Identification of ceRNA network based on a RNA-seq shows prognostic IncRNA biomarkers in human lung adenocarcinoma

XING LI1,2, BING LI1, PIXIN RAN3 and LANYING WANG2

1GMU-GIBH Joint School of Life Sciences, Guangzhou Medical University, Guangzhou, Guangdong 511400; 2Oncology Department, Gansu Provincial Hospital of Traditional Chinese Medicine, Lanzhou, Gansu 730050; 3State Key Laboratory of Respiratory Disease, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong 510030, P.R. China

Received October 21, 2017; Accepted July 5, 2018

DOI: 10.3892/ol.2018.9336

Abstract. Previous studies have emphasized the significant functions of long non-coding RNAs (lncRNAs) as competing endogenous RNAs (ceRNAs) in tumor biology. However, the functions of certain cancer lncRNAs in the IncRNA-related ceRNA network in lung adenocarcinoma (LUAD) are unknown. A systematic and integrative survey of RNA-seq data from The Cancer Genome Atlas (TCGA) was performed to identify candidate IncRNAs for the prognosis of LUAD. In total, 20,502 genes that contain 181 lncRNAs were evaluated in a cohort of 570 LUAD cases. Initially, 6,280 differentially expressed genes (fold-change >2, P<0.05) were obtained using R package, which includes 75 lncRNAs. Next, by univariate regression and multivariate Cox proportional hazards analysis, 32 genes were associated with survival in LUAD. Using these 29 mRNAs and 3 lncRNAs, a prognosis index (PI) was calculated to accurately estimate the survival in LUAD: PI=∑exp_risk gene x HR_risk gene. Furthermore, the 32-gene signature was an independent prognostic indicator for LUAD (HR >1; P<0.05, by multivariate analysis). Weighted gene co-expression network analysis (WGCNA) of three risk lncRNAs-FAM138B, NHEG1 and TLX1NB-was performed, based on the P-values of the associated genes, and the top 27 miRNAs that bound to these lncRNAs were predicted by Miranda as target miRNAs.

Next, these target miRNAs were transferred to the TarBase, miRTarBase, miRecords and starBase v2.0 databases to obtain their target genes. According to the previous miRNA-mRNA and miRNA-IncRNA data, three IncRNA-miRNA-mRNA ceRNA networks were established, based on the 29 prognostic mRNAs, forming a regulatory network in LUAD. The present study provided insight into the IncRNA-related ceRNA network in LUAD and has identified potential diagnostic and prognostic biomarkers.

Introduction

Lung adenocarcinoma (LUAD) is the most well-known histological sub-type of non-small cell lung cancer (NSCLC), which is the most significant cause of cancer-associated mortality worldwide (1). Since LUAD tends to form metastases at an early stage, the prognosis for patients with LUAD is generally poor, with average 5-year survival rates of <20% (2). Although recent advances in molecular pathology and targeted therapies have improved clinical therapies, the 5-year overall survival of patients with LUAD remains low (3), requiring continued research to identify novel biomarkers and effective targeted molecular therapies.

Recent high-throughput transcriptome analyses have demonstrated that >90% of the transcriptome is transcribed into non-coding RNAs, among which microRNAs (miRNAs/miRs) and long non-coding RNAs (lncRNAs) have been revealed to be involved in the malignant behaviors of LUAD. By definition, lncRNAs are non-coding RNAs that are >200 base pairs (4). According to recent studies, lncRNAs are important in various biological processes, including gene expression, cell proliferation and differentiation, the immune response, apoptosis and human diseases, including various types of cancer (5-8).

It has been reported that lncRNAs serve as competing endogenous RNAs (ceRNAs) or miRNA sponges, inhibiting miRNA response elements (MREs) from exchanging with miRNAs by competing for common miRNAs during the tumorigenic process (9). The ceRNA hypothesis has been proposed as a novel regulatory mechanism between non-coding RNAs and coding RNAs (10), thereby contributing toward the occurrence and progression of cancer (11,12). Based on ceRNA theory,
Akacakaya et al (13) demonstrated that miR-133b was significantly decreased in CRC and associated with overall survival and metastasis. Zhang et al (14) analyzed 372 IncRNAs from The Cancer Genome Atlas (TCGA) and NCBI GEO omnibus (GSE65485), and identified a complex cancer-specific ceRNA network in HCC. Zhang et al (15) revealed MALAT1 to be the direct target of miR-202 and reported that interfering with MALAT1 downregulates Gli2 through positive regulation of miR-202 to suppress the development of gastric cancer.

As a novel post-transcriptional regulatory system, ceRNAs provide a novel means of investigating the pathogenesis of cancer and detecting novel signatures with high accuracy in diagnosis. The development and survival of anomalous IncRNAs in different types of cancer (including LUAD) have been examined, revealing the potential of IncRNAs as prognostic cancer biomarkers (16-20). Nevertheless, the study of ceRNAs in LUAD is scarce, and there are few targets for tumor therapies. There are little data on large-scale samples, and microarray detection remains rare, and the association between cancer-specific IncRNAs and clinical features is unknown.

The TCGA database is a public platform that can be used to download the sequencing data of LUAD IncRNAs, miRNAs and mRNAs. Following obtaining RNA sequence data, IncRNAs, miRNAs and mRNAs with abnormal expression were identified, and an IncRNA/mRNA expression network was constructed. These results increase understanding of the molecular mechanisms and therapeutic targets of IncRNAs in the treatment of LUAD.

Materials and methods

TCGA dataset and analysis of the differentially expressed genes. LUAD individual RNA sequencing (RNA-Seq) data (level 3), computed on the Illumina HiSeq RNA-Seq platform, were obtained from the TCGA data portal (https://tcga-data.nci.nih.gov/tcga/), which contains 512 LUAD and 58 adjacent non-tumor lung tissue samples, until December 31, 2016. The raw RNA-Seq reads (IncRNA and mRNA) were post-processed and normalized on the TCGA RNASeqV2 system. The miRNA expression data (level 3) were collected by Illumina GA microRNA sequencing platforms (Illumina, Inc., Hayward, CA, USA), quantile-normalized prior to the analysis and expressed as reads per million (RPM).

The exclusion criteria were as follows: i) The histological diagnosis was not LUAD; ii) presence of a malignancy other than LUAD; and iii) patient samples lacking complete data for analysis.

A total of 493 patients with LUAD were included in the present study. IncRNA expression profiles for normal lung tissue samples were obtained from adjacent non-tumor lung tissues (n=58). In addition, there were 170 LUAD patients with lymph node metastases and 323 LUAD patients without lymph node metastases. According to the Union for International Cancer Control (UICC), there were 381 patients with stage I-II LUAD and 112 patients with stage III-IV LUAD (21).

The expression data of IncRNAs are displayed as RPM and the expression levels of each IncRNA were normalized by Deseq R package for further analysis. The data were obtained from TCGA, a community resource project that provides data for community research. Following low-expression genes being filtered, the LUAD RNA-Seq coverage data contained 20,321 mRNA and 181 IncRNAs compared with the GENCODE database (http://www.gencodegenes.org/). Next, the differentially expressed IncRNAs were counted using EdgeR and the DEseq package of R (Padj <0.05 and absolute log2FC>1). The final IncRNAs were selected using a combination of the two methods. Student's t-test (SPSS 20.0, Inc., IBM Corp., Armonk, NY, USA) was employed to analyze the difference in expression levels of these IncRNAs between LUAD and non-cancerous lung tissues. The present study meets the publication guidelines provided by TCGA (http://cancergenome.nih.gov/publications/publicationguidelines).

Functional enrichment analysis. The differentially expressed genes were analyzed by Gene Ontology (GO) (22) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (23) pathway enrichment using the cluster Profiler package of R. P-values <0.05 and q values <0.05 were considered to indicate statistically significant differences, and were based on similar functional visualization and clustering.

Preparation of the prognostic signature of differentially expressed IncRNAs for LUAD. The screening threshold criteria for differentially expressed genes were as follows: log2FC>1 and FDR<0.05. Additionally, the TCGA also collected clinical pathological features, including survival information. Cases with insufficient clinical data were omitted. Ultimately, the prognostic analysis included 493 samples, with 6,280 genes representing 6,205 mRNAs and 75 IncRNAs. The endpoints of the patients with LUAD in terms of overall survival (OS) were also studied. The mean follow-up period for these patients with LUAD was 28 months.

The prognostic value of the genes and the clinicopathological features were first evaluated by single-variable Cox proportional regression analysis (P<0.05). A multivariate Cox regression model with the survival and KMsurv packs (24) further confirmed the statistical indicators, including IncRNAs and clinicopathological features.

Based on the linear combination of multiple genes in the Cox regression model (β) and the linear combination of regression coefficients, the prognostic risk score was established based on a study undertaken by Tang et al (25). According to the definition of the median prognostic index (PI), patients with LUAD were classified by high and low scores, and the hazard ratio (HR) and 95% confidence interval (CI) were calculated. The time-dependent receiver operating characteristic (ROC) curve within 5 years was also analyzed as the defining point with the R package ‘survival ROC’ to evaluate the predictive accuracy of the prognostic model for the results of time-dependent diseases. Kaplan-Meier survival curves were drawn to assess the association between all parameters (clinical examination and prognostic risk score) and the OS of patients with LUAD (26).

Functional evaluation of IncRNAs with WGCNA. In analyzing the incorporation of differentially expressed IncRNAs into the network, WGCNA can describe the correlation patterns in gene expression profiles (27). WGCNA R package was used to assess the performance of an mRNA/IncRNA and
its module membership, and the correlation between each module and disease state was obtained through single-factor variance analysis. All mRNAs associated with an lncRNA in the module that were significantly associated with the disease were considered potential targets for the lncRNA. The function of an lncRNA was obtained by functional enrichment analysis of the target. Following an evaluation of the weighted correlations, the network feature was presented by Cytoscape 3.4.0. (http://www.cytoscape.org/).

**LncRNA function prediction.** The coding gene in the same WGCNA network module is a potential regulatory target gene for an lncRNA. The core enrichment lncRNA in the WGCNA network module, which is significantly associated with disease, is the risk lncRNA in the present study. GO and KEGG enrichment analyses of the risk lncRNAs and coding genes co-expressed in the same WGCNA module were performed using the R clusterProfiler package to obtain the functional information of disease risk lncRNAs.

**Network analysis of ceRNA.** lncRNAs can bind to miRNAs and affect ceRNA function. MiRanda (http://miranda.org.uk/) is used to predict the interaction between lncRNAs and miRNAs. miRanda is based mainly on the combination of circRNA and miRNA free energy size: The smaller the free energy, the stronger the binding capacity. Following application of filters of a threshold maximum energy ≤‑20 and score >160, numerous microRNAs that bound to lncRNA were predicted, but only microRNAs with the top 10 scores were studied.

Based on the TarBase (http://www.microrna.gr/tarbase), miRTarBase (http://miertarbase.mbc.nctu.edu.tw/php/index.php), miRecords (http://cl.accurascience.com/miRecords/) and starBase v2.0 (http://starbase.sysu.edu.cn/) databases, information on the downstream target genes of microRNAs in the ceRNA network was obtained. Cytoscape software was used to construct microRNA-target and microRNA-lncRNA integration networks that only showed the genes within the module being investigated. The mRNAs in the network were analyzed by GO BP enrichment using the clueGO plugins in Cytoscape software.

**The correlations between LUAD-specific lncRNAs and clinical features.** According to the bioinformatics analysis and the ceRNA network, prognostic lncRNAs were selected. Cox regression analysis was also performed on the results for the clinical factors, including sex, tumor pathological stage, Tumor-Node-Metastasis (TNM) stage (21) and the number of pack-years.

**Statistical analysis.** Data are presented as the mean ± standard deviation and Student's t-tests, two-sided log-rank test, single-factor cox proportional risk regression, multi-factor cox proportional risk regression, R package ‘survivalROC’ and ‘KMsurv’ were used to perform statistical analysis. P<0.05 was considered to indicate a statistically significant difference.

**Results**

**Data source and pre-processing.** The expression data on 20,502 genes were extracted, and R language (EdegR and DESeq) was used to calculate the expression data. The differentially expressed genes (n=6280; Table I) were selected to further investigate their prognostic value (Fig. 1). The differentially expressed genes were counted using EdgeR and the DEseq package of R (Padj <0.05 and absolute log2FC >1). The final genes were selected using a combination of the two methods. Student’s t-test was employed to analyze the difference in expression levels of genes between lung adenocarcinoma and non-cancerous lung tissues.
Table II. 32 genes serving as prognostic indicators.

<table>
<thead>
<tr>
<th>IncRNA</th>
<th>mRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAM138B</td>
<td>CRISP1</td>
</tr>
<tr>
<td></td>
<td>TP63</td>
</tr>
<tr>
<td></td>
<td>MYL10</td>
</tr>
<tr>
<td></td>
<td>CRISP3</td>
</tr>
<tr>
<td></td>
<td>NRG1</td>
</tr>
<tr>
<td></td>
<td>BCO92501</td>
</tr>
<tr>
<td></td>
<td>FAM9C</td>
</tr>
<tr>
<td></td>
<td>MYOC</td>
</tr>
<tr>
<td>TLX1NB</td>
<td>KRT9</td>
</tr>
<tr>
<td></td>
<td>STON1.GTF2A1L</td>
</tr>
<tr>
<td></td>
<td>ADAM30</td>
</tr>
<tr>
<td></td>
<td>SPANX1</td>
</tr>
<tr>
<td></td>
<td>ALDOB</td>
</tr>
<tr>
<td></td>
<td>FAM184B</td>
</tr>
<tr>
<td>NHEG1</td>
<td>PLXN1B3</td>
</tr>
<tr>
<td></td>
<td>BA1AP2L2</td>
</tr>
</tbody>
</table>

Single-factor Cox proportional risk regression and multi-factor Cox proportional risk regression analysis was used to evaluate the prognostic indicators in lung adenocarcinoma.

37.8 months, significantly less than that of the low-risk group (47.4 months; P=0.00409; Fig. 2A). In addition, the present study also assessed the difference in expression levels of 32 genes between the high-risk and low-risk groups. The prognostic risk index aided in predicting the 5-year survival rate of patients with LUAD, with an AUC value of 0.69 (Fig. 2B). The median prognostic risk index of ‘Risk score’ and ‘survival days’ is the separation for low-risk and high-risk patients with LUAD. (Fig. 2C and D).

WGCNA module division and lncRNA target gene prediction. The WGCNA network module was constructed for all differentially expressed mRNAs and lncRNAs, setting the number of genes in each module to be no less than 20. A total of 19 network modules were divided, and the correlation between the module and LUAD was analyzed by one-way analysis of variance. It was revealed that 11 modules were significantly correlated significantly with the disease. The present study focused on the network modules of the lncRNAs FAM138B (downregulated in LUAD), NHEG1 (downregulated in LUAD) and TLX1NB (upregulated in LUAD). FAM138B is located in the turquoise module containing 3,257 genes, 311 of which were co-expressed with FAM138B. We hypothesized that these 311 genes are potential FAM138B target genes. NHEG1 is located in the blue module containing 900 genes, 582 of which were co-expressed with NHEG1. Therefore, the 582 genes are potential NHEG1 target genes. TLX1NB lies in the turquoise module containing 3,257 genes, 2,262 of which are co-expressed with TLX1NB. We believe that these 2262 genes are potential TLX1NB target genes.

Functional analysis of the lncRNA target gene. From the WGCNA analysis of the FAM138B, NHEG1 and TLX1NB lncRNAs, the target genes of each lncRNA were obtained. The target genes of the lncRNAs were analyzed by GO and KEGG enrichment using the cluster Profiler package of R. The top three upregulated pathways of FAM138B and TLX1NB were ABC transporters, cAMP signaling and vascular smooth muscle contraction. The top five upregulated pathways of NHEG1 were cell cycle, DNA replication, homologous recombination, Fanconi anemia and p53 signaling (Fig. 3).

There were 14 prognostic risk genes among the target genes of FAM138B and 10 genes among the target genes of TLX1NB (Table III). The two lncRNAs FAM138B and TLX1NB shared MYOC, PLXN1B3 and ACACB among the eight prognostic risk genes as target genes (Fig. 4).

Construction of lncRNA-miRNA network. Miranda predicted microRNAs that were bound to the three lncRNAs, and 712 miRNAs were obtained at the thresholds. According to the predicted scores, the present study focused on the 27 miRNAs with the 10 highest scores (including duplicate scores) that bound to each lncRNA. The miRNA and lncRNA interaction network, which contained 715 nodes and 815 edges, was constructed (Fig. 5).

miRNA target gene prediction. Based on the TarBase, miRTarBase, miRecords and starBase v2.0 databases, the target genes of the aforementioned 27 key miRNAs were obtained, and the microRNA-mRNA network was constructed. The network contained 1,961 sides, 27 microRNAs and 1,229 mRNAs (816 upregulated and 413 downregulated; Fig. 6). Furthermore, GO BP enrichment analysis of the mRNAs in the network was performed (Fig. 7).

Construction of lncRNA-miRNA-mRNA ceRNA network. The miRNA-mRNA and lncRNA-miRNA pairs were used to build the lncRNA-miRNA-mRNA ceRNA network. The 29 previous prognostic risk mRNAs and their microRNAs were separately extracted, and a regulatory network was established. The network also added lncRNAs, which regulate potential miRNAs, and constructed an important lncRNA-miRNA-mRNA network (Fig. 8). As demonstrated, hsa-miR-619-5p has a regulatory association with three prognostic risk factors. Three upregulated mRNAs and one downregulated mRNA were potentially affected by the lncRNA-ceRNA network.

Correlations between LUAD-specific lncRNAs and clinical features. According to the single-factor and multi-factor Cox proportional risk regression analysis, three specific lncRNAs-FAM138B, NHEG1 and TLX1NB-were selected. Cox regression analysis for clinical factors, including sex, tumor pathological stage, TNM stage and number of pack-years, was then performed. The expression levels of these three lncRNAs were significantly altered by ‘T’ stage and primary tumor size (P<0.05; Table IV).
GEO validation. The prognostic value of these three new lncRNAs was then assessed using GEO data. However, there were not enough survival data available in the GEO datasets. Therefore, the prognostic value of these lncRNA signatures requires further verification using other approaches in a larger number of samples.

Discussion

Lung cancer has the highest incidence and mortality rate of any disease worldwide (30), and lung glandular cell cancer is one of the most common types of LUAD (31). Surgery is considered the most effective method of curing LUAD. However, the majority of patients with LUAD in hospital have missed the opportunity for surgery (32). Chemotherapy and radiotherapy are the typical second-line treatments for LUAD, but their effects are not satisfactory (33). In previous years, the diagnosis of lung cancer, surgical techniques and novel drug treatments have improved, but the 5-year response rate for lung cancer continues to be maintained at 10% (34), mainly because lung cancer cannot be diagnosed in the early stages. Although numerous tests can support lung cancer examination and staging, including tissue biopsy and thoracoscopy, these methods are time-consuming and will delay early diagnosis and treatment (35). Therefore, to improve this situation, more attention is being paid to the occurrence and development of lung cancer-related genes and their regulatory mechanisms.

With the development of high-throughput sequencing technology and bioinformatics, lncRNAs have been revealed to be involved in the development of numerous diseases, particularly tumors (36-39). However, the study of lncRNA expression profiles in LUAD has gradually increased. Using TNM I stage lung adenocarcinoma samples in the TCGA database, Tian et al. (40) demonstrated that FENDRR and LINC00312 had diagnostic value in patients with LUAD. A recent study has demonstrated that lncRNA can act as a ceRNA that adsorbs onto miRNAs to regulate mRNA expression (41). The ceRNA hypothesis is a novel regulatory mechanism that serves a role through miRNA competition (10,42). However, the study of ceRNAs in LUAD is not extensive. Li et al. (43) revealed MIAT, MEG3 and LINC00115 to be the basic therapeutic targets of LUAD as ceRNAs that regulate miRNA-mRNA pairs. Sui et al. (44) demonstrated that AFAP1-AS1 is associated with tumor pathological stage and that LINC00472 is associated with lymph node metastasis. Liu et al. (45) presented lncRNA landscapes and revealed highly altered, differentially expressed lncRNAs in LUSC and LUAD. Two lncRNAs (IGF2BP2-AS1 and DGCR5) were correlated with an improved OS in LUSC, and three (MIR31HG, CDKN2A-AS1 and LINC01600) predicted a poor OS in LUAD.

The public lncRNA data were mined in the latest study (46,47) from TCGA, which identified risk lncRNAs and mRNAs to construct a ceRNA network. The present study first obtained tumor tissue and their corresponding normal tissue.
LI et al: COMPREHENSIVE ANALYSIS OF LUNG ADENOCARCINOMA lncRNA EXPRESSION PROFILES

Table III. IncRNA FAM138B and TLX1NB target gene.

<table>
<thead>
<tr>
<th>LncRNA</th>
<th>Target gene</th>
<th>log2FC</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAM138B</td>
<td>MYOC</td>
<td>-4.357717939</td>
<td>0.0000000294</td>
</tr>
<tr>
<td></td>
<td>PLXNB3</td>
<td>3.878550766</td>
<td>0.0000000191</td>
</tr>
<tr>
<td></td>
<td>ACACB</td>
<td>-1.379164532</td>
<td>0.0000000121</td>
</tr>
<tr>
<td></td>
<td>MCC</td>
<td>-1.27636241</td>
<td>0.0000000488</td>
</tr>
<tr>
<td></td>
<td>RASGRF1</td>
<td>-2.055290402</td>
<td>0.0000000424</td>
</tr>
<tr>
<td></td>
<td>VSIG10L</td>
<td>2.102105505</td>
<td>0.000000327</td>
</tr>
<tr>
<td></td>
<td>SLC16A7</td>
<td>1.372035277</td>
<td>0.0000000011</td>
</tr>
<tr>
<td></td>
<td>NRGN</td>
<td>-1.203111742</td>
<td>0.0000000132</td>
</tr>
<tr>
<td></td>
<td>NRG1</td>
<td>-1.17178188</td>
<td>0.000000029</td>
</tr>
<tr>
<td></td>
<td>TLX1NB</td>
<td>2.102744205</td>
<td>0.0000335</td>
</tr>
<tr>
<td></td>
<td>LOC340074</td>
<td>1.475322606</td>
<td>0.000465275</td>
</tr>
<tr>
<td></td>
<td>STON1-GTF2A1L</td>
<td>-1.249288199</td>
<td>0.000666914</td>
</tr>
<tr>
<td></td>
<td>MYL10</td>
<td>1.376749843</td>
<td>0.02548923</td>
</tr>
<tr>
<td></td>
<td>C10orf27</td>
<td>1.038217387</td>
<td>0.03122004</td>
</tr>
<tr>
<td>TLX1NB</td>
<td>MYOC</td>
<td>-4.35771794</td>
<td>0.00000000272</td>
</tr>
<tr>
<td></td>
<td>PLXNB3</td>
<td>3.878550777</td>
<td>0.00000000943</td>
</tr>
<tr>
<td></td>
<td>ACACB</td>
<td>-1.379164532</td>
<td>0.00000000669</td>
</tr>
<tr>
<td></td>
<td>MCC</td>
<td>-1.27636241</td>
<td>0.00000000432</td>
</tr>
<tr>
<td></td>
<td>RASGRF1</td>
<td>-2.055290402</td>
<td>0.00000000571</td>
</tr>
<tr>
<td></td>
<td>NRGN</td>
<td>-1.203111742</td>
<td>0.0000000444</td>
</tr>
<tr>
<td></td>
<td>KRT9</td>
<td>2.93858028</td>
<td>0.000000755</td>
</tr>
<tr>
<td></td>
<td>FAM138B</td>
<td>-1.11756478</td>
<td>0.00016011</td>
</tr>
<tr>
<td></td>
<td>STON1-GTF2A1L</td>
<td>-1.2492882</td>
<td>0.000392098</td>
</tr>
<tr>
<td></td>
<td>MYL10</td>
<td>1.376749843</td>
<td>0.018236455</td>
</tr>
</tbody>
</table>

lncRNA, long non-coding RNA.

Figure 3. Significantly enriched Kyoto Encyclopedia of Genes and Genomes terms of the risk long non-coding RNA. The point color represents the P.adjust value, and the point size GeneRatio represents the ratio of the number of differential genes in the Gene Ontology item to the total number of differential genes.

Figure 4. lncRNA FAM138B, lncRNA TLX1NB and their target risk genes. lncRNA, long non-coding RNA.

lncRNAs and mRNAs, and subsequently, those RNAs that tended to have prognostic value were selected by univariate and multivariate analysis. Next, 29 mRNAs and three lncRNAs were yielded to construct the PI value (25). Notably, the PI value proved to be an independent prognostic indicator of LUAD. It was revealed that the PI, constructed from 32 differential genes, had good prognostic value in LUAD. The present study used TCGA data that were generated by RNA-Seq and used to assess the survival of a large number of patients with LUAD in 493 cases.
A total of 29 mRNAs and three IncRNAs were involved in LUAD risk genes. The present study investigated the molecular mechanisms of these IncRNAs. However, none of these IncRNAs was identified in reference records, and little was known regarding the functional mechanisms of these three IncRNAs. Therefore, a WGCNA study on IncRNA-associated genes was conducted, introducing the 14 candidates that could be the most relevant genes in LUAD into the network. The present study focused on the three risk IncRNAs, and genes within these modules that had a weighted association. These genes were the IncRNA target genes. Subsequently, the bioinformatics of abnormally expressed mRNAs was studied, and the most highly enriched KEGG pathway annotations could be the functions of IncRNAs. Based on the KEGG pathway analysis, several cancer-associated pathways were identified, including p53 signaling and the cell cycle.

Lu et al (48) assessed the expression and function of the IncRNA SOX21 antisense RNA 1 (sox21-as1) in 68 pairs of LUAD tissues and the corresponding non-tumor tissues, and demonstrated that sox21-as1 is associated with the development of LUAD. Furthermore, the knockdown of sox21-as1 inhibited the proliferation of LUAD cells in vitro and in vivo, inducing cell cycle stagnation and apoptosis. Wu et al (49) integrated two datasets to search for novel target genes and pathways to explain the pathogenicity of LUAD, and discovered that

![Image of IncRNA-miRNA co-expression network](image-url)
PPM1D and GADD45B could modulate the progression of LUAD through p53 signaling.

Furthermore, the present study investigated the target genes of three lncRNAs. Fourteen of the FAM138B target genes were prognostic risk genes, as were 10 TLX1NB target genes - 8 of which were shared. In the absence of calcium, NRGN, which combines calmodulin with synaptic protein kinase substrates, is reduced in monkeys and rats following exposure to welding smoke (50). MCC is a candidate colorectal tumor suppressor gene that is thought to negatively modulate cell cycle progression (51) and encodes guanine nucleotide exchange factor (GEF). RASGRF1 stimulates the dissociation of GDP from RAS. This gene is expressed in various tissues other than the brain, including the lung and pancreas, and several tumor cell lines (52). KRT9 serves an important role in mature palm and plantar skin tissue and in the morphological processes that
form these tissues, and is upregulated (53,54) in platycodin D (PD)-treated HepG2 cells. PLXNB3, a part of the coding family of plexiform proteins, acts as a receptor for signal 5A and serves a role in invasive growth, axon guidance and cell migration. Plexin and B3 serve vital roles in the invasion and metastasis of gastric cancer (55). ACACB may be involved in the regulation of fatty acid oxidation. It inhibits the synthesis of fatty acids and tumor development in non-small cell lung cancer in preclinical models (56).

lncRNAs can function as ceRNAs by binding miRNAs, regulating the expression of genes. LUAD-associated lncRNA-related ceRNA networks provide vital hints in detecting crucial RNAs of ceRNA-mediated gene regulation networks during the initiation and growth of LUAD. There is increasing evidence that lncRNAs are a key part of ceRNA networks by regulating the transcription of other RNAs (5,57‑61). For example, HOTAIR can use miR-331-3p as endogenous lodging to eliminate the inhibitory activity of the miRNA induced by HER2 3'-UTR and increase its post-transcriptional levels (62). Therefore, as potential connections to lncRNA, miRNA and mRNA may be involved in the initiation, as well as the development, of LUAD. The present study predicted a network involving three types of risk lncRNAs and 712 possible target miRNAs. To facilitate the next step, 27 miRNAs with a score of TOP10 were selected. Based on the TarBase, miRTarBase, miRecords and starBase v2.0 databases, 1,229 objective genes of these 27 miRNAs were obtained, and GO enrichment analysis was performed on these target genes. The GO enrichment results included positive regulation of the epithelial-to-mesenchymal transition, regulation of cell proliferation, positive regulation of the cell cycle and cell death, and multicellular organism development.

Figure 7. miRNA-target genes of Gene Ontology Biological Process enrichment results. The results were calculated as-log10 (P-value) (P<0.05). miRNA, microRNA.
Among the 1,229 genes, there were four prognostic risk genes. ALDOB is associated with the GO term phosphorus metabolic process. The main function is to catalyze the reversible transformation of fructose-1,6-diphosphate to dihydroxyacetone and glyceraldehyde 3-phosphate (63). Abnormal changes in ALDOB are associated with a number of diseases, including hereditary fructose intolerance, hepatitis, cirrhosis and cancer (64). Using data mined from the Gene Expression Omnibus for the colorectal cancer transcriptome (GSE35452) database, Tian et al. (65) reported that ALDOB is associated with glycolysis and is the most significantly increased transcript (GO: 0006096). It may serve a significant role in colorectal cancer progression and the response to neoadjuvant concurrent chemoradiotherapy, and as a novel prognostic biomarker. ACACB positively regulates organ growth. It catalyzes the ATP-dependent carboxylation of acetyl-CoA to malonyl-CoA (66). Recent studies have reported that ACACB is increased in a variety of human cancer types, most likely to promote fat generation and meet the requirement for rapid growth and proliferation (67-69). Jones et al. (70) demonstrated that knockout of ACACB changes the lipid content in glioblastoma multiform U87 cells, increases caspase activity and inhibits the growth of fat to generate tumor treatment effects. The F9 gene encodes vitamin K-dependent coagulation factor IX, which circulates in the blood as an inactive zymogen (71). Alterations in this gene cause blood coagulation factor IX deficiency, which is a recessive X-linked disorder, also called hemophilia B or Christmas disease. The presentation of hemophilia B is consistent with easy bruising, urinary tract bleeding and nosebleeds. We hypothesized that tumor-associated mutations in F9 may exacerbate bleeding symptoms due to necrosis of the tumor. CRISP1, also known as astrocyte elevated gene-1 (AEG-1), is an oncogene that is overexpressed in a wide range of cancer types and serves a key role in regulating carcinogenesis (72). AEG-1 serves an important role in regulating hepatocellular carcinoma (HCC). AEG-1-overexpression at the mRNA and protein expression levels has been observed in a high percentage (>90%) of patients with HCC. A progressive increase in the levels of AEG-1 has been observed in patients with HCC, correlating directly with disease stage, a higher rate of recurrence and a poorer overall survival (73,74).

The present study predicted a network consisting of IncRNAs, miRNAs and target genes, and these target genes are also cancer-related genes regulated by IncRNAs through the ceRNA network. The IncRNA FAM138B regulates ALDOB through hsa-miR-378a-3p and regulates F9 through hsa-miR-619-5p; the IncRNA NHEG1 also regulates F9 through hsa-miR-619-5p. The IncRNA TLX1NB regulates CRISP1 through hsa-miR-148b-3p and regulates ACACB through hsa-miR-1273g-3p. Therefore, hsa-miR-619-5p is at the center of three prognostic risk IncRNA ceRNA regulation networks and may be a key miRNA.

Therefore, three IncRNAs may be candidate prognostic biomarkers for LUAD patients with significant clinical significance. In summary, the present study examined IncRNAs in 493 patients with LUAD from TCGA and constructed three ceRNA networks as novel prognostic indicators of LUAD. However, further experiments are required to verify the clinical and biological functions of these three IncRNAs.

Acknowledgements

The authors would like to thank Dr Sean Kim (Yale University) for editing the English text of the first draft of this manuscript.

Funding

The present study was supported by funds from the Gansu Provincial Health Bureau Scientific Research Project (grant no. GSWSKY2018-09).
Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

PR and LW contributed toward the conception and design of the study; BL analyzed and interpreted the data; and LW and XL designed the experiments, drafted, and revised the manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

35. Langbein L, Heid HW, Moll I and Franke WW: Molecular charac-


45. Penhoet E, Rajkumar T and Rutter WJ: Multiple forms of fruc-


51. Swinnen JV, Brusselkens K and Verhoeven G: Increased lipog-

52. Jones JE, Esler WP, Patel R, Lanba A, Vera NB, Pfeifferkorn JA and Vernochet C: Inhibition of Acetyl-CoA carboxylase 1 (ACCI) and 2 (ACCI2) reduces proliferation and De Novo lipogenesis of EGF-R/H 


58. Penhoet E, Rajkumar T and Rutter WJ: Multiple forms of fruc-
