNAs (*AP002478.1, AP003064.2, C5orf64, C20orf197, LINC00460, LINC00518, MED4-AS1, MUC2*), 2 mRNAs (*E2F7, PFKP*), and
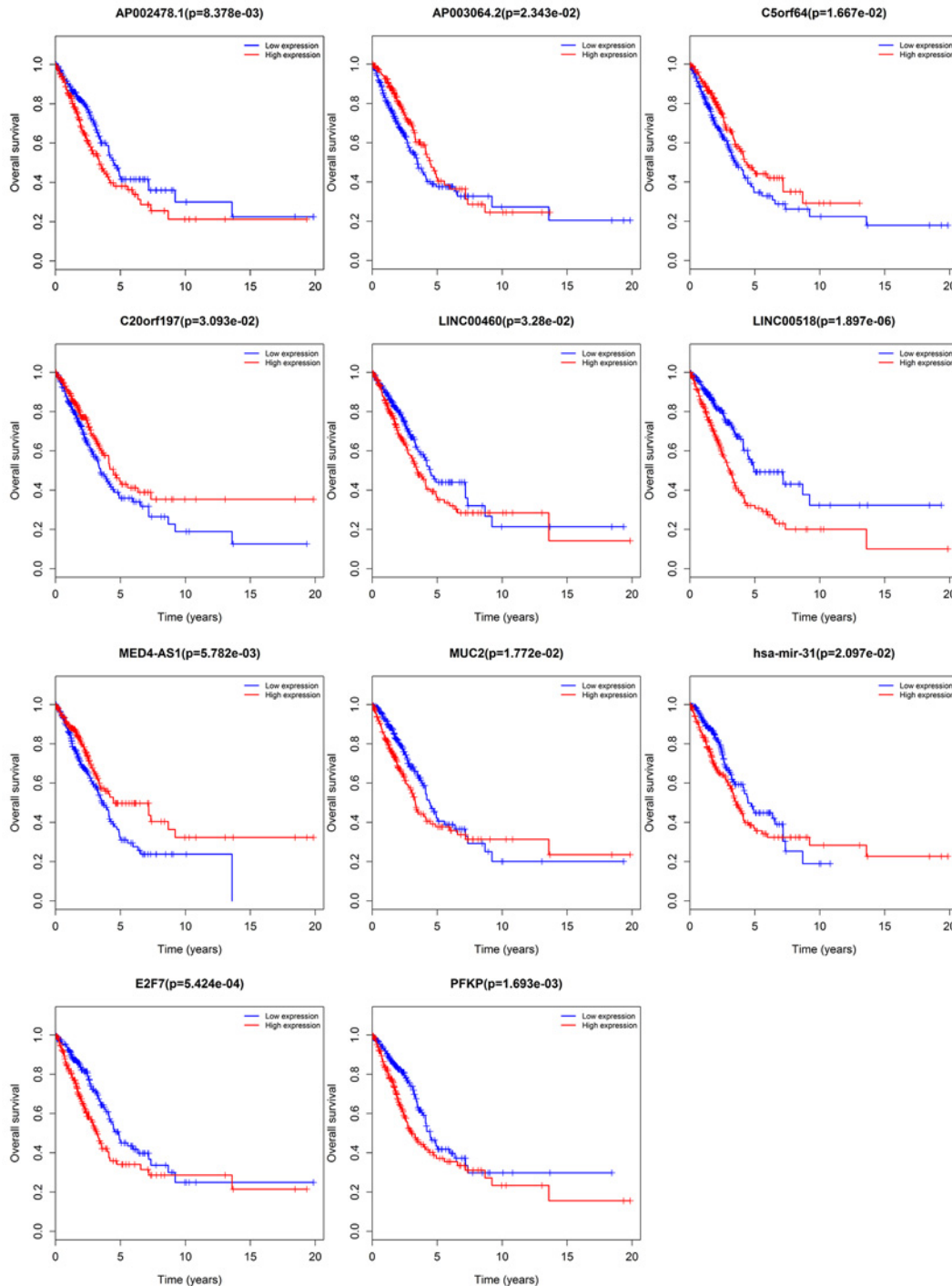
Figure 5. Kaplan–Meier survival curves and Log–rank for DElncRNAs, DEmiRNAs, and DEmRNAs as independent prognostic factors associated with OS in LUAD

Horizontal axis is OS time (years) and vertical axis stands for survival function.

1 miRNA (*hsa-mir-31*), and the relevant survival curves were showed in Figure 5. Using three signatures, we calculated the risk score for each patient separately and ranked them according to the increased risk score. According to the median risk score, LUAD patients were divided into high-risk and low-risk groups. K-M analysis showed significant differences of OS between high-risk and low-risk groups in all three models (Figures 6–8A). Our study predicted 1-, 3-, and 5-year survival rate accurately for LUAD patients. The time-dependent receiver operating characteristic
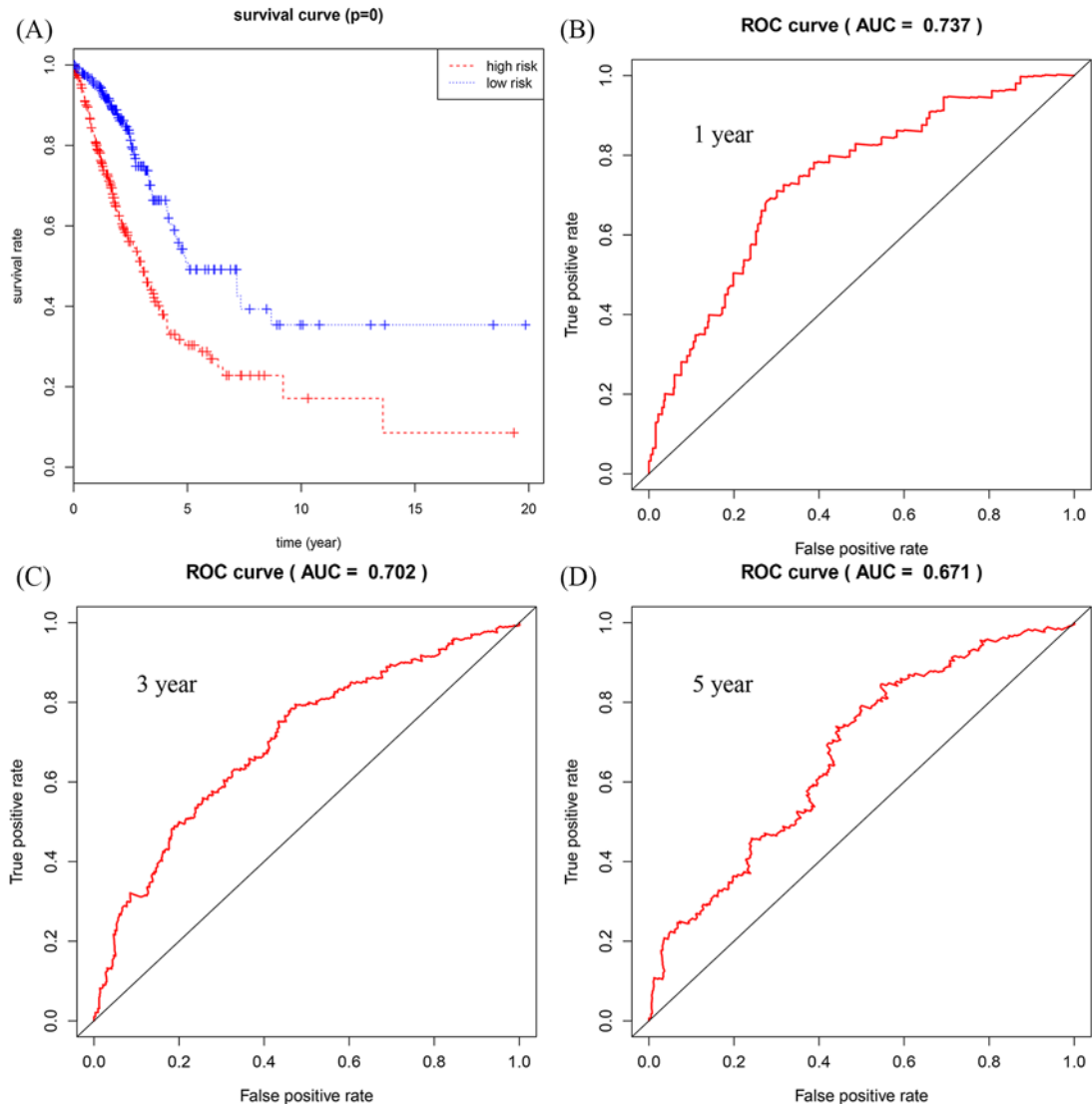
Figure 6. Construction of survival risk score system based on lncRNAs signature
Cox regression analysis for survival prediction by the 8-lncRNA signature screened from 22 most significantly expressed lncRNAs with $P < 0.01$. (**A**) The survival curve of patients with high risk and low risk. (**B–D**) The ROC curve in 1, 3, 5 years with AUC value.

(ROC) curves for lncRNAs (Figure 6B–D), mRNAs (Figure 7B–D), and miRNA (Figure 8B–D) signature have area under curve (AUC) values higher than 0.5. Among them, the PI of lncRNA showed good prognostic ability, which suggests that the prognostic model constructed by 8 prognostic lncRNA may be a good predictor of LUAD.

Discussion

Lung adenocarcinoma is a multifactorial disease with high heterogeneity and mixed genetic factors. Most patients with LUAD are diagnosed in the advanced stage and have missed the opportunity for surgery [26]. Despite the improvement of lung cancer diagnosis, surgical techniques, and new drug treatments, the survival rate is still at a lower level [27]. With recent advancements, we are moving to a new era of personalized target therapies. It is becoming increasingly clear that many complex diseases, especially cancer, can rarely be attributed to a single genomic variation [28]. Therefore, it is very important to develop genomic detection of lung adenocarcinoma and predict the survival rate of patients according to their individual characteristics. A growing body of experimental evidence suggests that lncRNA plays an important role in many biological processes and, as a ceRNA, its involved in the occurrence and development of LUAD and can become a potential diagnostic and prognostic marker [29,30]. The ceRNA hypothesis
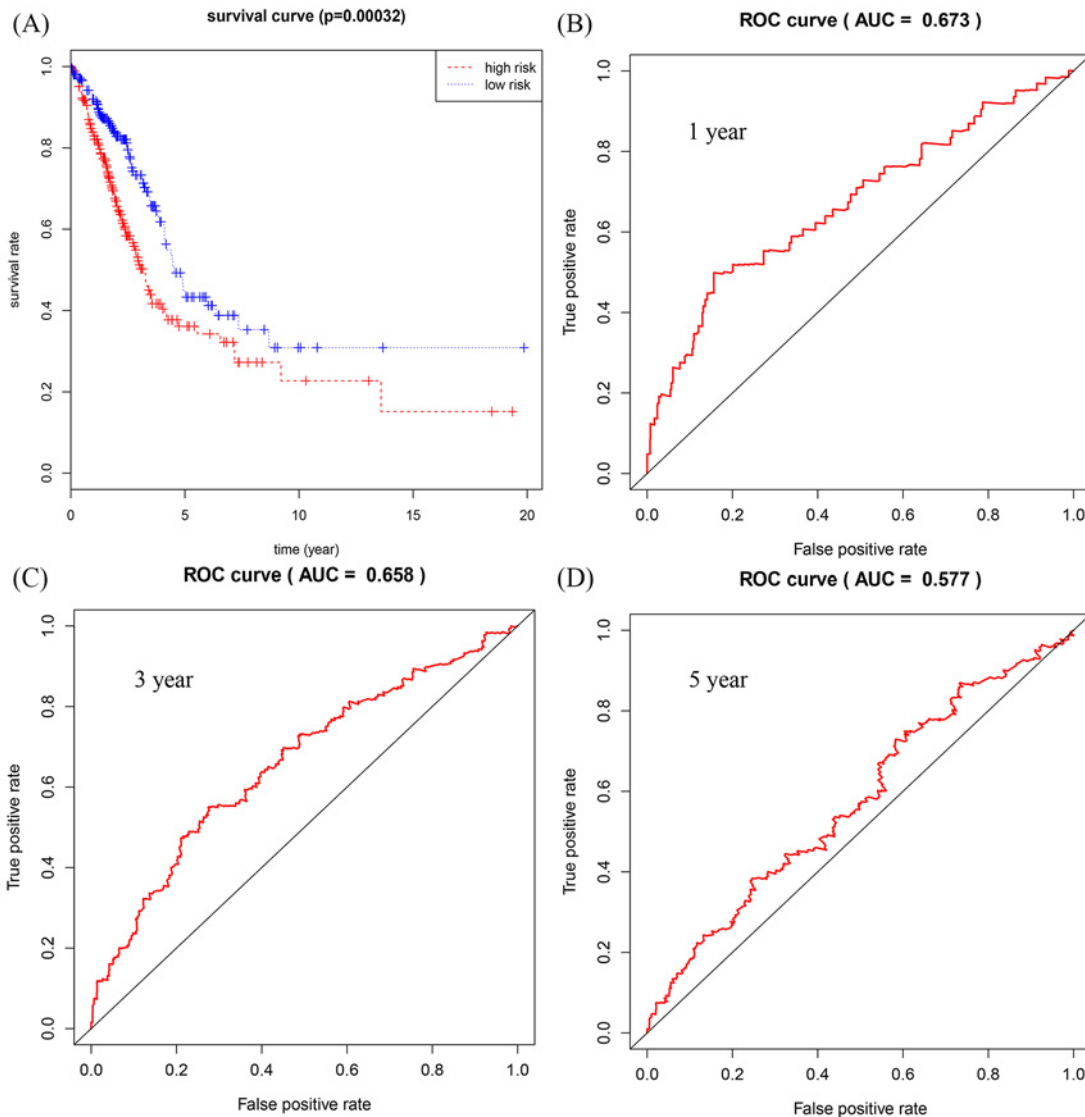
Figure 7. Construction of survival risk score system based on mRNAs signature

Cox regression analysis for survival prediction by the 2-mRNA signature screened from 9 most significantly expressed mRNAs with $P < 0.01$. (**A**) The survival curve of patients with high risk and low risk. (**B–D**) The ROC curve in 1, 3, 5 years with AUC value.

reveals a new mechanism of RNA interaction and provides important clues and theoretical guidance for further understanding the tumorigenesis mechanism [12,15,31]. However, the expression pattern of ceRNA against LUAD and its prognostic value have not been thoroughly studied. In addition, compared with single-gene analysis, the prognostic characteristics based on multiple genes can help to predict the survival rate more accurately.

In our study, we identified DElncRNAs, DEmRNAs, and DEmiRNAs according to a large number of LUAD tissue samples from TCGA database, and constructed a ceRNA network, based on which we developed a genome-wide prognosis model. Our study has a deeper understanding of LUAD overall and also provides a new dimension for predicting the prognosis of individual patients with LUAD. In order to further study the potential functions of ceRNA networks, we performed GO and KEGG functional enrichment on mRNAs that were abnormally expressed in the network. The GO analysis revealed that the function of the aberrantly expressed mRNAs mainly concentrates on the nucleus, centrosome, and heterochromatin, which could participate in DNA replication and damage detection, and regulate the activity of cyclin-dependent protein serine/threonine kinases. Through KEGG pathway analysis, we detected a number of cancer-related pathways, including the p53 signaling pathway, the PI3K-Akt signaling pathway, and microRNAs in cancer. Studies have shown that *PPM1D* and *GADD45B* may regulate the progress of LUAD
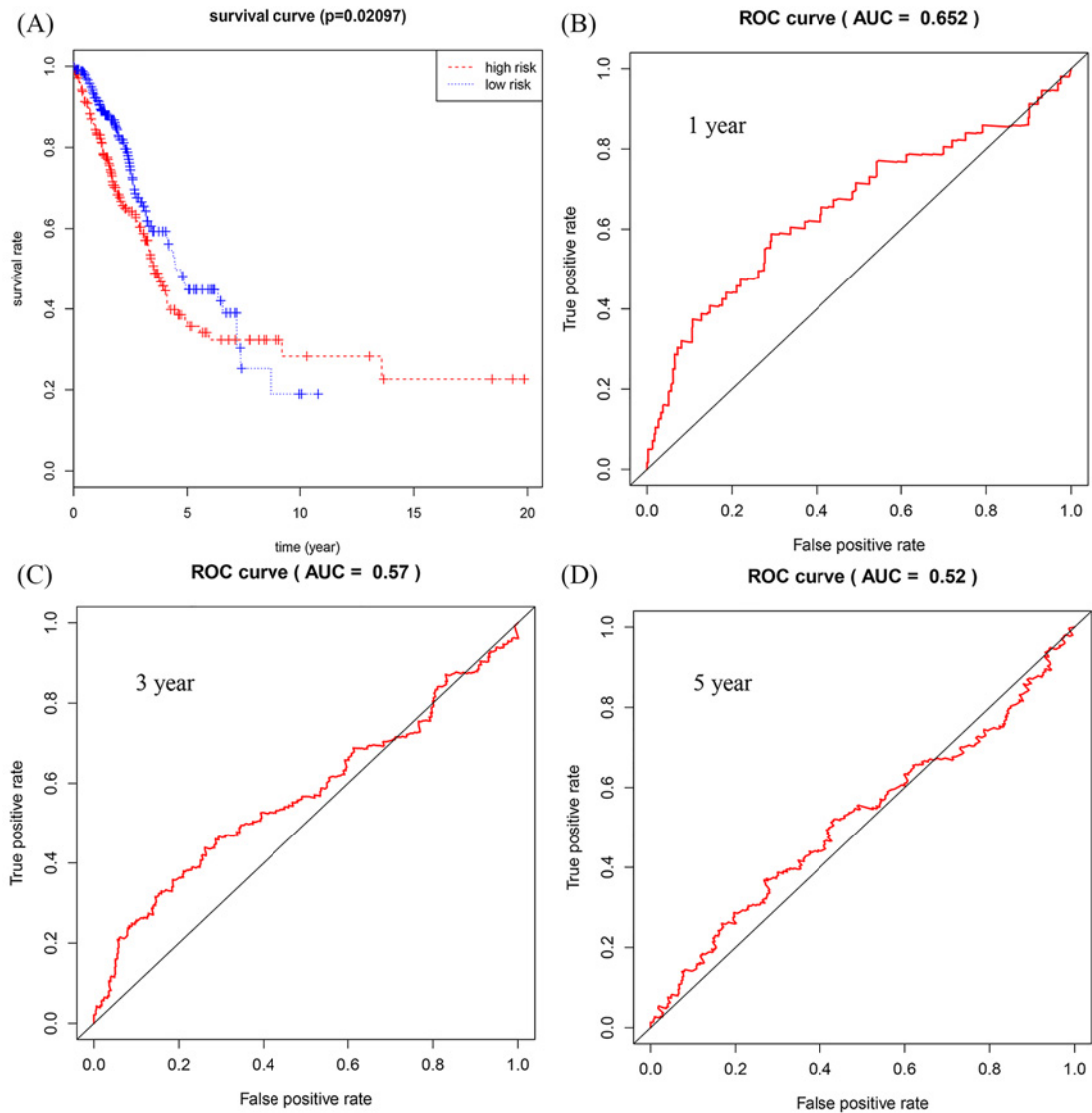
Figure 8. Construction of survival risk score system based on miRNAs signature

Cox regression analysis for survival prediction by the 1-miRNA signature with $P < 0.01$. (**A**) The survival curve of patients with high risk and low risk. (**B–D**) The ROC curve in 1, 3, 5 years with AUC value.

through the p53 signaling pathway [32], and that down-regulation of *CDKN2B-AS1* and *CDKN2A* may activate the p53 signaling pathway to promote lung cancer formation [33]. BOS-102 induces apoptosis and cell cycle arrest through reactive oxygen species (ROS)-mediated activation of the PI3K-Akt signaling pathway [34]. All these above views demonstrate that our ceRNA network reflects vital mechanisms of LUAD.

In addition, through the survival analysis of the RNAs in the network and the construction of the prognosis model, we found that the model constructed by 8 lncRNAs can be used as an independent predictor of OS in patients with LUAD. In the present study, there were eight prognostic lncRNAs and seven other lncRNAs besides *C20orf19* had shorter overall survival. To date, none of the lncRNAs *C20orf197*, *C5orf64*, *MED4*-AS1, or *AP003064.2* have been reported to have disease associations in any medical field. For the first time, we found that these four novel lncRNAs play a vital role in LUAD. Studies have shown that highly expressed *AP002478.1* has a poor prognosis in gastric cancer and hepatocellular carcinoma [35,36], and our study in LUAD also showed the same results, this suggests that *AP002478.1* plays an important role as a carcinogenic factor in many kinds of cancers. *LINC00518* was first discovered and developed as a non-invasive 2-gene molecule in the detection of cutaneous melanoma [37]. *LINC00518*

was considered as a competitive endogenous RNA candidate for predicting breast cancer prognosis and could reduce multidrug resistance [25,38]. Its overexpression promotes paclitaxel resistance in prostate cancer by isolating $miR-216b-5$p [39]. MUC2 is the major mucin in the formation of intestinal epithelium [40]. Numerous studies have shown that the expression of MUC2 is related to the invasion and metastasis of various malignant tumors, including gastric cancer, gallbladder cancer, breast cancer, ovarian cancer, and lung cancer [41–45]. *LINC00460*, located on chromosome 13q33.2. has been shown to be up-regulated in cancers such as non-small cell lung cancer, esophageal squamous cell carcinoma, nasopharyngeal carcinoma, and thyroid carcinoma [46–49]. It can promote the migration and invasion of non-small cell lung cancer cells, induce epithelial–mesenchymal transition, and inhibit the growth of esophageal squamous cell carcinoma cells by regulating cell proliferation and apoptosis. These reports suggest that lncRNA can have a profound impact on the prognosis of LUAD, improve the accuracy of prognosis prediction, and emphasize the prognostic value of lncRNAs.

We built a genome-wide prognostic model based on a large number of samples from the TCGA database. However, several limitations must be noted. First, a longer follow-up time is needed to verify the reliability of our model. Second, these eight prognostic lncRNAs need to be validated experimentally, which will be carried out in our follow-up study.

Conclusion

In conclusion, we identified prognostic biomarkers of lung adenocarcinoma based on competitive endogenous RNA network and Cox regression analysis, and constructed a model containing eight lncRNAs to reliably predict the prognosis of LUAD patients. Our research will contribute in further understanding the pathogenesis of LUAD and lay the foundation for future clinical research.

Competing Interests

Yan Yao and Changgang Sun conceived and designed the experiments; Tingting Zhang and lingyu Qi performed the experiments; Ruijuan Liu ang Gongxi Liu analyzed the data; Jia Wang and Qi Song contributed reagents/materials/analysis tools; Tingting Zhang prepared the figures and/or tables; Yan Yao drafted the work or revised it critically for important content; Changgang Sun approved the final draft of the manuscript submitted for review and publication.

Funding

Author Contribution

The authors declare that there are no competing interests associated with the manuscript.

Abbreviations

AUC, area under curve; BP, biological process; CC, cellular component; ceRNA, competing endogenous RNA; DElncRNA, differential expression of lncRNA; DEmiRNA, differential expression of miRNA; DEmRNA, differential expression of mRNA; GO, Gene ontology; KEGG, Kyoto Encyclopedia of Gene and Genome; LUAD, lung adenocarcinom; lncRNA, long non-coding RNA; MF, molecular function; miRNA, microRNA; mRNA, messenger RNA; OS, overall survival; TCGA, the cancer genome atlas.

References

1 Torre, L.A., Bray, F., Siegel, R.L., Ferlay, J., Lortet-Tieulent, J. and Jemal, A. (2015) Global cancer statistics, 2012. *CA Cancer J. Clin.* **65**, 87–108, https://doi.org/10.3322/caac.21262

2 Siegel, R.L. and Miller, K.D. (2019) Cancer statistics, 2019. *CA Cancer J. Clin.* **69**, 7–34, https://doi.org/10.3322/caac.21551

3 Alam, H., Li, N., Dhar, S.S., Wu, S.J., Lv, J., Chen, K. et al. (2018) HP1gamma Promotes Lung Adenocarcinoma by Downregulating the Transcription-Repressive Regulators NCOR2 and ZBTB7A. *Cancer Res.* **78**, 3834–3848, https://doi.org/10.1158/0008-5472.CAN-17-3571

4 Forouzanfar, M.H., Alexander, L.T., Anderson, H.R. et al. (2016) Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1659–1724, A.A., https://doi.org/10.1016/S0140-6736(16)31679-8

5 Morgensztern, D., Ng, S.H., Gao, F. and Govindan, R. (2010) Trends in stage distribution for patients with non-small cell lung cancer: a National Cancer Database survey. *J. Thorac. Oncol.* **5**, 29–33, https://doi.org/10.1097/JTO.0b013e3181c5920c

6 Hu, Z., Fan, C., Oh, D.S., Marron, J.S., He, X., Qaqish, B.F. et al. (2006) The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* **7**, 96, https://doi.org/10.1186/1471-2164-7-96

7 Rajendran, V., Gopalakrishnan, C. and Sethumadhavan, R. (2018) Pathological role of a point mutation (T315I) in BCR-ABL1 protein-A computational insight. *J. Cell. Biochem.* **119**, 918–925, https://doi.org/10.1002/jcb.26257

8 Jathar, S., Kumar, V., Srivastava, J. and Tripathi, V. (2017) Technological Developments in lncRNA Biology. *Adv. Exp. Med. Biol.* **1008**, 283–323, https://doi.org/10.1007/978-981-10-5203-3˙10

9 Xie, W., Yuan, S., Sun, Z. and Li, Y. (2016) Long noncoding and circular RNAs in lung cancer: advances and perspectives. *Epigenomics* **8**, 1275–1287, https://doi.org/10.2217/epi-2016-0036

10 Yang, G., Lu, X. and Yuan, L. (2014) LncRNA: a link between RNA and cancer. *Biochim. Biophys. Acta* **1839**, 1097–1109, https://doi.org/10.1016/j.bbagrm.2014.08.012

11 Hung, T. and Chang, H.Y. (2010) Long noncoding RNA in genome regulation: prospects and mechanisms. *RNA Biol.* **7**, 582–585, https://doi.org/10.4161/rna.7.5.13216

12 Salmena, L., Poliseno, L., Tay, Y., Kats, L. and Pandolfi, P.P. (2011) A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* **146**, 353–358, https://doi.org/10.1016/j.cell.2011.07.014

13 Ergun, S. and Oztuzcu, S. (2015) Oncocers: ceRNA-mediated cross-talk by sponging miRNAs in oncogenic pathways. *Tumour Biol.* **36**, 3129–3136, https://doi.org/10.1007/s13277-015-3346-x

14 Sanchez-Mejias, A. and Tay, Y. (2015) Competing endogenous RNA networks: tying the essential knots for cancer biology and therapeutics. *J. Hematol. Oncol.* **8**, 30

15 Tay, Y., Rinn, J. and Pandolfi, P.P. (2014) The multilayered complexity of ceRNA crosstalk and competition. *Nature* **505**, 344–352, https://doi.org/10.1038/nature12986

16 Wang, Y., Zhang, F., Wang, J., Hu, L., Jiang, F., Chen, J. et al. (2018) lncRNA LOC100132354 promotes angiogenesis through VEGFA/VEGFR2 signaling pathway in lung adenocarcinoma. *Cancer Manag.Res.* **10**, 4257–4266, https://doi.org/10.2147/CMAR.S177327

17 Lu, Q.C., Rui, Z.H., Guo, Z.L., Xie, W., Shan, S. and Ren, T. (2018) LncRNA-DANCR contributes to lung adenocarcinoma progression by sponging miR-496 to modulate mTOR expression. *J. Cell. Mol. Med.* **22**, 1527–1537, https://doi.org/10.1111/jcmm.13420

18 Peng, F., Wang, R., Zhang, Y., Zhao, Z., Zhou, W., Chang, Z. et al. (2017) Differential expression analysis at the individual level reveals a lncRNA prognostic signature for lung adenocarcinoma. *Mol. Cancer* **16**, 98, https://doi.org/10.1186/s12943-017-0666-z

19 Guo, L.L., Song, C.H., Wang, P., Dai, L.P., Zhang, J.Y. and Wang, K.J. (2015) Competing endogenous RNA networks and gastric cancer. *World J. Gastroenterol.* **21**, 11680–11687, https://doi.org/10.3748/wjg.v21.i41.11680

20 Xu, Y., Yang, X., Wang, T., Yang, L., He, Y.Y., Miskimins, K. et al. (2018) Knockdown delta-5-desaturase in breast cancer cells that overexpress COX-2 results in inhibition of growth, migration and invasion via a dihomo-gamma-linolenic acid peroxidation dependent mechanism. *BMC Cancer* **18**, 330, https://doi.org/10.1186/s12885-018-4250-8

21 Li, J.H., Liu, S., Zhou, H., Qu, L.H. and Yang, J.H. (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **42**, D92–D97, https://doi.org/10.1093/nar/gkt1248

22 Wong, N. and Wang, X. (2015) miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.* **43**, D146–D152, https://doi.org/10.1093/nar/gku1104

23 Chou, C.H., Shrestha, S., Yang, C.D., Chang, N.W., Lin, Y.L., Liao, K.W. et al. (2018) miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **46**, D296–D302, https://doi.org/10.1093/nar/gkx1067

24 Agarwal, V., Bell, G.W., Nam, J.W. and Bartel, D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4**, 1–38, https://doi.org/10.7554/eLife.05005

25 Fan, C.N., Ma, L. and Liu, N. (2018) Systematic analysis of lncRNA-miRNA-mRNA competing endogenous RNA network identifies four-lncRNA signature as a prognostic biomarker for breast cancer. *J. Transl. Med.* **16**, 264, https://doi.org/10.1186/s12967-018-1640-2

26 Kerr, K.M. (2009) Pulmonary adenocarcinomas: classification and reporting. *Histopathology* **54**, 12–27, https://doi.org/10.1111/j.1365-2559.2008.03176.x

27 Aokage, K., Yoshida, J., Hishida, T., Tsuboi, M., Saji, H., Okada, M. et al. (2017) Limited resection for early-stage non-small cell lung cancer as function-preserving radical surgery: a review. *Jpn. J. Clin. Oncol.* **47**, 7–11, https://doi.org/10.1093/jjco/hyw148

28 Nibbe, R.K., Chowdhury, S.A., Koyuturk, M., Ewing, R. and Chance, M.R. (2011) Protein-protein interaction networks and subnetworks in the biology of disease, Wiley interdisciplinary reviews. *Syst. Biol. Med.* **3**, 357–367

29 Kumar, M.S., Armenteros-Monterroso, E., East, P., Chakravorty, P., Matthews, N., Winslow, M.M. et al. (2014) HMGA2 functions as a competing endogenous RNA to promote lung cancer progression. *Nature* **505**, 212–217, https://doi.org/10.1038/nature12785

30 Wang, X., Ding, Y., Da, B., Fei, Y. and Feng, G. (2018) Identification of potential prognostic long noncoding RNA signatures based on a competing endogenous RNA network in lung adenocarcinoma. *Oncol. Rep.* **40**, 3199–3212

31 Tang, J., Zhuo, H., Zhang, X., Jiang, R., Ji, J., Deng, L. et al. (2014) A novel biomarker Linc00974 interacting with KRT19 promotes proliferation and metastasis in hepatocellular carcinoma. *Cell Death. Dis.* **5**, e1549

32 Wu, X., Zang, W., Cui, S. and Wang, M. (2012) Bioinformatics analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes. *Eur. Rev. Med. Pharmacol. Sci.* **16**, 1582–1587

33 Du, Y., Hao, X. and Liu, X. (2018) Low expression of long noncoding RNA CDKN2B-AS1 in patients with idiopathic pulmonary fibrosis predicts lung cancer by regulating the p53-signaling pathway. *Oncol. Lett.* **15**, 4912–4918

34 Guo, C.L. and Wang, L.J. (2018) A Novel Bromophenol Derivative BOS-102 Induces Cell Cycle Arrest and Apoptosis in Human A549 Lung Cancer Cells via ROS-Mediated PI3K/Akt and the MAPK Signaling Pathway. *Marine Drugs* **16**, 43, https://doi.org/10.3390/md16020043

35 Liu, Y., Zhu, J., Ma, X., Han, S., Xiao, D. and Jia, Y. (2018) ceRNA network construction and comparison of gastric cancer with or without Helicobacter pylori infection. *J. Cell. Physiol.* 7128–7140

36 Yue, C., Ren, Y., Ge, H., Liang, C., Xu, Y., Li, G. et al. (2019) Comprehensive analysis of potential prognostic genes for the construction of a competing endogenous RNA regulatory network in hepatocellular carcinoma. *OncoTargets Ther.* **12**, 561–576, https://doi.org/10.2147/OTT.S188913

37 Gerami, P., Alsobrook, II, J.P., Palmer, T.J. and Robin, H.S. (2014) Development of a novel noninvasive adhesive patch test for the evaluation of pigmented lesions of the skin. *J. Am. Acad. Dermatol.* **71**, 237–244, https://doi.org/10.1016/j.jaad.2014.04.042

38 Chang, L., Hu, Z., Zhou, Z. and Zhang, H. (2018) Linc00518 Contributes to Multidrug Resistance Through Regulating the MiR-199a/MRP1 Axis in Breast Cancer. *Cell. Physiol. Biochem.* **48**, 16–28, https://doi.org/10.1159/000491659

39 He, J., Sun, M., Geng, H. and Tian, S. (2018) Long non-coding RNA Linc00518 promotes paclitaxel resistance of the human prostate cancer by sequestering miR-216b-5p. *Biol. Cell* **111**, 39–50

40 Rodriguez-Pineiro, A.M., Bergstrom, J.H., Ermund, A., Gustafsson, J.K., Schutte, A., Johansson, M.E. et al. (2013) Studies of mucus in mouse stomach, small intestine, and colon. II. Gastrointestinal mucus proteome reveals Muc2 and Muc5ac accompanied by a set of core proteins. *Am. J. Physiol. Gastrointest. Liver Physiol.* **305**, G348–G356, https://doi.org/10.1152/ajpgi.00047.2013

41 Xiao, L.J., Zhao, S., Zhao, E.H., Zheng, X., Gou, W.F., Xing, Y.N. et al. (2012) Clinicopathological and prognostic significance of MUC-2, MUC-4 and MUC-5AC expression in japanese gastric carcinomas. *Asian Pac. J. Cancer Prev.* **13**, 6447–6453, https://doi.org/10.7314/APJCP.2012.13.12.6447

42 Hiraki, T., Yamada, S., Higashi, M., Hatanaka, K., Yokoyama, S., Kitazono, I. et al. (2017) Immunohistochemical expression of mucin antigens in gallbladder adenocarcinoma: MUC1-positive and MUC2-negative expression Is associated with vessel invasion and shortened survival. *Histol. Histopathol.* **32**, 585–596

43 Patel, D.S., Khandeparkar, S.G.S., Joshi, A.R., Kulkarni, M.M., Dhande, B., Lengare, P. et al. (2017) Immunohistochemical Study of MUC1, MUC2 and MUC5AC Expression in Primary Breast Carcinoma. *J. Clin. Diagnostic Res.: JCDR* **11**, Ec30–Ec34

44 He, Y.F., Zhang, M.Y., Wu, X., Sun, X.J., Xu, T., He, Q.Z. et al. (2013) High MUC2 expression in ovarian cancer is inversely associated with the M1/M2 ratio of tumor-associated macrophages and patient survival time. *PLoS One* **8**, e79769, https://doi.org/10.1371/journal.pone.0079769

45 Awaya, H., Takeshima, Y., Yamasaki, M. and Inai, K. (2004) Expression of MUC1, MUC2, MUC5AC, and MUC6 in atypical adenomatous hyperplasia, bronchioloalveolar carcinoma, adenocarcinoma with mixed subtypes, and mucinous bronchioloalveolar carcinoma of the lung. *Am. J. Clin. Pathol.* **121**, 644–653, https://doi.org/10.1309/U4WGE9EBFJN6CM8R

46 Li, K., Sun, D., Gou, Q., Ke, X., Gong, Y., Zuo, Y. et al. (2018) Long non-coding RNA linc00460 promotes epithelial-mesenchymal transition and cell migration in lung cancer cells. *Cancer Lett.* **420**, 80–90, https://doi.org/10.1016/j.canlet.2018.01.060

47 Liang, Y., Wu, Y., Chen, X., Zhang, S., Wang, K., Guan, X. et al. (2017) A novel long noncoding RNA linc00460 up-regulated by CBP/P300 promotes carcinogenesis in esophageal squamous cell carcinoma. *Biosci. Rep.* **37**, https://doi.org/10.1042/BSR20171019

48 Kong, Y.G., Cui, M., Chen, S.M., Xu, Y., Xu, Y. and Tao, Z.Z. (2018) LncRNA-LINC00460 facilitates nasopharyngeal carcinoma tumorigenesis through sponging miR-149-5p to up-regulate IL6. *Gene* **639**, 77–84, https://doi.org/10.1016/j.gene.2017.10.006

49 Feng, L., Yang, B. and Tang, X.D. (2018) Long noncoding RNA LINC00460 promotes carcinogenesis via sponging miR-613 in papillary thyroid carcinoma. *J. Cell Physiol.* **234**, 11431–11439