Comprehensive analysis of dysregulated long non-coding RNA – microRNA – mRNA involved in the progression of human lung adenocarcinoma

Dan Yang
China Medical University
Yang He
The first affiliated hospital of China Medical University
Bo Wu
The first affiliated hospital of China Medical University
Yan Deng
China Medical University
Ruxi Liu
The first affiliated hospital of China Medical University
Nan Wang
China Medical University
Tieting Wang
China Medical University
Yannan Luo
China Medical University
Yunda Li
China Medical University
Yang Liu (yangliu@cmu.edu.cn)
China Medical University https://orcid.org/0000-0003-0866-0380

Primary research

Keywords: ceRNA, lung adenocarcinoma, bioinformatics, lncRNA, miRNA, risk score

Posted Date: March 9th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-16347/v1

Abstract

Background: Lung adenocarcinoma (LUAD) is the most common histological subtype of lung cancer worldwide. Until now, the molecular mechanisms underlying LUAD progression have not been fully explained. This study aimed to identify a competing endogenous RNA (ceRNA) network in LUAD.

Methods: Differentially expressed lncRNAs (DELs), miRNAs (DEMs), and mRNAs (DEGs) were identified from The Cancer Genome Atlas (TCGA) database with a $|\log_2 FC| > 1.0$ and a false discovery rate (FDR) < 0.05. Then, these DELs, DEMs, and DEGs were used to construct the initial ceRNA network. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), protein-protein interaction (PPI) network, and survival analyses were performed to analyse these DEGs involved in the ceRNA network. Subsequently, the drug-gene interaction database (DGIdb) was utilized to select candidate LUAD drugs interacting with significant DEGs. Then, lasso-penalized Cox regression and multivariate Cox regression models were used to construct the risk score system. Kaplan-Meier (K-M) survival curves and receiver operating characteristic (ROC) curves were utilized to validate the reliability of the risk score system. Finally, based on the correlations between DELs and DEGs involved in the risk score system, the final ceRNA network was identified. Meanwhile, the GEPIA2 database and immunohistochemical (IHC) results were utilized to validate the expression levels of selected DEGs. GEPIA2 was further used to verify the correlations between DEGs and DELs.

Results: A total of 340 DELs, 29 DEMs, and 218 DEGs were selected to construct the initial ceRNA network. Functional enrichment analyses indicated that 218 DEGs were significantly enriched in the GO terms “nucleoplasm”, “transcription factor complex”, “protein binding”, and “metal ion binding”, whereas these DEGs were associated with the KEGG pathway terms “microRNAs in cancer”, “pathways in cancer”, “cell cycle”, “HTLV-1 infection”, and the “PI3K-Akt signalling pathway”. K-M survival analysis of all differentially expressed genes involved in the ceRNA network identified 24 DELs, 4 DEMs, and 29 DEGs, all of which were significantly correlated with LUAD progression ($P < 0.05$). Furthermore, 15 LUAD drugs interacting with 29 DEGs were selected. After lasso-penalized Cox regression and multivariate Cox regression modelling, 4 DEGs, PRKCE, DLC1, LATS2, and DPY19L1, were incorporated into the risk score system. The area under the curve (AUC) values of the time-dependent ROC curves at 3 years and 5 years were both higher than 0.5. Finally, the correlation coefficients between these 4 DEGs and their corresponding DELs involved in the ceRNA network suggested that there were 2 DEL-DEG pairs, NAV2-AS2 – PRKCE (r = 0.430, $P < 0.001$) and NAV2-AS2 – LATS2 (r = 0.338, $P < 0.001$). Considering the previously constructed ceRNA network, NAV2-AS2 – mir-31 – PRKCE and NAV2-SA2 – mir-31 – LATS2 were identified.

Conclusions: The lncRNA-miRNA-mRNA ceRNA network plays an essential role in LUAD. These results may improve our understanding and provide novel mechanistic insights to explore diagnostics, tumourigenesis, prognosis, and therapeutic drugs for LUAD patients.

Background

Lung cancer remains the most commonly diagnosed cancer and has the highest morbidity among all cancer types worldwide, with approximately 2.1 million new cases and 1.8 million deaths in 2018 [1]. Lung adenocarcinoma (LUAD) is the most common histological subtype of lung cancer. It usually accounts for nearly 40% of cancer cases, and the average 5-year survival rate is less than 20% in most countries [2–5]. Although we have made great progress in early diagnosis and novel therapy, the majority of lung cancer deaths are related to secondary diseases or metastatic progression [6]. Risk factors, such as age, gender, cigarette smoking, and pollution, are closely associated with the initiation of LUAD [7]. Additionally, genetic factors also contribute to the development of LUAD [7]. However, the underlying mechanisms of genetic factors and LUAD are not well understood. Because of the lack of sensitive and specific biomarkers, most patients with LUAD are diagnosed at a very late stage, which results in poor prognosis [8]. Therefore, more efforts need to be invested in the identification and understanding of novel biomarkers and specific targets of LUAD, which are considered the key to the development of early diagnosis, better treatment and better overall prognosis of LUAD.

Recently, the emergence of high-throughput transcriptome analysis has received much attention and has revealed that many types of noncoding RNAs (ncRNAs) without a protein-coding ability; these ncRNAs mainly consist of microRNAs (miRNAs) and long noncoding RNAs (lncRNAs), which are involved in the malignant behaviours of LUAD [9]. MicroRNAs are usually 19–25 nucleotides in length [10]. They perform transcriptional or post-transcriptional regulatory functions through binding to targeted mRNAs and further influence the degradation and translation of mRNAs [11]. Many studies have also considered miRNAs as diagnostic or prognostic biomarkers in cancer treatment [12]. lncRNAs are longer than 200 nucleotides in length[13]. The latest research illustrates that lncRNAs exert multiple functional roles in lung cancer, such as cancer metastasis, recurrence, and poor prognosis [14]. In addition, lncRNAs have been identified to perform different functions in the cell. For example, they regulate targeted gene expression in either cis or trans by recruiting a chromatin-modifying complex [14].

Salmena and his colleagues first proposed competing endogenous RNA (ceRNA, including lncRNAs, miRNAs, and mRNAs) as a tumour-specific regulatory pathway that affects the protein expression levels [15]. The principal mechanism of ceRNA is that lncRNAs can be used as endogenous molecular sponges harbouring miRNA response elements (MREs) to indirectly regulate mRNA expression levels [16]. Furthermore, numerous experiments have validated the ceRNA hypothesis, which is involved in the occurrence, recurrence, metastasis, and poor prognosis of LUAD [17–19]. For example, Cai Y et al. found that the lncRNA HMMR-AS1 can promote proliferation and inhibit apoptosis of human LUAD via the miR-138-SIRT6 axis [19]. Thus, it is essential and necessary to perform in depth integrated and comprehensive analyses to explore the regulatory functions of the lncRNA-miRNA-mRNA ceRNA network in tumourigenesis and prognosis.

The Cancer Genome Atlas (TCGA) dataset publicly provides transcriptome profiles with mRNA, miRNA and lncRNA data spanning 33 cancer types. This platform can contribute to the construction of a ceRNA network for LUAD. In this study, differentially expressed mRNAs (DEGs), lncRNAs (DELs), and miRNAs (DEMs) were first identified by comparing RNA-Seq and miRNA-Seq data from tumour and normal tissues

from TCGA. Subsequently, integrated bioinformatics analyses, including Gene Ontology (GO) term analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis, and protein-protein interaction (PPI) construction based on intersecting mRNAs, were performed. Furthermore, survival analysis was performed, and the differentially expressed genes (including DELs, DEMs, and DEGs) significantly affecting LUAD patient prognosis were then screened. We also constructed a ceRNA network on the basis of all the above genes. Finally, four DEGs were identified after lasso-penalized Cox regression analysis and receiver operating characteristic (ROC) curve analysis. Targeted small molecular drugs for LUAD were also identified. This study might contribute to exploring the mechanisms of LUAD tumourigenesis, prognosis, candidate biomarkers, and therapeutic drugs.

Materials And Methods

Data sources and processing

The RNA-Seq data (including lncRNA and mRNA level 3, Illumina HiSeq RNA-Seq platform), miRNA-Seq data (Illumina HiSeq miRNA-Seq platform), and the corresponding clinical information of LUAD patients were downloaded from TCGA database using The GDC Data Portal (https://portal.gdc.cancer.gov/repository) on February 5, 2020. Inclusion criteria were set as follows: (1) histological diagnosis on LUAD; (2) the disease type only included adenomas and adenocarcinomas. The mRNA and lncRNA expression data included a total of 551 samples consisting of 497 LUAD cases and 54 normal controls. The miRNA expression data included 1056 samples consisting of 966 LUAD cases and 90 normal controls. All three RNA expression profiles were extracted by using Perl and the “R Bioconductor” package in R software (version: 3.6.2). Because this study strictly followed the publication guidelines approved by TCGA (https://cancergenome.nih.gov/publications/publicationguidelines), no ethical issues were involved.

Identification of DEGs, DELs, and DEMs

The “edgeR” package in R software was used to identify crucial DEGs (cDEGs), DELs (cDELs), and DEMs (cDEMs) between LUAD cases and normal controls [20]. The FDR < 0.05 and $|\log_2$fold change ($\log_2$FC)| > 1.0 were set as the cutoff criteria [20]. Then, all the cDEGs, cDELs, and cDEMs meeting the criteria were shown in volcano plots using the “gplots” package in R software. Finally, the “pheatmap” package in R software was used to perform hierarchical clustering analysis for cDEGs, cDELs, and cDEMs, as shown in the heatmaps.

Construction of the ceRNA network

The miRcode database (http://www.mircode.org/) was used to predict the interactions between DELs and DEMs [21]. Then, the miRDB, miRTarBase, and TargetScan databases were applied to identify the targeted mRNAs of DEMs [22–24]. To increase the reliability of the results, only mRNAs found in all three databases were considered candidate mRNAs. Subsequently, these candidate mRNAs intersected with cDEGs to identify the DEGs targeted by the DEMs involved in the ceRNA network. Finally, lncRNA-miRNA-

mRNA ceRNA networks were constructed between DELs and DEMs and between DEMs and DEGs. These networks were then visualized by Cytoscape software (version: 3.7.1).

Functional enrichment analysis and PPI network construction of DEGs in the ceRNA network

The Database for Annotation, Visualization and Integrated Discovery (DAVID; https://david.ncifcrf.gov/; version: 6.8) database was used to perform GO terms and KEGG analyses of these DEGs involved in the above ceRNA network. Upregulated and downregulated genes were submitted to the DAVID online program. The top 10 items of GO categories including biological process (BP), cellular component (CC), and molecular function (MF) and KEGG pathways were then sorted and displayed in the form of bubble maps. These bubble maps were drawn by the “ggplot2” R package based on the P value through R software. In this study, $P < 0.05$ was considered to be statistically significant. Additionally, PPI networks are the networks of protein complexes, which provide valuable clues for understanding the molecular mechanisms underlying carcinoma progression [25]. These networks are always formed by biochemical or electrostatic forces [25]. In this study, we used the online database Search Tool for the Retrieval of Interacting Genes (STRING; https://string-db.org/) to construct PPI networks for the overlapping DEGs with the criteria of interaction score > 0.4 and text mining, experiments, databases, coexpression, neighbourhood, gene fusion, and co-recurrence as the active interaction sources [26].

Survival analysis

Patient survival analyses related to expression of each DEG, DEL, and DEM in the ceRNA network were performed via Kaplan-Meier (K-M) survival curve and log-rank test analyses by using the “survival” package in R software. Patients were classified into a high-expression group and a low-expression group based on the median expression of each gene in LUAD tissues. The overall survival (OS) as the endpoint was compared between the above two groups for each gene among all LUAD patients. Specifically, for the survival analysis of DEGs, Kaplan-Meier Plotter database (http://kmplot.com/analysis/) was utilized with a survival curve to identify the significant DEGs related to the prognosis of LUAD patients. The DEGs validated by either TCGA database or Kaplan-Meier Plotter database were considered statistically significant [27]. Similarly, the significant DEMs were identified either in the TCGA database or in the OncomiR database (http://www.oncomir.org/) [28]. $P < 0.05$ was considered to indicate statistical significance. Finally, genetic alterations in those significant DEGs involved in survival analysis were displayed using the cBio Cancer Genomics Portal (https://www.cbioportal.org/; version: 3.0.2).

Drug-hub gene interaction

The Drug-Gene Interaction Database (DGIdb; http://www.dgidb.org/search_interactions; version: 3.0.2 – sha1 ec916b2) was used to select drugs based on the statistically significant DEGs that served as promising targets. Drugs supported by multiple databases or PubMed references were considered potential drugs. The final promising drug list only involved drugs that were approved by the Food and Drug Administration (FDA).

Construction of the risk score system and ROC curve analysis

The above DEGs related to LUAD patient survival were incorporated into lasso-penalized Cox regression to reduce the number of genes and to remove confounding factors. The penalized maximum likelihood method was applied to a Cox model, and the regression coefficients (β) of the multivariate Cox regression model were predicted on the basis of the best lambda, which was validated by ten-fold cross-validation to minimize the mean cross-validated error [16]. The above regression analyses were conducted with the "glmnet" and "survival" packages in R software. Subsequently, a prognostic risk score system based on significant DEGs (these significant DEGs were determined by lasso-penalized Cox regression) was established. The prognostic index (PI) of OS in LUAD patients was predicted as follows: PI = (β1 × expression level of DEG1) + (β2 × expression level of DEG2) + (β3 × expression level of DEG3) + …+ (βn × expression level of DEGn). We considered the median PI as the cutoff for the following grouping. All the LUAD samples were divided into two groups, namely, the low- and high-risk groups, based on the optimal cutoff value of PI. To perform the survival analysis between these two risk groups, a K-M survival curve was drawn by using the "survival" package in R software. Furthermore, to estimate the area under the ROC curve and further validate the reliability of the risk score system, time-dependent ROC curves within 3 and 5 years were illustrated by using the "survivalROC" package in R software. Finally, the DEG expression levels and the risk scores of LUAD samples with survival time were visualized with a heatmap by using the "pheatmap" package in R software.

Univariate and multivariate Cox regression analyses

Univariate Cox regression analysis was conducted to test whether the clinical characteristics, including age, gender, pathological stage, TNM classification, and risk score, were related to progression in LUAD patients. Only those factors with a $P < 0.01$ were considered candidate prognosis factors. Then, those candidate factors were further analysed by multivariate Cox regression analysis. Both univariate and multivariate Cox regression analyses were conducted by using the "survival" package in R software. $P < 0.05$ in the multivariate Cox regression analysis was considered statistically significant. The forest map was drawn by STATA software (version: 15.0; StataCorp, College Station, Texas, USA).

Correlation analysis between the expression level of DEGs and DELs

Correlation analyses of the expression levels between DEGs involved in the risk score system and their related DELs in the ceRNA network were conducted. These correlation results were visualized by using the "corrplot" package in R software. $P < 0.05$ and $r > 0.30$ were considered statistically significant [16].

Validation of the DEGs

The expression levels of the selected significant DEGs involved in correlation analysis between LUAD cases and normal controls were further validated by Gene Expression Profiling Interactive Analysis

(GEPIA2; http://gepia2.cancer-pku.cn/#index). Then, GEPIA2 was also used to verify the correlations between the selected DEGs and DELs based on TCGA normal and GTEx data. Finally, immunohistochemical (IHC) results of the selected DEGs in LUAD were obtained from the Human Protein Atlas (https://www.proteinatlas.org/; version: 19.2). IHC images in LUAD and normal lung were visualized based on the Human Protein Atlas.

Results

Identification of DEGs, DEMs and DELs in LUAD patients based on TCGA data

The LUAD transcriptome profiling data and corresponding clinical information were downloaded from the TCGA database. By using the “edgeR” package in R with the cutoff value of $|\log_2 FC| > 1$ and FDR < 0.05, a total of 5595 cDEGs were considered in LUAD samples compared with those in normal samples, including 3750 (67.02%) upregulated and 1845 (32.98%) downregulated genes (Fig. 1a). A total of 3942 cDELs were then identified, including 3257 (82.62%) upregulated and 685 (17.38%) downregulated lncRNAs (Fig. 1b). Meanwhile, a total of 265 cDEMs were further screened, including 159 (77.56%) upregulated and 46 (22.44%) downregulated miRNAs (Fig. 1c). Additionally, the expression levels of these above genes are visualized in the form of heatmaps in Fig. 2.

Intersecting lncRNAs and mRNAs and construction of the ceRNA network in LUAD

A flowchart of the ceRNA network construction is shown in Fig. 3. As demonstrated in this flowchart, 34 potential miRNAs that interacted with 340 lncRNAs were first predicted on the basis of the miRcode database (Fig. 3). Then, these 34 miRNAs were applied to the further selection of targeted mRNAs. After using all three databases, including miRDB, miRTarBase, and TargetScan, 5 miRNAs were removed (namely, hsa-mir-508, hsa-mir-489, hsa-mir-301b, hsa-mir-184, and hsa-mir-187), whereas 943 targeted mRNAs remained (Fig. 3). Subsequently, at the intersection of these 943 targeted mRNAs and 5377 cDEGs (only those genes identified by R), 218 overlapping DEGs were identified (Fig. 4). Then, 340 DELs, 29 DEMs, and 218 DEGs were incorporated into a final LUAD ceRNA network (Fig. 5). The respective genes involved in the ceRNA network are presented in Table S1.

Functional enrichment analysis and PPI network construction on DEGs involved in ceRNA

The online tool DAVID database was applied to perform GO and KEGG pathway enrichment analyses. GO analysis of the above 218 DEGs grouped DEGs into three functional groups, namely, BP, CC, and MF. The top 5 significant terms from the GO enrichment analysis showed that in the BP category, the DEGs were involved in transcription, DNA-templated (18.81%, $P < 0.001$), positive regulation of transcription from RNA polymerase II promoter (13.30%, $P < 0.001$), regulation of transcription, DNA-templated (12.84%, $P = 0.029$), signal transduction (10.55%, $P = 0.028$), and negative regulation of transcription from RNA polymerase II promoter (9.63%, $P = 0.001$) (Fig. 6a). For the CC category, the DEGs were correlated with nucleus (40.83%, $P < 0.001$), cytoplasm (34.86%, $P = 0.014$), nucleoplasm (23.85%, $P < 0.001$), transcription factor complex (5.50%, $P < 0.001$), and proteinaceous extracellular matrix (3.67%, $P = 0.035$)

(Fig. 6b). For the MF group, the DEGs were enriched for protein binding (61.00%, $P < 0.001$), metal ion binding (18.81%, $P = 0.002$), DNA binding (15.60%, $P = 0.003$), ATP binding (12.84%, $P = 0.021$), transcription factor activity, and sequence-specific DNA binding (11.93%, $P < 0.001$) (Fig. 6c). For KEGG pathway enrichment, the DEGs were associated with hsa05206: microRNAs in cancer (8.72%, $P < 0.001$), hsa05200: pathways in cancer (8.26%, $P < 0.001$), hsa04110: cell cycle (5.96%, $P < 0.001$), hsa05166: HTLV-1 infection (5.96%, $P < 0.001$), and hsa04151: PI3K-Akt signaling pathway (5.05%, $P = 0.041$) (Fig. 6d). Furthermore, the specific information on the DEGs identified in each category in the functional enrichment analysis is presented in Table S2.

The PPI network of the DEGs involved in the ceRNA network was constructed on the basis of the information from the STRING database. A total of 218 DEGs were mapped to the PPI network (Fig. S1). 218 nodes and 455 edges were included in this PPI network, and its PPI enrichment P value was lower than 1.0e – 16 (Fig. S1). In addition, the average local clustering coefficient was 0.38 (Fig. S1).

Survival analysis of ceRNA network-associated genes

To validate whether the potential DEGs, DELs, and DEMs were significantly associated with the progression of patients with LUAD, K-M survival analyses and log-rank tests were conducted for each gene involved in the above ceRNA network. After validation, 29 DEGs, 24 DELs, and 4 DEMs were considered oncogenes because all of them had statistical significance in the survival analysis of LUAD patients ($P < 0.05$; Table 1). K-M survival curves of the partial DEGs, DELs, and DEMs (DEMs: PRKCE, LATS2; DELs: NAV2-AS2; DEMs: hsa-mir-375) are shown in Fig. 7. The survival analyses of PRKCE and LATS2 were performed through K-M plotter database (Fig. 7a, b), whereas the survival results of NAV2-AS2 and hsa-mir-375 were based on TCGA database (Fig. 7c, d). The other DEGs, DELs, and DEMs are presented in Figs. S2-S3. Specifically, except hsa-mir-375, the other DEMs including hsa-mir-200a, hsa-mir-21. and hsa-mir-31 were performed on the OncomiR dataset (Table S3). In addition, we further constructed a ceRNA network based on the DEGs, DELs, and DEMs that were significantly related to the progression of LUAD patients (Fig. 8).

Genetic information and drug-gene interactions

Subsequently, cBioPortal was used to determine the genetic alterations of the 29 DEGs (Fig. 9). As presented in Fig. 9a, these queried genes were altered in 300 (59%) queried patients or samples. PTPRD was mutated most often (24%) (Fig. 9a). These mutations included missense mutations, truncating mutations, fusions, amplifications, and deep deletions (Fig. 9a). Among the different kinds of alterations, multiple alterations accounted for the highest percentages, followed by mutation, deep deletion, and amplification (Fig. 9b).

Furthermore, a total of 15 potential drugs approved by the FDA for treating LUAD patients were screened when drug-gene interactions were conducted (Table 2). In this study, the DEGs, including UBASH3B, ZEB1, SELE, PRKCE, and TBXA2R, were selected as the potential targets of the 15 drugs on the basis of the significant results of the above survival analysis (Table 2). Most potential drugs might interact with

TBXA2R (10/15) as agonists (ABACAVIR, ILOPROST, DINOPROSTONE, and DINOPROST), antagonists (FELBINAC and ACETAMINOPHEN), or in some unknown manners (CYCLOSPORINE, MORPHINE, FUROSEMIDE, and VINBLASTINE) (Table 2). In addition, protein kinase C epsilon (PRKCE) is always considered an activator that interacts with INGENOL MEBUTATE and MEPROBAMATE (Table 2).

Construction of the mRNA-associated risk score system

Lasso-penalized Cox regression and multivariate Cox regression analyses were conducted to screen the potential prognosis-related mRNAs based on the 29 DEGs that were significantly associated with OS in the above survival analysis. We also used relative coefficients to weigh the contributions of these DEGs (Figs. 10a, b). Then, according to the minimum of the mean cross-validated error, only 6 DEGs were left for the follow-up analysis, namely, PRKCE, DLC1, LATS2, RALGPS2, ZNF367, and DPY19L1 (Figs. 10a, b). Next, these 6 DEGs were further incorporated into the multivariate Cox regression model using “both” directions, and the mRNA-associated risk score system was constructed as follows: PI = (-0.32648 × expression level of PRKCE) + (-0.12901 × expression level of DLC1) + (0.43411 × expression level of LATS2) + (0.20825 × expression level of DPY19L1). Among these 4 DEGs, PRKCE and DLC1 had negative coefficients in the multivariate Cox regression analysis, whereas LATS2 and DPY19L1 had positive coefficients. Specifically, the median risk score in LUAD patients was 0.97315. After we classified the LUAD samples into two groups (namely, low- and high-risk groups) with the median PI as the cutoff value, the survival curve and log-rank test were performed (Fig. 10c). As presented in Fig. 10c, for LUAD patients in the low-risk group, the 3-year survival rate was 66.5% (95% CI: 57.94% - 76.20%), whereas the 3-year survival rate of the high-risk group was 46.83% (95% CI: 38.27% - 57.30%). Furthermore, the 5-year survival rates of the low- and high-risk groups were 45.2% (95% CI: 34.81% - 58.60%) and 22.05% (95% CI: 14.11% - 34.40%), respectively (Fig. 10c). Additionally, the area under the curve (AUC) values of the time-dependent ROC curves at 3 years and 5 years were both higher than 0.5 (3 years: AUC = 0.64; 5 years: AUC = 0.664; Figs. 10d & S4), which means that the PI constructed by the 4 mRNAs had a good prognostic ability. Finally, the DEG expression levels and the risk scores of LUAD samples with survival time were visualized through a heatmap (Fig. 10e).

Univariate and multivariate analyses

We subsequently used univariate and multivariate Cox regression analyses to select the potential factors related to OS from 316 LUAD patients with clinical information based on TCGA database. The univariate results indicated that the pathological stage, tumour (T), and lymph node (N) classification were significantly associated with the prognosis of LUAD, similar to the risk score results (Fig. 11). For the multivariate results, the tumour and lymph node classifications were no longer related to the progression of LUAD, whereas the risk score system and pathological stage were still significantly correlated with the survival time of LUAD patients (Fig. 11). Compared to the low-risk patients, the high-risk patients had a hazard ratio (HR) of 1.74 (95% CI 1.17 – 2.59, P = 0.006; Fig. 11).

Correlations between DEGs and DELs

After the construction of the risk score system, 4 significant DEGs were identified, namely, PRKCE, LATS2, DPY19L1, and DLC1. DEGs are always positively regulated by DELs by directly interacting with DEMs. To verify this hypothesis in LUAD patients, a correlation analysis between these 4 significant DEGs and their corresponding DELs involved in the ceRNA network (Fig. 8) was performed. As shown in Fig. 12, we identified 2 DEL-DEG pairs, namely, NAV2-AS2 – PRKCE (r = 0.430, $P < 0.001$; Fig. 12a), and NAV2-AS2 – LATS2 (r = 0.338, $P < 0.001$; Fig. 12b). In addition, the correlations between the above 2 DEL and DEG pairs were further verified by the online database GEPIA2 (Fig. S5). These results were consistent with Fig. 12. Meanwhile, the IHC results of LATS2 (Fig. 13) and PRKCE (Fig. S6). As shown in Fig. 13, it is easily to see that most malignant cells presented moderate cytoplasmic and membranous positivity of LATS2 in LUAD. However, PRKCE was not detected in both LUAD tissues and normal lung tissues (Fig. S6).

Considering the ceRNA network shown in Fig. 8, hsa-mir-31 was selected as the key gene. Finally, in this study, two ceRNA networks were selected, namely, NAV2-AS2 – mir-31 – PRKCE and NAV2-SA2 – mir-31 – LATS2.

Discussion

Although there has been great progress in surgical and medical therapy for lung cancer, the latest global cancer statistics (2018) reported that lung cancer is still the most frequently diagnosed cancer, and its mortality ranks first globally [1]. As LUAD is the most common histological subtype of lung cancer, this low 5-year survival rate is far from satisfactory. LUAD is often diagnosed at an advanced or metastatic stage, which makes early detection and treatment impossible. However, cancer mortality can be vastly reduced if cases are detected or treated early [29]. Therefore, it is urgent and necessary to seek innovative biomarkers and precise molecular mechanisms for the early diagnosis, treatment, and prognosis of LUAD.

In recent years, due to the applications of RNA-Seq data and microarray-based expression profiling data, the ceRNA hypothesis proposes a new RNA interaction mechanism, which further contribute to the better understanding of the tumourigenesis and prognosis of LUAD at the molecular level [20]. In our study, DEGs, DELs, and DEMs were first identified by using the raw sequencing data of LUAD cases and normal controls from the TCGA database. Then, the DEL-DEM-DEG ceRNA network was constructed on the basis of DEL-DEM interactions and DEM-DEG interactions. This initial ceRNA network consisted of 340 DELs, 29 DEMs, and 218 DEGs. Subsequently, GO and KEGG pathway analyses of those 218 DEGs were conducted by DAVID. The GO analysis of DEGs indicated that the DEGs were significantly enriched in “nucleoplasm”, “transcription factor complex”, “protein binding”, and “metal ion binding”, which further suggested that LUAD might be considered a metabolism-related disease. The KEGG pathway results showed that the DEGs involved in the ceRNA network were associated with microRNAs in cancer, pathways in cancer, cell cycle, HTLV-1 infection, and the PI3K-Akt signalling pathway. These results also provide significant clues to explore the molecular mechanisms of tumourigenesis and prognosis in LUAD patients. Indeed, many studies have illustrated that the cell cycle, HTLV-1 infection, and PI3K-Akt

signalling pathway are highly related to various types of cancer, especially lung cancer [30–33]. For example, Weimiao Li et al. proposed that overexpression of cell cycle-related proteins in tumours is always related to tumour proliferation behaviours and poor prognosis in non-small-cell lung cancer (NSCLC) [30]. Furthermore, Hiromitsu M et al. in 1990 validated that human T-cell leukaemia virus type 1 (HTLV-1) is associated with small cell lung cancer [31], while no studies have explored the relationship between HTLV-1 infection and LUAD. In addition, the PI3K-Akt signalling pathway regulates the normal physiological activities of cells. Some evidence has also validated that aberrant activation of the PI3K-Akt signalling pathway always leads to tumourigenesis and metastasis in many types of cancer, such as LUAD [32], gastric carcinoma [34], and bladder carcinoma [35]. For LUAD, the PI3K-Akt pathway may affect cell apoptosis and proliferation [36]. Additionally, the PPI network in this study consisted of 218 nodes and 455 edges. This provides useful information on the PPIs involved in our ceRNA network. All of the above evidence suggests that our ceRNA network might play essential roles in exploring the mechanisms of LUAD.

After conducting these functional analyses and PPI construction, K-M survival analysis of all differentially expressed genes involved in the ceRNA network screened 24 DELs, 4 DEMs, and 29 DEGs, all of which were significantly correlated with progression in LUAD patients ($P < 0.05$). Then, a new ceRNA network was constructed based on the significant genes in the survival analysis. Moreover, for the residual 29 DEGs, we used cBioPorta to explore the alteration information and the drug-gene interactions. In this study, we found that multiple alterations, deep deletions, and amplifications occurred frequently for those 29 DEGs in LUAD patients. It is well known that the tumourigenesis and prognosis of LUAD is mostly the consequence of multiple and cooperative genomic alterations [37]. For example, the famous tumour suppressor TP53 is often deactivated and deleted in the majority of LUAD patients [37, 38]. In this study, the large tumour suppressor 2 (LATS2) gene suffered from deep deletion in LUAD patients (Fig. 9a). Furthermore, from the results of K-M survival analysis, a low expression level of LATS2 was associated with poor survival time (Fig. 7b), which was consistent with the study of Jang SH et al [39]. In addition, 15 drugs interacting with 29 DEGs against LUAD were further selected (Table 2). From this result, we found that TBXA2R interacted mostly with many types of drugs approved by the FDA as agonists or antagonists. However, none of these drugs have been reported to be related to the treatment of lung cancer through interactions with TBXA2R. Nevertheless, there is much evidence indicating that iloprost [40], acetaminophen [41], morphine [42], furosemide [43], and vinblastine [44] are broadly applied in chemotherapy for lung cancer patients. Nevertheless, this study still provides novel insights into LUAD pathogenesis and treatment. Moreover, the selected conventional drugs might find potentially innovative use in the future.

In addition, lasso-penalized Cox regression and multivariate Cox regression analyses were combined to assist with the construction of the risk score system. Among the remaining 4 DEGs, namely, PRKCE, DLC1, LATS2, and DPY19L1, three of them have been studied before. For example, deleted in liver cancer-1 (DLC1) is a tumour suppressor gene that has been reported to be involved in the genetic and epigenetic mechanisms of various human cancers, such as lung, colorectum, breast, and prostate carcinomas [45–

47]. Xiaolan Qian et al. indicated that inactivation of DLC1 and downregulation of $p15^{INK4b}$ and $p16^{Ink4a}$ were combined to cause neoplastic transformation and poor prognosis in human cancer [48]. The DLC1 gene encodes a Rho GTPase-activating protein (RhoGAP), and it is also an important tumour suppressor gene in lung cancer [49]. Loss or downregulation of DLC1 expression always causes abnormal functions of Rho GTPases [49]. In this study, we also found that deep deletion frequently occurred in DLC1 (Fig. 9a), and the survival analysis further validated that LUAD patients with lower expression levels of DLC1 were often related to poorer OS (P = 0.022) (Fig. S2). Specifically, we first identified DPY19L1 as the targeted mRNA that was associated with the tumourigenesis and prognosis of LUAD patients in this study. Furthermore, to validate the reliability of our risk score system, survival analysis and ROC curve analysis based on risk scores were conducted. Because the AUC values of the time-dependent ROC curves at 3 years and 5 years were both higher than 0.5, we considered our risk score system to be effective. After validation by univariate and multivariate Cox regression analyses based on the full clinical information, the risk score system derived from the expression levels of the 4 DEGs and the pathological stages might be treated as the only independent prognostic factors of OS in LUAD patients.

Then, based on the correlation results among the remaining 4 DEGs and their corresponding DELs, only two ceRNA networks were selected, namely, NAV2-AS2 – mir-31 – PRKCE and NAV2-SA2 – mir-31 – LATS2. Chang SH and Wang LH once suggested that miRNAs could cause mRNA degradation by binding to the 3'-untranslated region (3'-UTR) of the target genes [50]. Surprisingly, in this study, the expression level of hsa-mir-31 was upregulated in LUAD patients, whereas the expression levels of PRKCE, LATS2, and NAV2-AS2 were downregulated (Fig. 8). In addition, the online database GEPIA2 was also utilized to further validate the expression level of PRKCE (Fig. S7a) and LATS2 (Fig. S7b) between LUAD and normal tissues based on the TCGA normal and GTEx data. The results from GEPIA2 were consistent with the results of Fig. 8. Furthermore, on the basis of the results of survival analysis, lower expression levels of PRKCE, LATS2, and NAV2-AS2 were always correlated with poorer OS in LUAD patients (Fig. 7). All this evidence in our study improved the reliability of the ceRNA network.

In the above two ceRNA networks, NAV2-AS2, an upstream lncRNA, has never been studied; in contrast, the downstream mRNAs PRKCE and LATS2 related to LUAD have been studied extensively [51–53]. For example, PRKCE, a phorbol ester receptor, has been validated to be associated with various types of cell functions, such as cell cycle progression [54], ion channel control [55], cytokinesis [56], and regulation of transcription factor activity [57]. The onset and progression of various types of chronic diseases have been indicated to be associated with PRKCE. These diseases include heart failure, obesity, diabetes, neurological diseases, and cancer [58]. Regarding the mechanisms of NSCLC, Li Ding et al. indicated that PRKCE could keep lung cancer cells from undergoing apoptosis and further promote cell survival through the dysregulation of the mitochondrial caspase pathway [51]. Furthermore, Junjie Wu et al. demonstrated that PRKCE regulated cell proliferation and was targeted by hsa-mir-129 [59]. In the above section, we explored the alterations in LATS2. LATS2 is a presumed tumour suppressor gene that encodes a serine or threonine kinase [60]. Dysregulation of the LATS2 functions has been validated to be related to a series of malignancies, including lung cancer [39], breast cancer [61], prostate cancer [62], and malignant

mesothelioma [63]. There are multiple mechanisms of action for LATS2 in various kinds of cancers, such as the regulation of the cell cycle by controlling G1/S and G2/M transition [64], induction of apoptosis by downregulating Bcl-2 and Bcl-X_2 [65], and conservation of genetic stability by interacting with p53 [66]. Furthermore, Susan Y. Luo et al. suggested that LATS2 might promote tumour growth via different signalling pathways, especially in EGFR mutant and wild-type LUAD [53]. As the centre of the ceRNA network, hsa-mir-31 has been demonstrated to be a new diagnostic microRNA classifier for lung squamous cell carcinoma (LUSC) [67]. In that research, Xiaogang Tan et al. also explored the functions between hsa-mir-31 and its potential targeted gene LATS2 in LUSC patients. However, he found that expression of hsa-mir-31 in the SK-MES-1 cell line did not regulate the activity of LATS2 [67]. Nevertheless, more studies need to be conducted to explore the associations between hsa-mir-31 and the tumourigenesis and prognosis of LUAD.

The strength of our study is that we first the constructed NAV2-AS2 – mir-31 – PRKCE and NAV2-SA2 – mir-31 – LATS2 ceRNA networks based on a risk score system and explored their potential biological functions in the tumourigenesis and prognosis of LUAD patients. Considering the crucial roles of these genes, further studies might be focused on validating and testing the predicted ceRNA network and exploring their precise mechanisms of LUAD. Indeed, our study has several limitations. One is that all the data used in this study were derived from public databases and were not generated by the authors of this article. The clinical information did not include the family exposure history and other environmental factors, which will bias the results of Cox regression analysis. The other is that future experimental studies and verifications must be conducted to explore the mechanisms in depth. Despite these limitations, this study may provide novel insights into the molecular mechanisms of LUAD and assist with exploring potential and innovative targets and biomarkers of tumourigenesis, prognosis, diagnosis, and therapeutic drugs for LUAD patients.

Conclusions

In summary, two novel ceRNA networks, namely, NAV2-AS2 – mir-31 – PRKCE and NAV2-SA2 – mir-31 – LATS2, were first identified. All these results not only provide a comprehensive analysis network but also narrow the scope of research and enhance the prediction accuracy for ceRNA networks. Our findings will improve our understanding of these candidate biomarkers for the diagnosis, tumourigenesis, prognosis, and therapeutic drugs of LUAD patients.

Declarations

Acknowledgements

We thank the TCGA, DAVID, KEGG, STRING, Kaplan-Meier Plotter, DGIdb, cBio Cancer Genomics Portal, and OncomiR databases for providing their platforms and contributors for their valuable data sets.

Authors' contributions

DY, YH, BW, RXL collected and analyzed the data. YL designed and supervised the study. DY, YH, and YD drafted the first version manuscript. DY, YH, NW, and YL reviewed and revised the manuscript. DY, YDL, TTW, and YNL made the diagrams and tables of the article. All authors read and approved the final manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials

The authors declare that the data supporting the findings of this study are available within the article.

Ethics approved and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Abbreviations

LUAD: lung adenocarcinoma; miRNA: mircoRNA; lncRNA: long non-coding RNA; DELs: differentially expressed lncRNAs; DEMs: differentially expressed miRNAs; DEGs: differentially expressed mRNAs; TCGA: The Cancer Genome Atlas; GO: gene ontology; KEGG: Kyoto encyclopaedia of genes and genomes; PPI: protein-protein interaction; ceRNA: competing endogenous RNA; DAVID: database for annotation, visualization and integrated discovery; K-M: Kaplan-Meier; ROC: receiver operating curve; AUC: area under the curve; MREs: miRNA response elements; FDR: false discovery rate; BP: biological process; CC: cellular component; MF: molecular function; STRING: search tool for the retrieval of interacting genes; OS: overall survival; DGIdb: drug gene interaction database; FDA: food and drug administration; PI: prognostic index; HTLV-1: human T-cell leukemia virus type 1; LATS2: large tumor suppressor; DLC1: deleted in liver cancer-1; RhoGAP: Rho GTPase-activating protein; 3’-UTR: 3’-untranslated region; PRKCE: protein kinase C epsilon; NSCLC: non-small-cell lung cancer; LUSC: lung squamous cell carcinoma.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries**. *CA*

Cancer J Clin 2018, **68**(6):394-424.

2. Maemura K, Watanabe K, Ando T, Hiyama N, Sakatani T, Amano Y, Kage H, Nakajima J, Yatomi Y, Nagase T *et al*: **Altered editing level of microRNAs is a potential biomarker in lung adenocarcinoma**. *Cancer Sci* 2018, **109**(10):3326-3335.
3. Saito M, Shiraishi K, Kunitoh H, Takenoshita S, Yokota J, Kohno T: **Gene aberrations for precision medicine against lung adenocarcinoma**. *Cancer Sci* 2016, **107**(6):713-720.
4. Siegel R.L. MKD: **Cancer statistics**. *CA Cancer J Clin* 2019, **69**:7-34.
5. Alam H, LI, N., Dhar, S.S., Wu, S.J., Lv, J., etal: **HP1gamma Promotes Lung Adenocarcinoma by Downregulating the Transcription-Repressive Regulators NCOR2 and ZBTB7A**. *Cancer Res* 2018, **78**:3834-3848.
6. Li JD, FQ, Qi Y, Cui G, Zhao S: **PPARGC1A is upregulated and facilitates lung cancer metastasis**. *Exp Cell Res* 2017, **359**:356-360.
7. de Groot P, Munden RF: **Lung cancer epidemiology, risk factors, and prevention**. *Radiol Clin North Am* 2012, **50**(5):863-876.
8. Guo T, Ma H, Zhou Y: **Bioinformatics analysis of microarray data to identify the candidate biomarkers of lung adenocarcinoma**. *PeerJ* 2019, **7**:e7313.
9. Chen J, WR, Zhang K, Chen LB.: **Long non-coding RNAs in non-small cell lung cancer as biomarkers and therapeutic targets**. *J Cell Mol Med* 2014, **12**:2425-2436.
10. V A: **The functions of animal microRNAs**. *Nature* 2004, **431**:350-355.
11. Kong YW CI, de Moor CH, Hill K, Garside PG, Hamilton TL, et al.: **The mechanism of micro-RNA-mediated translation repression is determined by the promoter of the target gene**. *Proc Natl Acad Sci USA* 2008, **105**:8866-8871.
12. Calin GA CC: **MicroRNA signatures in human cancer**. *Nat Rev Cancer* 2006, **6**:857-866.
13. Jathar S KV, Srivastava J, Tripathi V.: **Technological developments in lncRNA biology**. *AdvExpMedBiol* 2017, **1008**:283-323.
14. Xie W, YS, Sun Z, Li Y: **Long noncoding and circular RNAs in lung cancer: advances and perspectives**. *Epigenomics* 2016, **8**:1275-1287.
15. Salmena L PL, Tay Y, Kats L, Pandolfi P.P.: **A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?***Cell* 2011, **146**:353-358.
16. Bai Y, Long J, Liu Z, Lin J, Huang H, Wang D, Yang X, Miao F, Mao Y, Sang X *et al*: **Comprehensive analysis of a ceRNA network reveals potential prognostic cytoplasmic lncRNAs involved in HCC progression**. *J Cell Physiol* 2019, **234**(10):18837-18848.
17. Cong Z, Diao Y, Xu Y, Li X, Jiang Z, Shao C, Ji S, Shen Y, De W, Qiang Y: **Long non-coding RNA linc00665 promotes lung adenocarcinoma progression and functions as ceRNA to regulate AKR1B10-ERK signaling by sponging miR-98**. *Cell Death Dis* 2019, **10**(2):84.
18. Lu QC, Rui ZH, Guo ZL, Xie W, Shan S, Ren T: **LncRNA-DANCR contributes to lung adenocarcinoma progression by sponging miR-496 to modulate mTOR expression**. *J Cell Mol Med* 2018, **22**(3):1527-

1537.

19. Cai Y, SZ, Chen Y, Wang J: **LncRNA HMMR-AS1 promotes proliferation and metastasis of lung adenocarcinoma by regulating MiR-138/sirt6 axis**. *Aging (Albany NY)* 2019, **11(10)**:3041-3054.
20. Li L, Peng M, Xue W, Fan Z, Wang T, Lian J, Zhai Y, Lian W, Qin D, Zhao J: **Integrated analysis of dysregulated long non-coding RNAs/microRNAs/mRNAs in metastasis of lung adenocarcinoma**. *J Transl Med* 2018, **16**(1):372.
21. Jeggari A MD, Larsson E.: **miRcode: a map of putative microRNA target sites in the long non-coding transcriptome**. *Bioinformatics* 2012, **28(15)**:2062-2063.
22. Agarwal V BG, Nam JW, Bartel DP.: **Predicting effective microRNA target sites in mammalian mRNAs**. *eLife* 2015, **4**:e05005.
23. Chou CH, CN, Shrestha S, et al: **miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database**. *Nucleic Acids Res* 2016, **44(D1)**:D239-D247.
24. Wong N, WX: **miRDB: an online resource for microRNA target prediction and functional annotations**. *Nucleic Acids Res* 2015, **43 (Database issue)**:D146-D152.
25. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C *et al*: **STRING v9.1: protein-protein interaction networks, with increased coverage and integration**. *Nucleic Acids Res* 2013, **41**(Database issue):D808-815.
26. Huang H, Huang Q, Tang T, Zhou X, Gu L, Lu X, Liu F: **Differentially Expressed Gene Screening, Biological Function Enrichment, and Correlation with Prognosis in Non-Small Cell Lung Cancer**. *Med Sci Monit* 2019, **25**:4333-4341.
27. Gyorffy B SP, Budczies J, Lanczky A: **Online survival analysis software to access the prognostic value of biomarkers using transcriptome data in non-small-cell lung cancer**. *PLoS One* 2013, **8(12)**:e82241.
28. Wong N, CY, Chen S, Wang X: **OncomiR: an online resource for exploring pan-cancer microRNA dysregulation**. *Bioinformatics* 2017, **34(4)**:713-715.
29. WHO: **Health Topics: Cancer.**https://wwwwhoint/news-room/fact-sheets/detail/cancer February 16, 2020.
30. Li W, Zheng G, Xia J, Yang G, Sun J, Wang X, Wen M, Sun Y, Zhang Z, Jin F: **Cell cycle-related and expression-elevated protein in tumor overexpression is associated with proliferation behaviors and poor prognosis in non-small-cell lung cancer**. *Cancer Sci* 2018, **109**(4):1012-1023.
31. Hiromitsu Matsuzaki YK, Hiroyuki Hata, Takeshi Yoshinaga, Etsuo Kinuwaki, et al: **Human T-cell leukemia virus type 1 associated with small cell lung cancer**. *Cancer* 1990, **66(8)**:1763-1768.
32. Fumarola C, Bonelli MA, Petronini PG, Alfieri RR: **Targeting PI3K/AKT/mTOR pathway in non small cell lung cancer**. *Biochem Pharmacol* 2014, **90**(3):197-207.
33. Xuan YW, LM, Zhai WL, Peng LJ, Tang Y: **MicroRNA-381 inhibits lung adenocarcinoma cell biological progression by directly targeting LMO3 through regulation of the PI3K/Akt signaling pathway and epithelial-to-mesenchymal transition**. *Eur Rev Med Pharmacol Sci* 2019, **23(19)**:8411-8421.

34. Singh SS, Yap WN, Arfuso F, Kar S, Wang C, Cai W ea: **Targeting the PI3K/Akt signaling pathway in gastric carcinoma: A reality for personalized medicine?** *World J Gastroenterol* 2015, **21(43)**:12261-12273.
35. Sun M, Chen Z, Tan S, Liu C, W Z: **Knockdown of fibrous sheath interacting protein 1 expression reduces bladder urothelial carcinoma cell proliferation and induces apoptosis via inhibition of the PI3K/AKT pathway**. *Onco Targets Ther* 2018, **11**:1961-1971.
36. Liu Y, Yang H, Chen T, Luo Y, Xu Z, Li Y, Yang J: **Silencing of Receptor Tyrosine Kinase ROR1 Inhibits Tumor-Cell Proliferation via PI3K/AKT/mTOR Signaling Pathway in Lung Adenocarcinoma**. *PLoS One* 2015, **10**(5):e0127092.
37. Cancer Genome Atlas Research N: **Comprehensive molecular profiling of lung adenocarcinoma**. *Nature* 2014, **511**(7511):543-550.
38. Consortium APG: **AACR Project GENIE: Powering Precision Medicine through an International Consortium**. *Cancer Discov* 2017, **7**(8):818-831.
39. Jang SH, Oh MH, Cho H, Lee JH, Lee H, Ahn H, H. K: **Low LATS2 expression is associated with poor prognosis in non-small cell lung carcinoma**. *Pol J Pathol* 2019, **70(3)**:189-197.
40. Pehlivan Y, Turkbeyler IH, Balakan O, Sevinc A, Yilmaz M, Bakir K, Onat AM: **Possible anti-metastatic effect of iloprost in a patient with systemic sclerosis with lung cancer: a case study**. *Rheumatol Int* 2012, **32**(5):1437-1441.
41. Gai C, Yu M, Li Z, Wang Y, Ding D, Zheng J, al e: **Acetaminophen sensitizing erastin-induced ferroptosis via modulation of Nrf2/heme oxygenase-1 signaling pathway in non-small-cell lung cancer**. *J Cell Physiol* 2020, **235(4)**:3329-3339.
42. Minchom A, Punwani R, Filshie J, Bhosle J, Nimako K, J M: **A randomised study comparing the effectiveness of acupuncture or morphine versus the combination for the relief of dyspnoea in patients with advanced non-small cell lung cancer and mesothelioma**. *Eur J Cancer* 2016, **61**:102-110.
43. Makimoto G, Ichihara E, Hotta K, Ninomiya K, Oze I, Minami D, al e: **Randomized Phase II Study Comparing Mannitol with Furosemide for the Prevention of Renal Toxicity Induced by Cisplatin-based Chemotherapy with Short-term Low-volume Hydration in Advanced Non-small Cell Lung Cancer: The OLCSG1406 Study Protocol**. *Acta Med Okayama* 2018, **72(3)**:319-323.
44. Blumenreich MS, Woodcock TM, Gentile PS, Barnes GR, Jose B, Sherrill EJ, al e: **High-dose cisplatin and vinblastine infusion with or without radiation therapy in patients with advanced non-small-cell lung cancer**. *J Clin Oncol* 1987, **5(11)**:1725-1730.
45. Kim TY, Vigil D, Der CJ, Juliano RL: **Role of DLC-1, a tumor suppressor protein with RhoGAP activity, in regulation of the cytoskeleton and cell motility**. *Cancer Metastasis Rev* 2009, **28**(1-2):77-83.
46. Durkin ME, Yuan B-Z, Zhou X, Zimonjic DB, Lowy DR, Thorgeirsson SS, Popescu NC: **DLC-1:a Rho GTPase-activating protein and tumour suppressor**. *Journal of Cellular and Molecular Medicine* 2007, **11**(5):1185-1207.

47. Seng TJ, Low JS, Li H, Cui Y, Goh HK, Wong ML, Srivastava G, Sidransky D, Califano J, Steenbergen RD *et al*: **The major 8p22 tumor suppressor DLC1 is frequently silenced by methylation in both endemic and sporadic nasopharyngeal, esophageal, and cervical carcinomas, and inhibits tumor cell colony formation**. *Oncogene* 2007, **26**(6):934-944.
48. Qian X, Durkin ME, Wang D, Tripathi BK, Olson L, Yang XY, Vass WC, Popescu NC, Lowy DR: **Inactivation of the Dlc1 gene cooperates with downregulation of p15INK4b and p16Ink4a, leading to neoplastic transformation and poor prognosis in human cancer**. *Cancer Res* 2012, **72**(22):5900-5911.
49. Yang X, Popescu NC, Zimonjic DB: **DLC1 interaction with S100A10 mediates inhibition of in vitro cell invasion and tumorigenicity of lung cancer cells through a RhoGAP-independent mechanism**. *Cancer Res* 2011, **71**(8):2916-2925.
50. Chan SH, Wang LH: **Regulation of cancer metastasis by microRNAs**. *J Biomed Sci* 2015, **22**:9.
51. Ding L, Wang H, Lang W, Xiao L: **Protein kinase C-epsilon promotes survival of lung cancer cells by suppressing apoptosis through dysregulation of the mitochondrial caspase pathway**. *J Biol Chem* 2002, **277**(38):35305-35313.
52. Wan L, Sun M, Liu GJ, Wei CC, Zhang EB, Kong R, Xu TP, Huang MD, Wang ZX: **Long Noncoding RNA PVT1 Promotes Non-Small Cell Lung Cancer Cell Proliferation through Epigenetically Regulating LATS2 Expression**. *Mol Cancer Ther* 2016, **15**(5):1082-1094.
53. Luo SY, Sit KY, Sihoe AD, Suen WS, Au WK, Tang X, Ma ES, Chan WK, Wistuba, II, Minna JD *et al*: **Aberrant large tumor suppressor 2 (LATS2) gene expression correlates with EGFR mutation and survival in lung adenocarcinomas**. *Lung Cancer* 2014, **85**(2):282-292.
54. Mesquita RF, Paul MA, Valmaseda A, Francois A, Jabr R, Anjum S, Marber MS, Budhram-Mahadeo V, Heads RJ: **Protein kinase Cepsilon-calcineurin cosignaling downstream of toll-like receptor 4 downregulates fibrosis and induces wound healing gene expression in cardiac myofibroblasts**. *Mol Cell Biol* 2014, **34**(4):574-594.
55. Zhao R, Pei GX, Cong R, Zhang H, Zang CW, Tian T: **PKC-NF-kappaB are involved in CCL2-induced Nav1.8 expression and channel function in dorsal root ganglion neurons**. *Biosci Rep* 2014, **34**(3).
56. Saurin AT, Durgan J, Cameron AJ, Faisal A, Marber MS, Parker PJ: **The regulated assembly of a PKCepsilon complex controls the completion of cytokinesis**. *Nat Cell Biol* 2008, **10**(8):891-901.
57. Kampfer S, Windegger M, Hochholdinger F, Schwaiger W, Pestell RG, Baier G, Grunicke HH, Uberall F: **Protein kinase C isoforms involved in the transcriptional activation of cyclin D1 by transforming Ha-Ras**. *J Biol Chem* 2001, **276**(46):42834-42842.
58. Wang H, Gutierrez-Uzquiza A, Garg R, Barrio-Real L, Abera MB, Lopez-Haber C, Rosemblit C, Lu H, Abba M, Kazanietz MG: **Transcriptional regulation of oncogenic protein kinase C (PKC) by STAT1 and Sp1 proteins**. *J Biol Chem* 2014, **289**(28):19823-19838.
59. Wu J, Qian J, Li C, Kwok L, Cheng F, Liu P, Perdomo C, Kotton D, Vaziri C, Anderlind C *et al*: **miR-129 regulates cell proliferation by downregulating Cdk6 expression**. *Cell Cycle* 2010, **9**(9):1809-1818.

60. Hori T, Takaori-Kondo A, Kamikubo Y, T U: **Molecular cloning of a novel human protein kinase, kpm, that is homologous to warts/lats, a Drosophila tumor suppressor**. *Oncogene* 2000, **19(27)**:3101-3109.
61. Takahashi Y, Miyoshi Y, Takahata C, Irahara N, Taguchi T, Tamaki Y, al e: **Down-regulation of LATS1 and LATS2 mRNA expression by promoter hypermethylation and its association with biologically aggressive phenotype in human breast cancers**. *Clin Cancer Res* 2005, **11(4)**:1380-1385.
62. Powzaniuk M, McElwee-Witmer S, Vogel RL, Hayami T, Rutledge SJ, Chen F, Harada S, Schmidt A, Rodan GA, Freedman LP *et al*: **The LATS2/KPM tumor suppressor is a negative regulator of the androgen receptor**. *Mol Endocrinol* 2004, **18**(8):2011-2023.
63. Mizuno T, Murakami H, Fujii M, Ishiguro F, Tanaka I, Kondo Y, Akatsuka S, Toyokuni S, Yokoi K, Osada H *et al*: **YAP induces malignant mesothelioma cell proliferation by upregulating transcription of cell cycle-promoting genes**. *Oncogene* 2012, **31**(49):5117-5122.
64. Kamikubo Y, Takaori-Kondo A, Uchiyama T, Hori T: **Inhibition of cell growth by conditional expression of kpm, a human homologue of Drosophila warts/lats tumor suppressor**. *J Biol Chem* 2003, **278**(20):17609-17614.
65. Ke H, Pei J, Ni Z, Xia H, Qi H, Woods T, Kelekar A, Tao W: **Putative tumor suppressor Lats2 induces apoptosis through downregulation of Bcl-2 and Bcl-x(L)**. *Exp Cell Res* 2004, **298**(2):329-338.
66. Aylon Y, Michael D, Shmueli A, Yabuta N, Nojima H, Oren M: **A positive feedback loop between the p53 and Lats2 tumor suppressors prevents tetraploidization**. *Genes Dev* 2006, **20**(19):2687-2700.
67. Tan X, Qin W, Zhang L, Hang J, Li B, Zhang C, Wan J, Zhou F, Shao K, Sun Y *et al*: **A 5-microRNA signature for lung squamous cell carcinoma diagnosis and hsa-miR-31 for prognosis**. *Clin Cancer Res* 2011, **17**(21):6802-6811.

Tables

Table 1. The DEGs, DELs, and DEMs related to significant OS in LUAD patients.

DEGs in survival analysis ($P < 0.05$)*	Gene name
DEGs (29)	ELAVL4, UBASH3B, SCD5, CCNE2, ZEB1, MACC1, DPY19L1, DLC1, ZEB2, PTPRD, SELE, TIMP3, PIK3R1, SASH1, HOXC13, SKP2, PRKCE, ZNF367, LATS2, EPM2A, KLF13, PRICKLE2, CCL20, RALGPS2, BMPR2, TBXA2R, SMAD7, FZD3, RECK
DELs (24)	AC011483.1, AC022148.1, AC107021.1, AL158151.1, ALMS1-IT1, AP000525.1, C1orf195, CARS-AS1, FAM181A-AS1, H19, HOTTIP, LINC00114, LINC00221, LINC00336, LINC00518, LINC00525, LSAMP-AS1, MALAT1, NAV2-AS2, OSBPL10-AS1, STEAP2-AS1, SYNPR-AS1, TTC3-AS1, ZEB2-AS1
DEMs (4)	hsa-mir-21, hsa-mir-200a, hsa-mir-31, hsa-mir-375

* To validate whether the potential DEGs, DELs, and DEMs were significantly associated with the progression of patients with LUAD, K-M survival analyses and log-rank tests were conducted for each gene involved in the above ceRNA network. $P < 0.05$ was considered statistically significant. DEGs: differentially expressed mRNAs; DELs: differentially expressed lncRNAs; DEMs: differentially expressed miRNAs; K-M: Kaplan-Meier; ceRNA: competing endogenous RNA; LUAD: lung adenocarcinoma.

Table 2. Candidate drugs interacted with the significant DEGs.

No.	Gene	Drug	Interaction types	Sources	PMIDS
1	UBASH3B	GEMCITABINE	-	PharmGKB	-
2	ZEB1	DOXORUBICIN	-	CIViC	24013721
3	SELE	CARVEDILOL	inhibitor	DrugBank	15374848
4	PRKCE	INGENOL MEBUTATE	activator	TdgClinicalTrial; GuideToPharmacologyInteractions	-
5	PRKCE	MEPROBAMATE	activator	TdgClinicalTrial; DrugBank; TTD	-
6	TBXA2R	ABACAVIR	agonist	GuideToPharmacologyInteractions	-
7	TBXA2R	ILOPROST	agonist	GuideToPharmacologyInteractions	-
8	TBXA2R	DINOPROSTONE	agonist	GuideToPharmacologyInteractions	-
9	TBXA2R	DINOPROST	agonist	GuideToPharmacologyInteractions	-
10	TBXA2R	FELBINAC	antagonist	GuideToPharmacologyInteractions	-
11	TBXA2R	ACETAMINOPHEN	antagonist	GuideToPharmacologyInteractions	-
12	TBXA2R	CYCLOSPORINE	-	NCI	1385733
13	TBXA2R	MORPHINE	-	NCI	3435201
14	TBXA2R	FUROSEMIDE	-	NCI	7589163
15	TBXA2R	VINBLASTINE	-	NCI	8632653

These data were based on the DGIdb. These targeted DEGs were all statistically significant in the survival analysis of LUAD patients. DGIdb: drug-gene interaction database; DEGs: differentially expressed mRNAs; LUAD: lung adenocarcinoma.

Figures

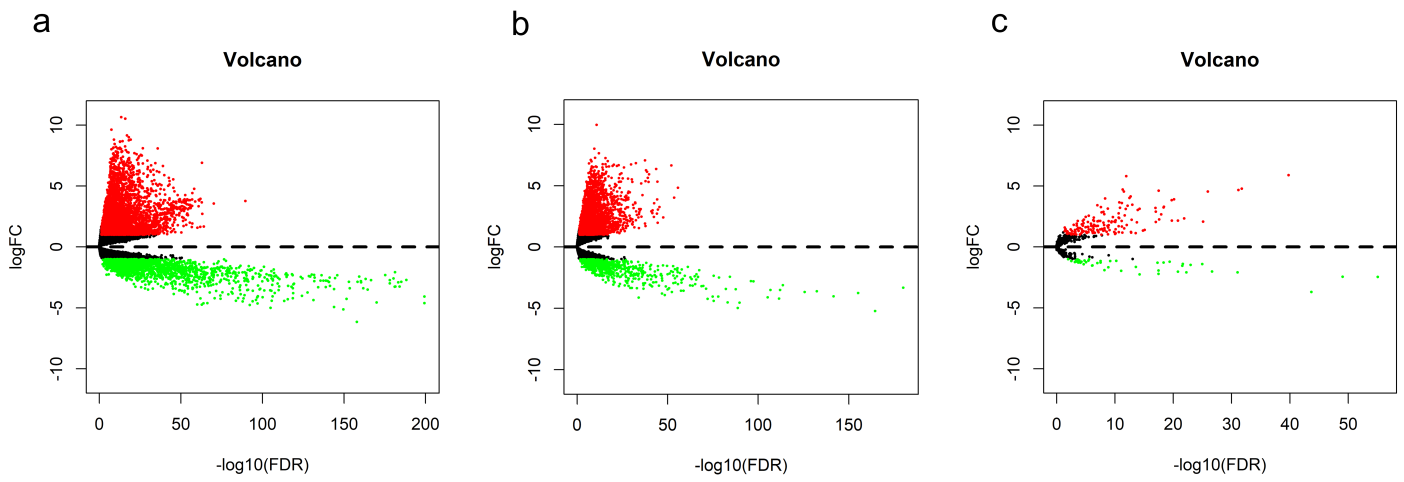

Figure 1

Selection of cDEGs, cDELs, and cDEMs. Volcano maps of differentially expressed (a) mRNAs, (b) lncRNAs, and (c) miRNAs were generated to visualise the differences between LUAD cases and normal controls. Compared to the expression levels of these genes in normal controls, red dots indicate upregulation in LUAD, while green dots indicate downregulation in LUAD. FC: fold change; FDR: false discovery rate; LUAD: lung adenocarcinoma.

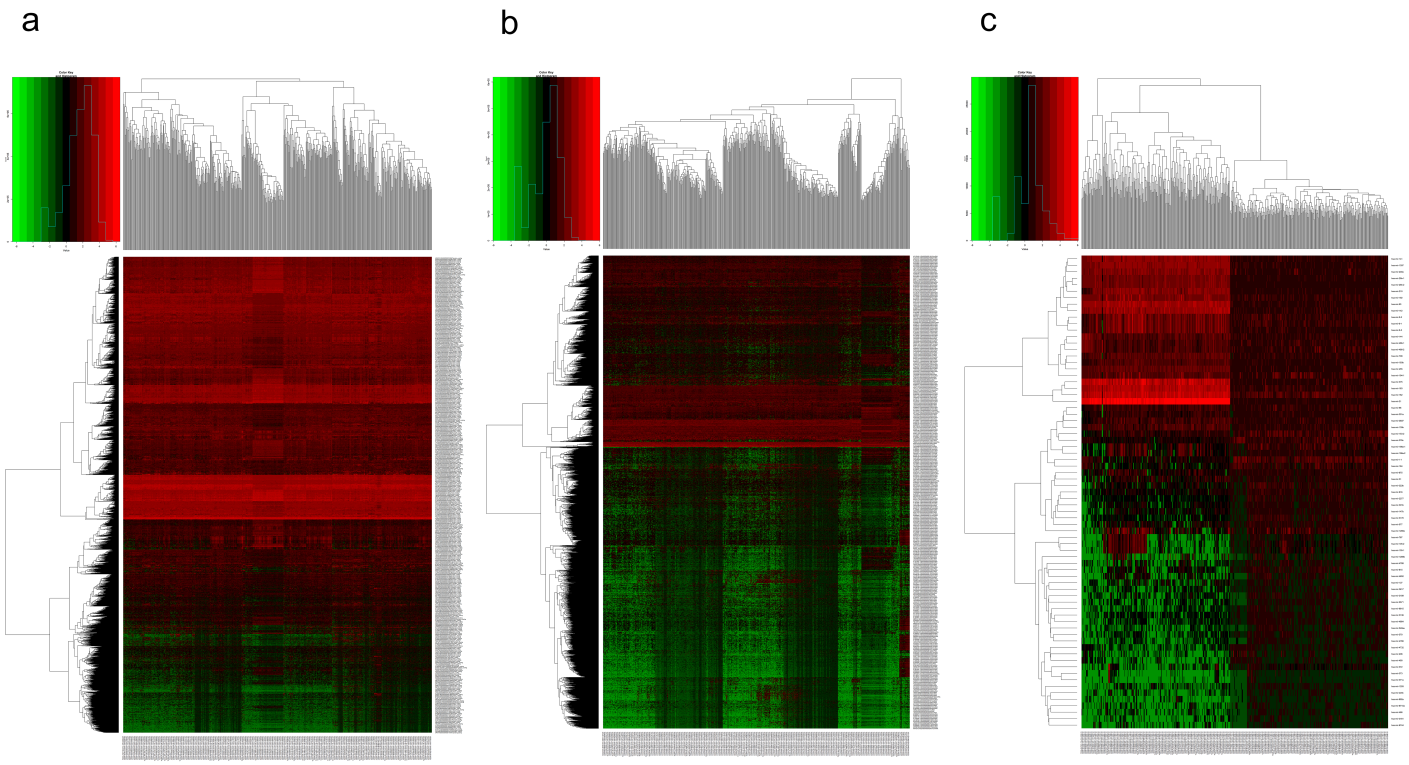

Figure 2

Heatmaps of the cDEGs, cDELs, and cDEMs between LUAD cases and normal controls. Hierarchical clustering analyses for cDEGs, cDELs, and cDEMs were conducted through the “pheatmap” package in R. Red and green colours represent high and low expression levels of (a) cDEGs, (b) cDELs, and (c) cDEMs, respectively.

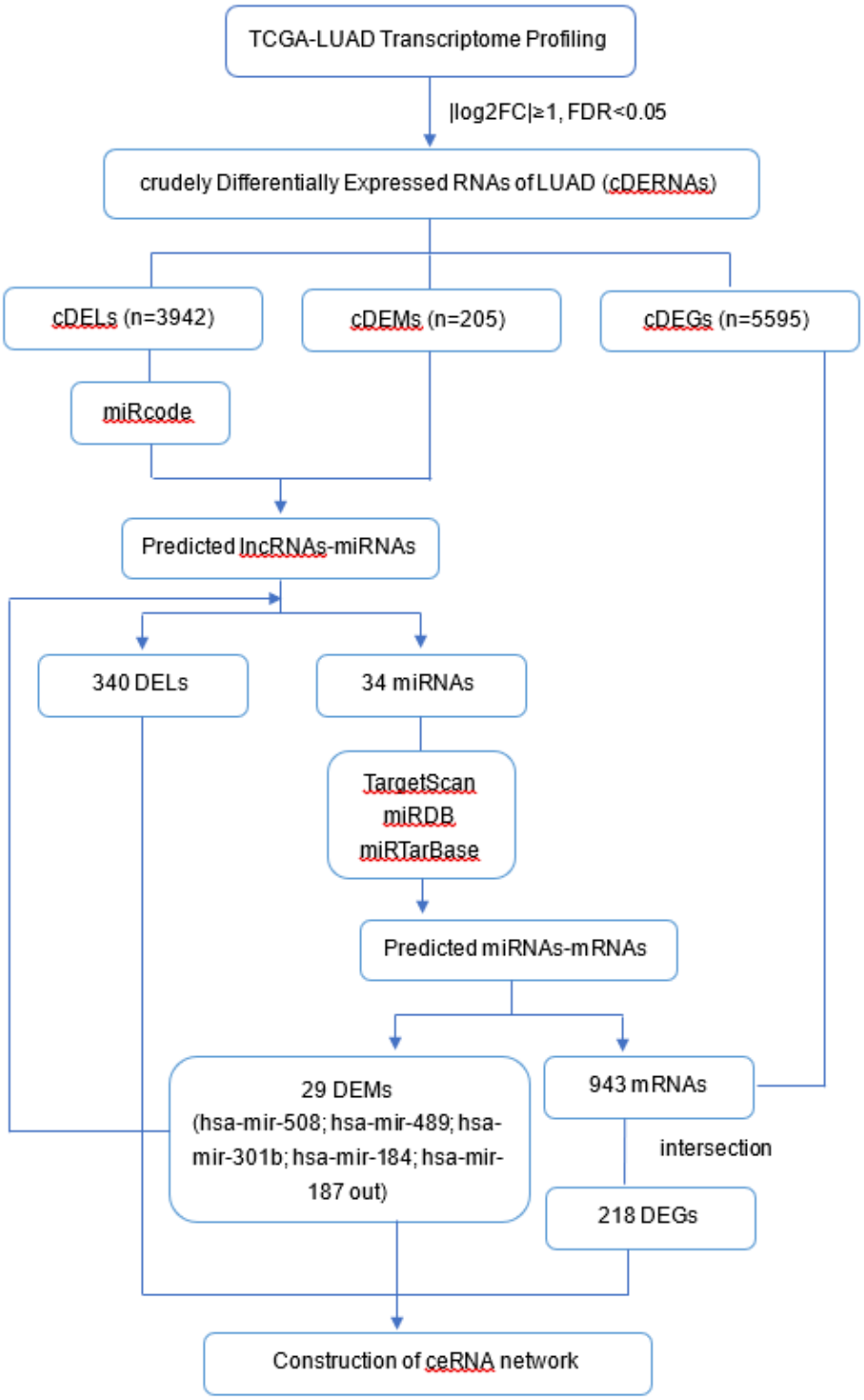

Figure 3

A flowchart on the construction of the ceRNA network.

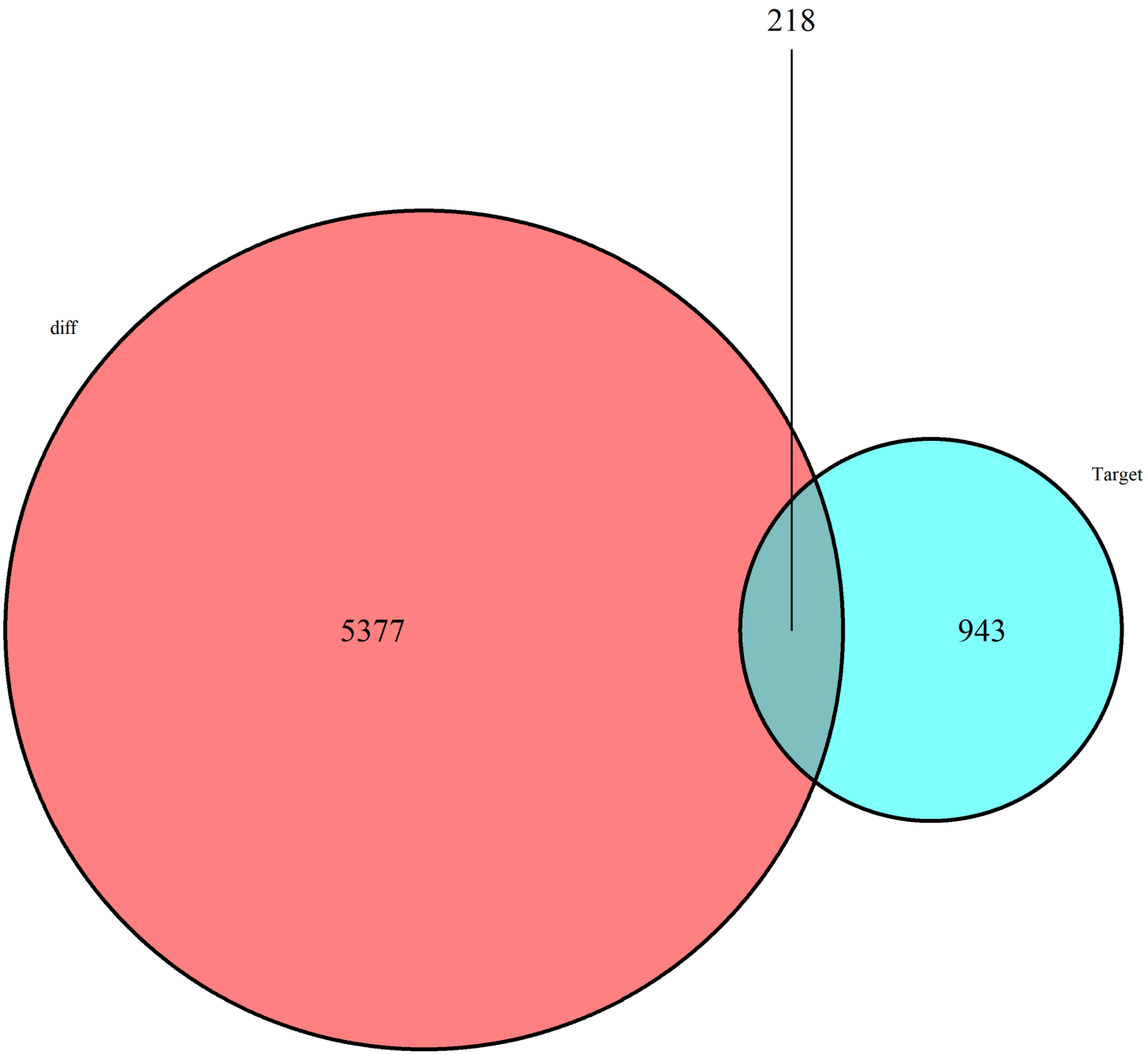

Figure 4

The intersection of the candidate mRNAs and cDEGs.

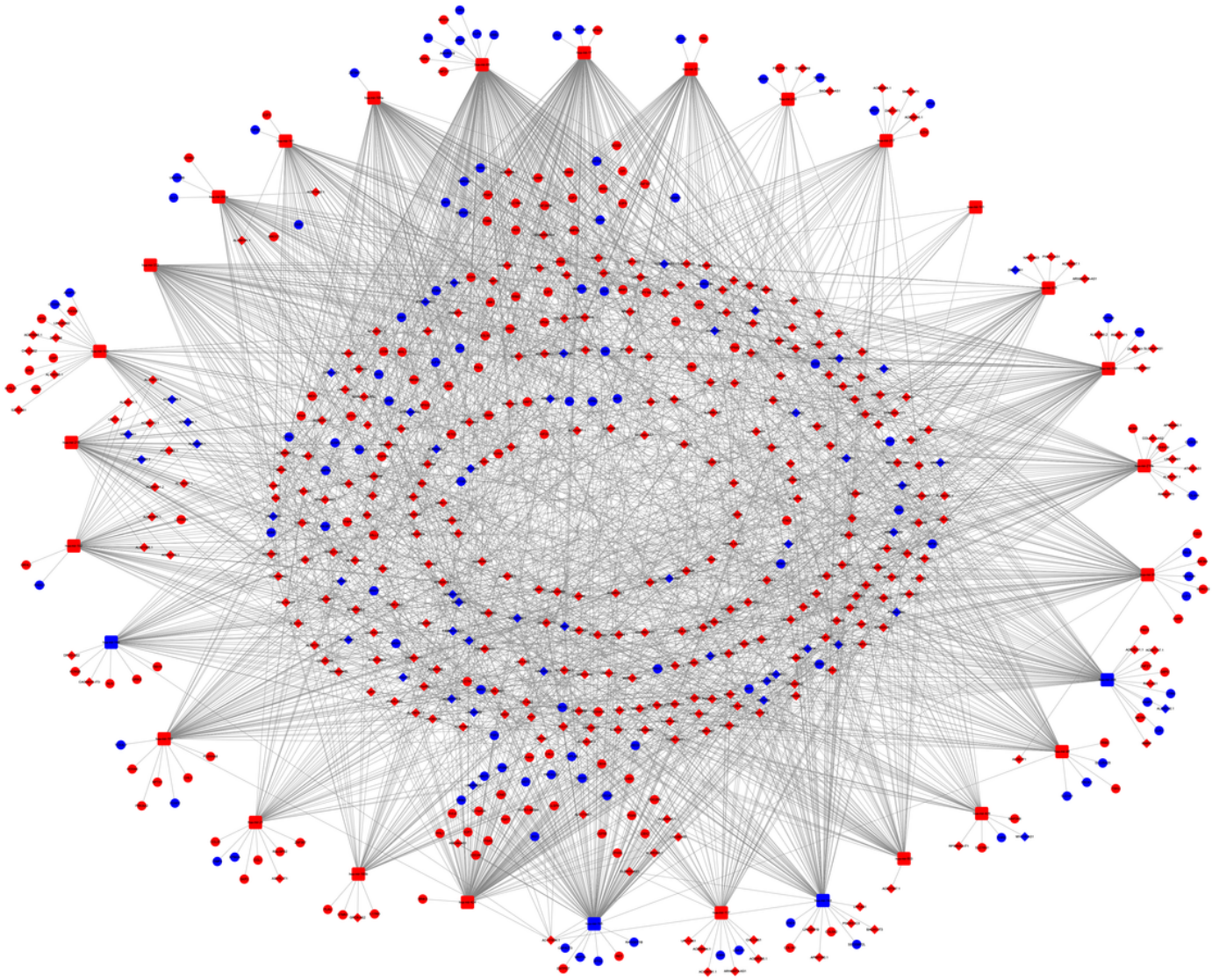

Figure 5

The initial lncRNA-miRNA-mRNA ceRNA network in LUAD. The forms of diamond, square, and circle represent lncRNAs, miRNAs, and mRNAs, respectively. The colour of red indicates upregulated genes, whereas the colour of blue indicates downregulated genes. ceRNA: competing endogenous RNA.

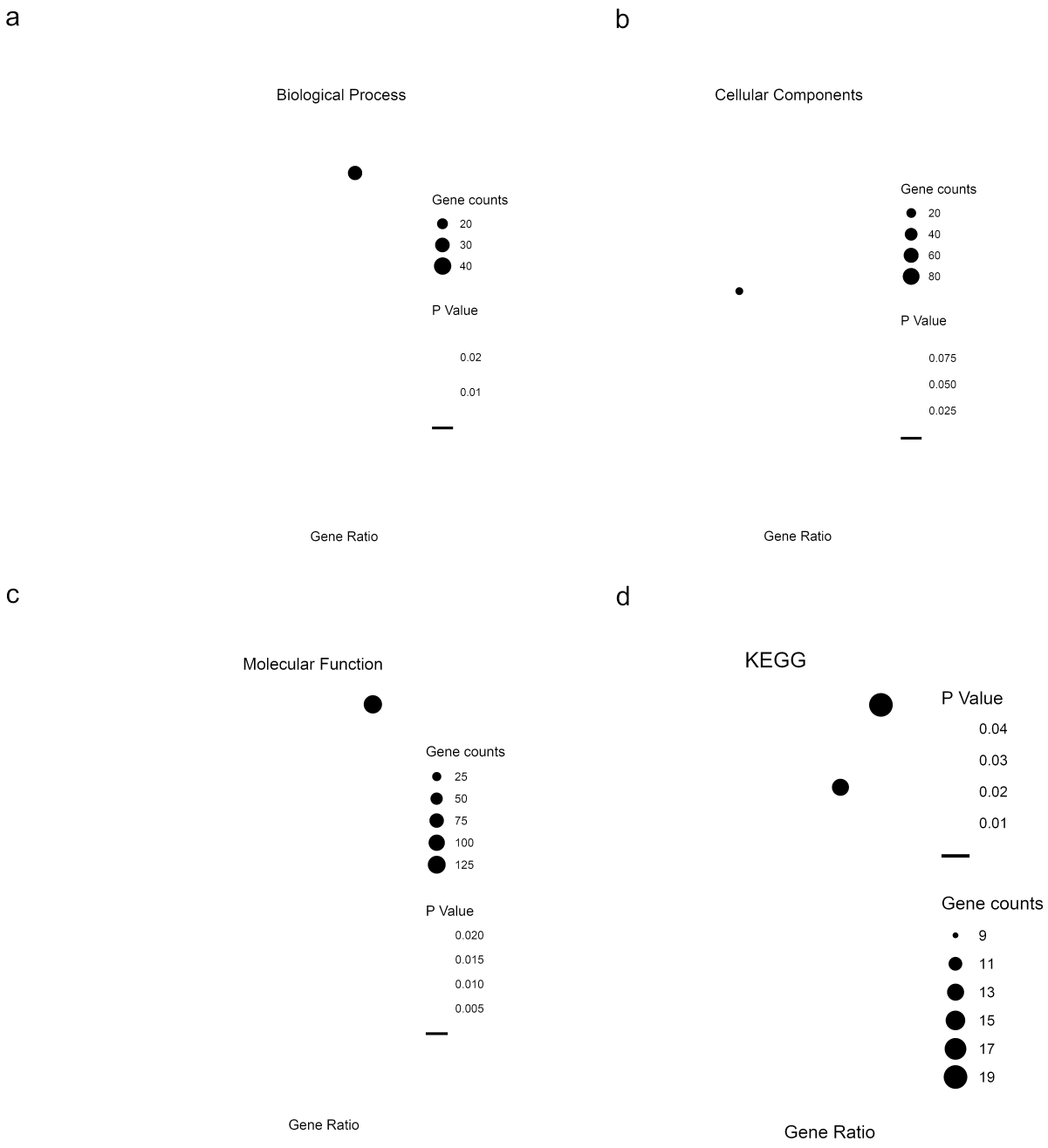

Figure 6

Bubble maps for GO and KEGG analyses of DEGs involved in the ceRNA network. The DAVID database (https://david.ncifcrf.gov/; version: 6.8) was used to perform GO and KEGG analyses. The top 10 items of these enrichment analyses are conducted by using the “ggplot2” package in R software. $P < 0.05$ was considered statistically significant. (a) BP; (b) CC; (c) MF; (d) KEGG pathways. GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; DAVID: database for annotation, visualization and integrated discovery; BP: biological process; CC: cellular component; MF: molecular function.

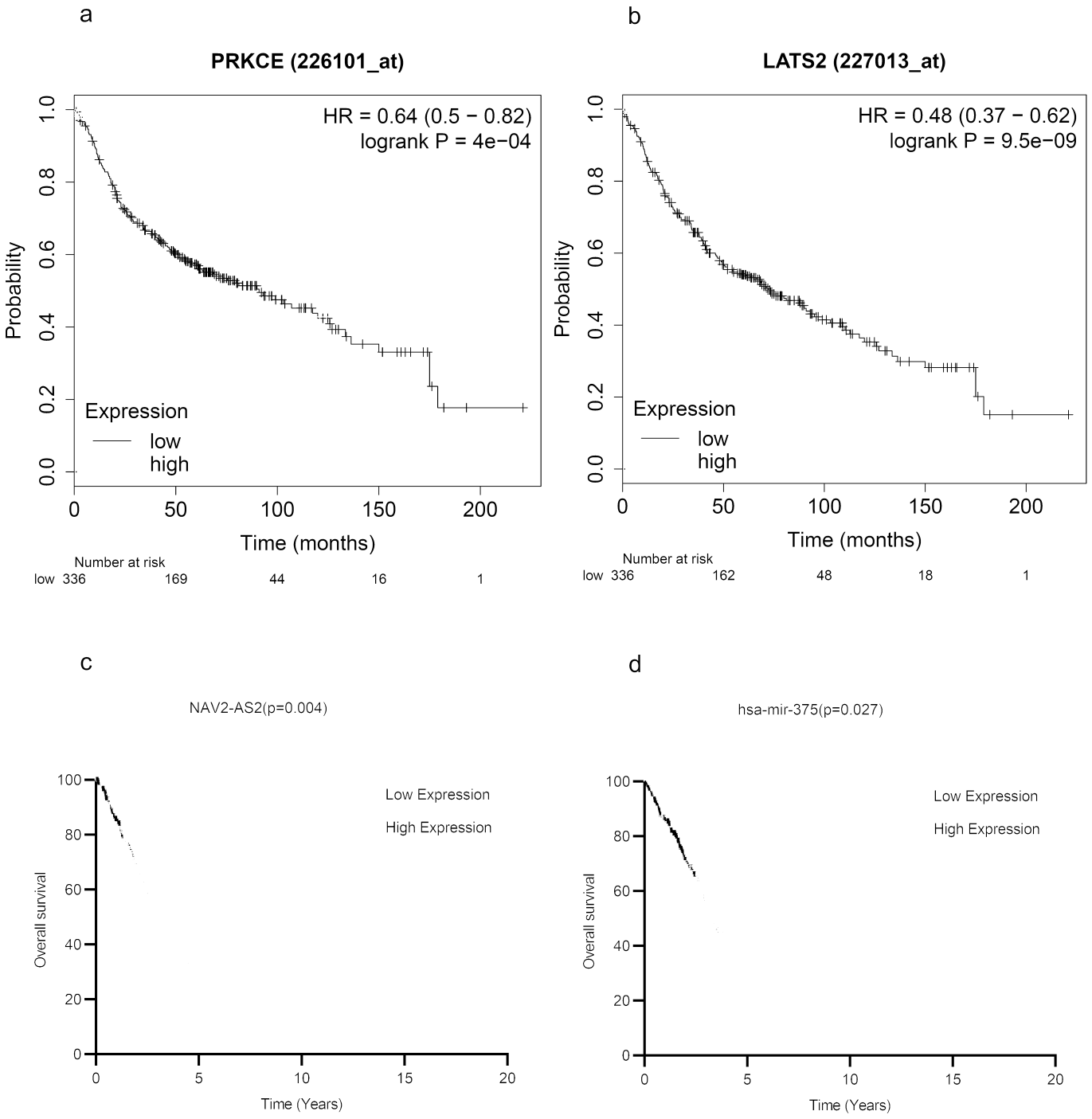

Figure 7

OS analyses of DEGs, DELs, and DEMs in patients with LUAD. The survival analyses for (a) PRKCE and (b) LATS2 are conducted by Kaplan-Meier Plotter database (http://kmplot.com/analysis/). The survival analyses for (c) NAV2-AS2 and for (d) hsa-mir-375 are conducted by using the “survival” package in R software based on the data from TCGA database. Survival curves were constructed based on the low and high expression of DEGs, DELs, and DEMs in LUAD patients. The P value of log-rank test less than 0.05

was considered statistically significant. DEGs: differentially expressed mRNAs; DELs: differentially expressed lncRNAs; DEMs: differentially expressed miRNAs; LUAD: lung adenocarcinoma.

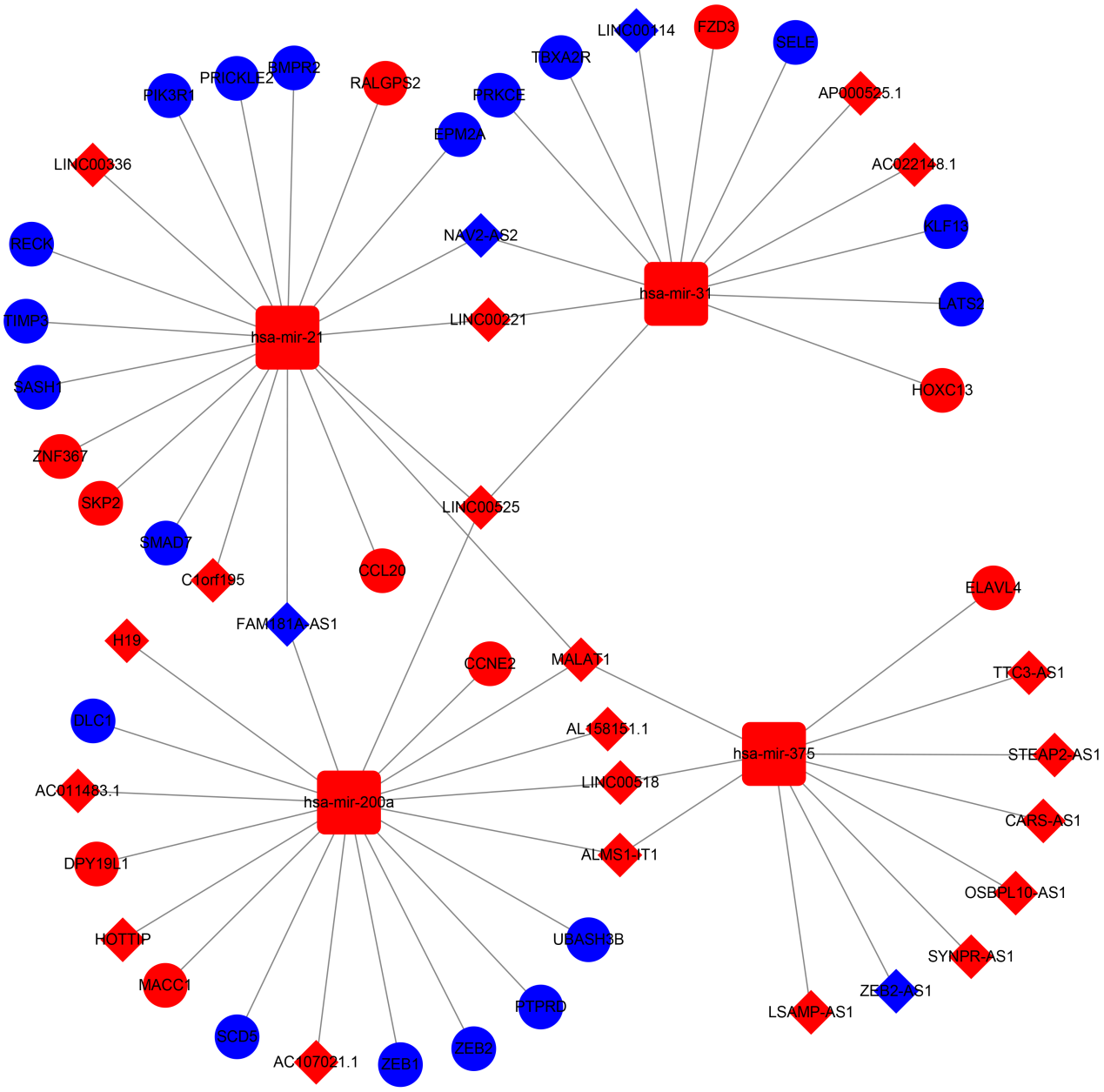

Figure 8

The significant lncRNA-miRNA-mRNA ceRNA network. All DELs, DEMs, and DEGs involved in the significant ceRNA network were validated by survival analysis, and the results were all statistically significant. The diamonds, squares, and circles represent lncRNAs, miRNAs, and mRNAs, respectively. Red indicates upregulated genes, whereas blue indicates downregulated genes. DEGs: differentially expressed mRNAs; DELs: differentially expressed lncRNAs; DEMs: differentially expressed miRNAs; ceRNA: competing endogenous RNA.

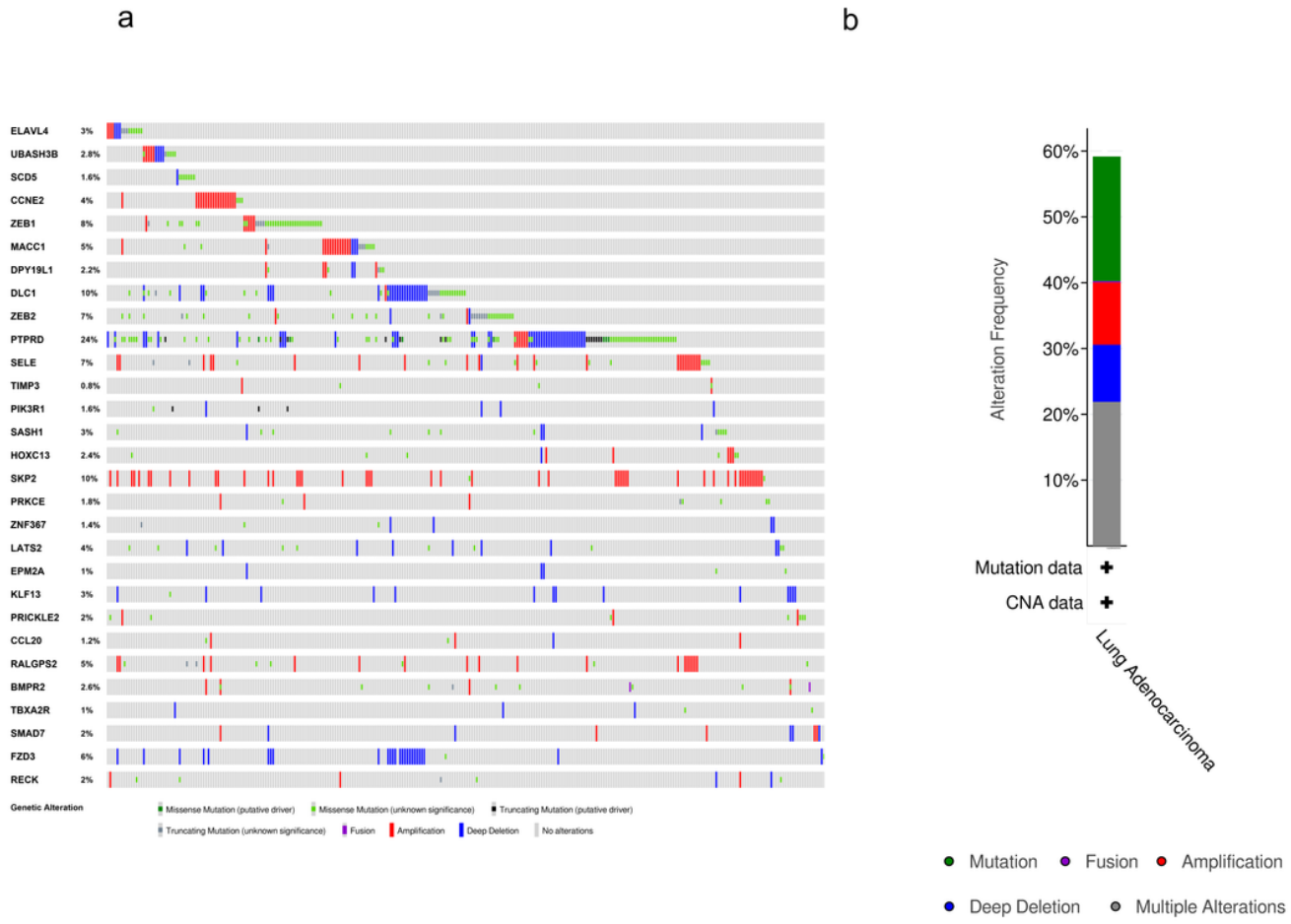

Figure 9

The alteration information on significant DEGs. (a) The genetic alterations associated with the significant DEGs are presented in a visual summary across a set of LUAD samples (data from TCGA, PanCancer Atlas). (b) An overview of the main types of alteration information on the significant DEGs in the genomics datasets of LUAD in the TCGA database. DEGs: differentially expressed mRNAs.

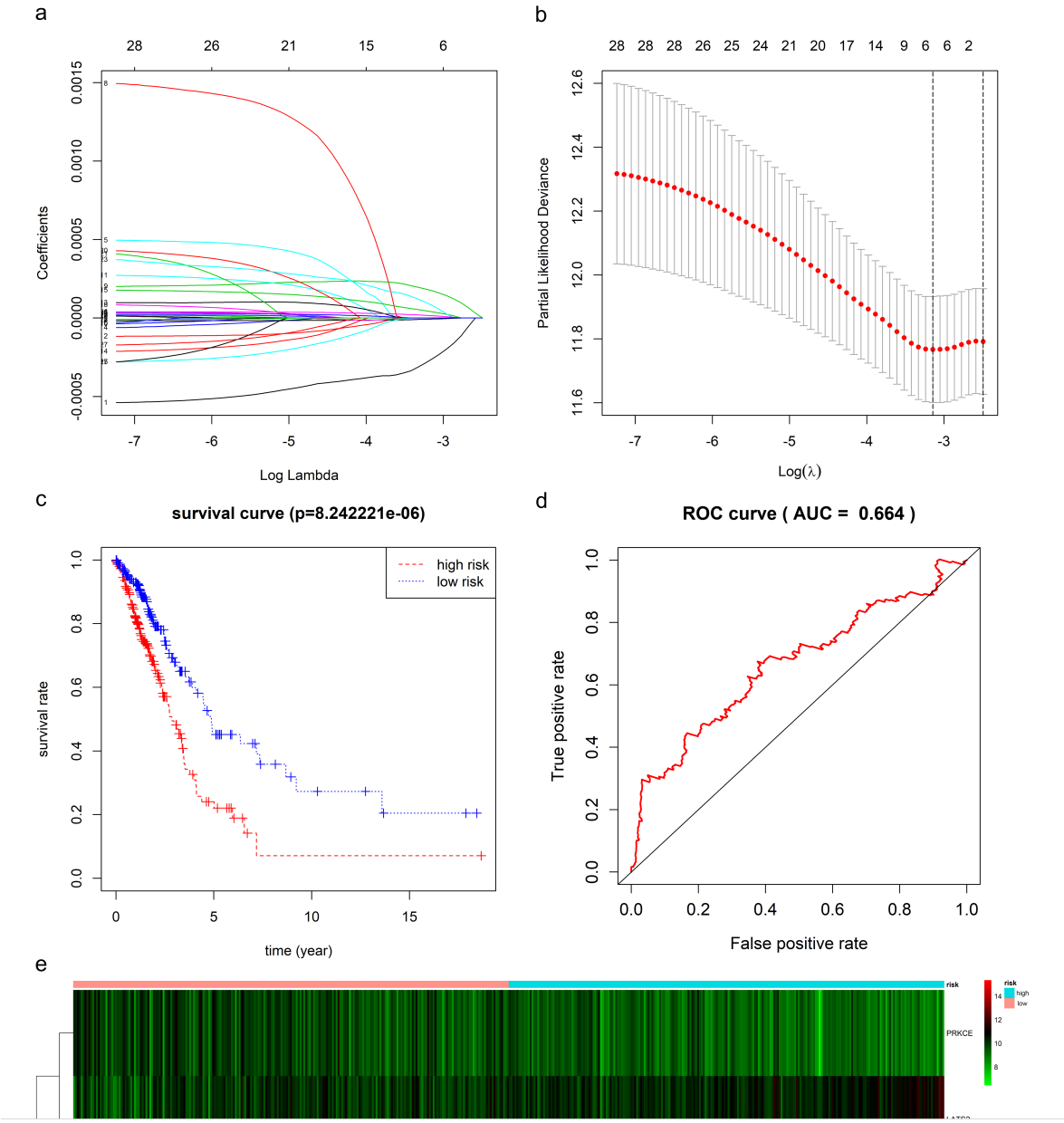

Figure 10

The alteration information on significant DEGs. (a) Shows the coefficient values at different levels of allowed error, and each curve represents an mRNA. (b) The best lambda was validated by ten-fold cross-validation to minimize the mean cross-validated error. The red dots and solid vertical lines indicate the partial likelihood deviances and their corresponding 95% CIs, respectively. (c) Presents the K-M survival curve with log-rank test between low- and high-risk groups. (d) Illustrates the time-dependent ROC curve

based on the risk score system. (e) Expression levels of the DEGs involved in the risk score system between low- and high-risk groups are shown in the form of a heatmap. The light red and light blue in the above heatmap represent high- and low-risk groups, respectively. The red and green in the main body of the heatmap represent high- and low expression levels of DEGs involved in the risk score system, respectively. K-M: Kaplan-Meier; ROC: receiver operating curve. DEGs: differentially expressed mRNAs.

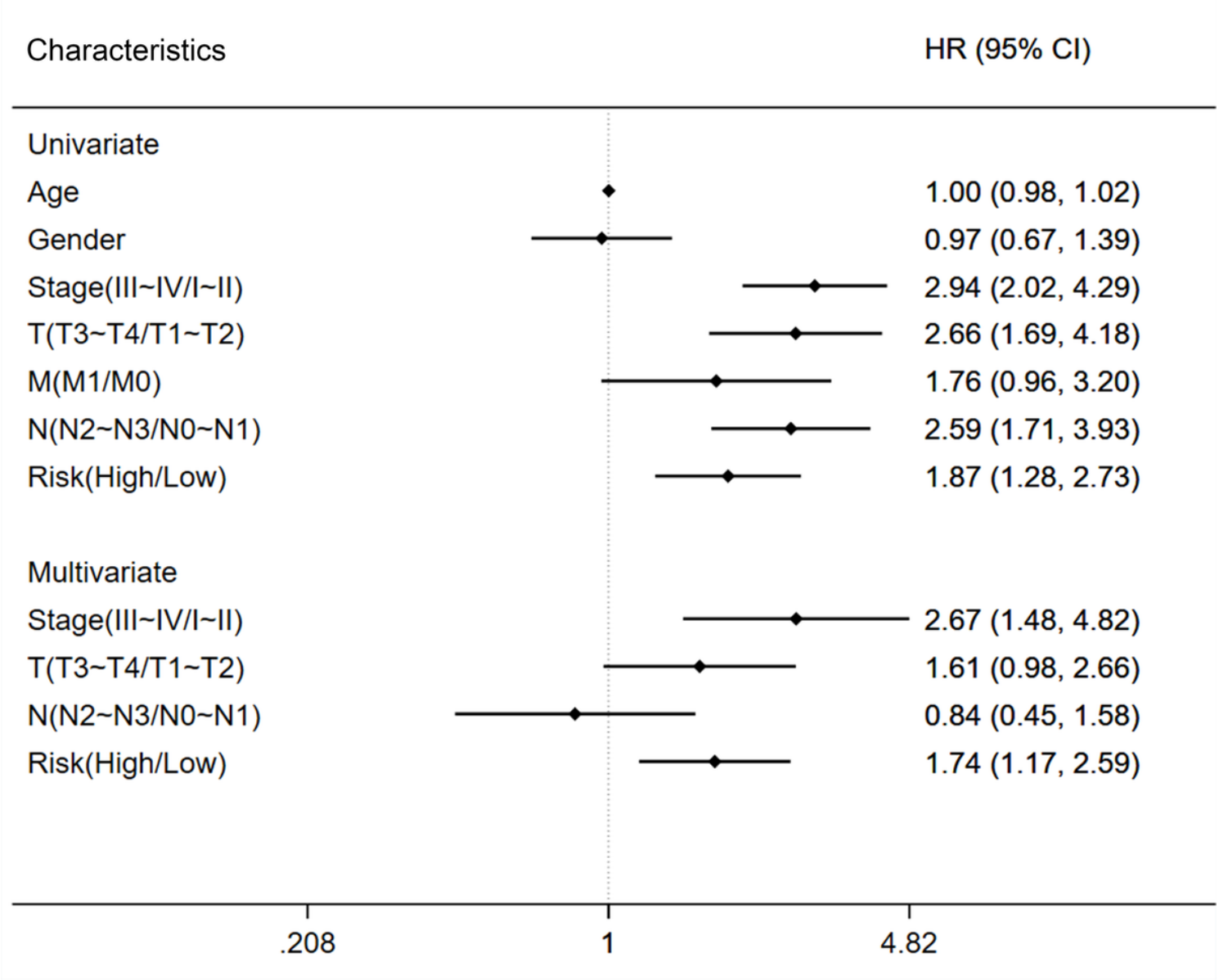

Figure 11

Univariate and multivariate Cox regression analyses in LUAD patients. Survival prognosis forest map of clinical information related to prognosis in lung adenocarcinoma (LUAD) patients. Each point in the forest plot represents the HR of the indicator, and the line on both sides of the point represents the 95% confidence interval (95% CI).

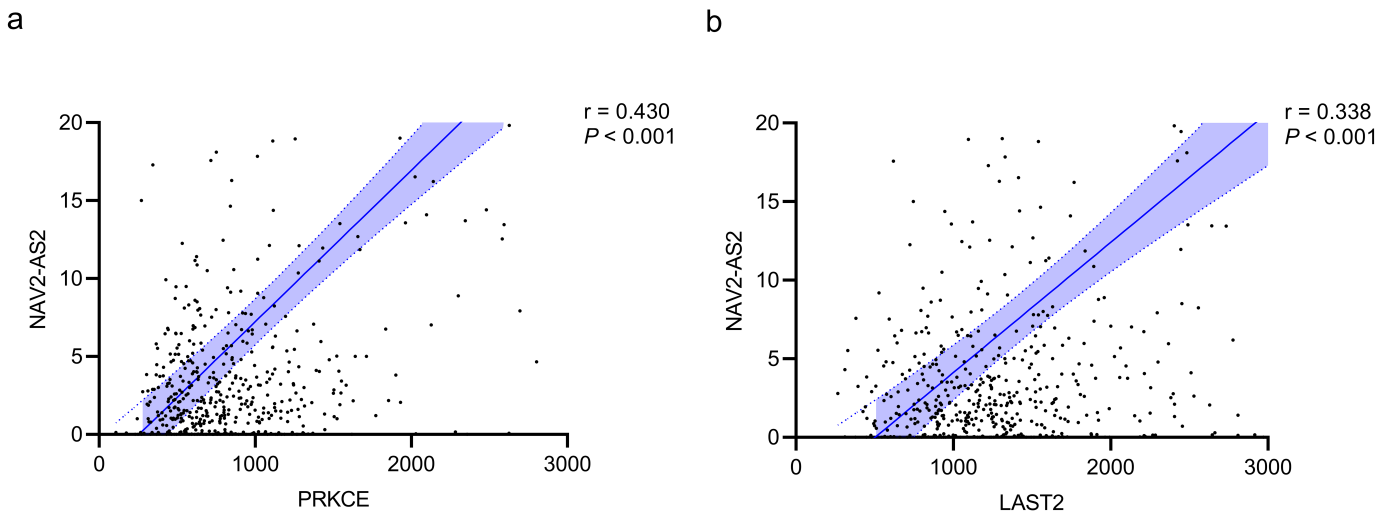

Figure 12

Correlations between DEGs involved in the risk score system and their corresponding DELs. Linear regression and correlation analyses were conducted between the expression levels of DEGs and their corresponding DELs. The blue shadow around the blue line represents the 95% CI. P < 0.05 and r > 0.30 were considered statistically significant. DEGs: differentially expressed mRNAs; DELs: differentially expressed lncRNAs.

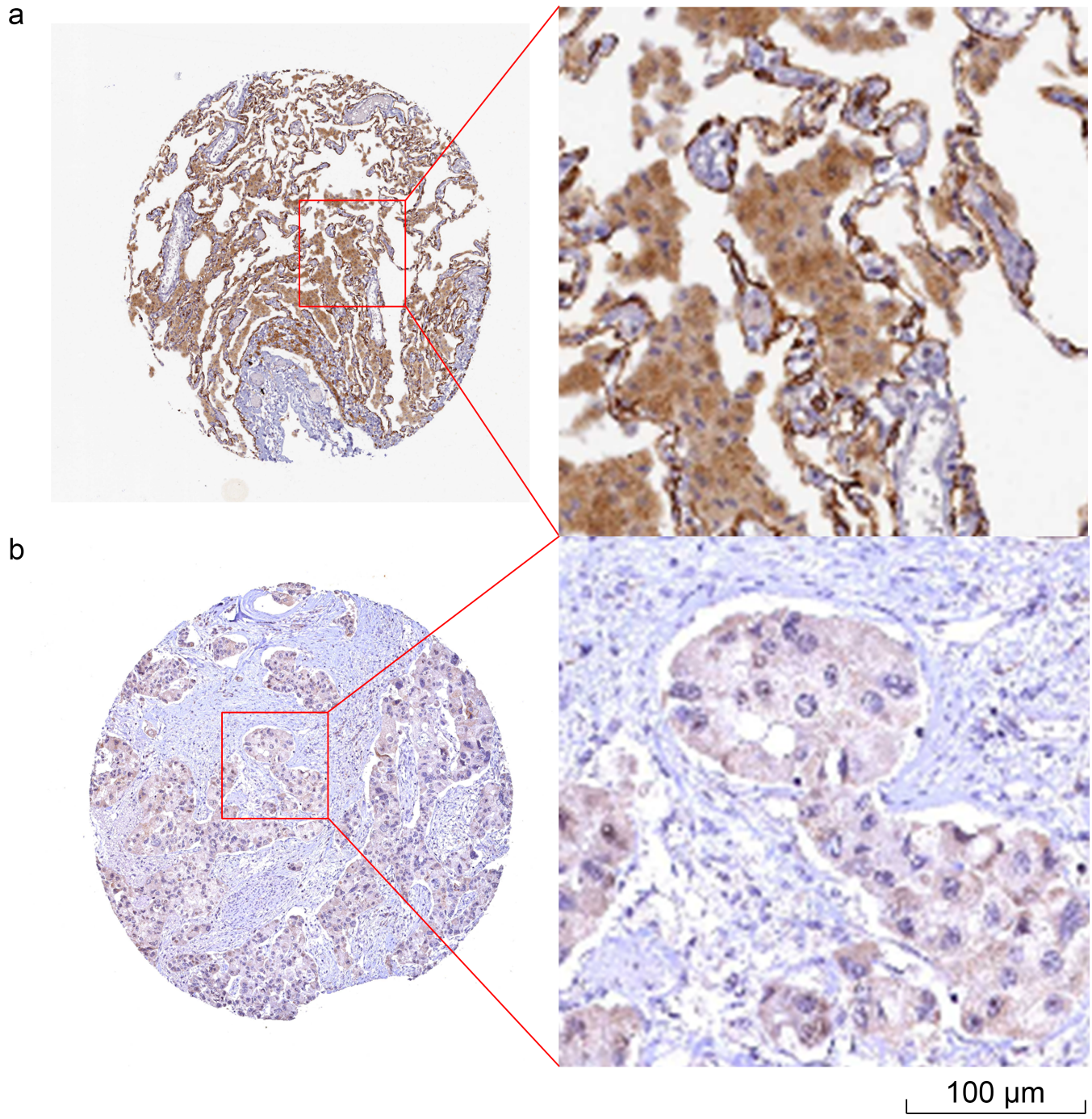

Figure 13

IHC results of LATS2 protein levels in normal lung and LUAD from the Human Protein Atlas. (a) IHC results of LATS2 protein in normal lung. Staining: medium; Intensity: moderate; Quantity: > 75%; Location: cytoplasmic / membranous. (b) IHC results of LATS2 protein in LUAD. Staining: not detected; Intensity: weak; Quantity: < 25%; Location: nuclear.

Supplementary Files

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

- Supplementarymaterials.docx