

Construction of a metastasis-associated ceRNA network reveals a prognostic signature in lung cancer

Qing Cao

Northwest University

Zewen Dong

Northwest University

Shuzhen Liu

Northwest University

Guoyan An

Northwest University

Bianbian Yan

Northwest University

Lei Lei (✉ leilei@nwu.edu.cn)

Northwest University <https://orcid.org/0000-0002-3457-1907>

Primary research

Keywords: lung cancer, metastasis, ceRNA, lncRNA, LINC01010

Posted Date: April 2nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-20435/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Cancer Cell International on June 3rd, 2020. See the published version at <https://doi.org/10.1186/s12935-020-01295-8>.

Abstract

Background: Lung cancer is the most common cancer worldwide, and metastasis is the leading cause of lung cancer related death. However, the molecular network involved in lung cancer metastasis remains incompletely described. Here, we aimed to construct a metastasis-associated ceRNA network and identify a lncRNA prognostic signature in lung cancer.

Methods: RNA expression profiles were downloaded from The Cancer Genome Atlas (TCGA) database. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses and gene set enrichment analysis (GSEA) were performed to investigate the function of these genes. Using Cox regression analysis, we found that a 6 lncRNA signature may serve as a candidate prognostic factor in lung cancer. Finally, we used Transwell assays with lung cancer cell lines to verify that LINC01010 acts as a tumor suppressor.

Results: We identified 1249 differentially expressed (DE) mRNAs, 440 DE lncRNAs and 26 DE miRNAs between nonmetastatic and metastatic lung cancer tissues. GO and KEGG analyses confirmed that the identified DE mRNAs are involved in lung cancer metastasis. Using bioinformatics tools, we constructed a metastasis-associated ceRNA network for lung cancer that includes 117 mRNAs, 23 lncRNAs and 22 miRNAs. We then identified a 6 lncRNA signature (LINC01287, SNAP25-AS1, LINC00470, AC104809.2, LINC00645 and LINC01010) that had the greatest prognostic value for lung cancer. Furthermore, we found that suppression of LINC01010 promoted lung cancer cell migration and invasion.

Conclusions: This study might provide insight into the identification of potential lncRNA biomarkers for diagnosis and prognosis in lung cancer.

Background

Lung cancer is the most common cancer worldwide and the leading cause of cancer-related death in men and the second in women [1, 2]. In recent years, several studies have shown that abnormalities in noncoding genes are associated with lung cancer pathogenesis [3–6], but the mechanism whereby noncoding genes affect lung cancer metastasis remains incompletely understood.

The ENCODE (Encyclopedia of DNA Elements) Consortium revealed that less than 2% of the human genome is comprised of protein coding genes, while a dominant portion of transcripts are noncoding genes, which includes long noncoding RNAs (lncRNAs), pseudogenes and microRNAs (miRNAs) [7–9]. lncRNAs used to be considered transcriptional noise that have no biological function. Recently, increasing studies have revealed that lncRNAs are involved in many cellular processes, such as myocyte differentiation, immune response, cancer cell metastasis, proliferation, and drug resistance [10–12]. For instance, overexpression of lncRNA HAND2-AS1 inhibited migration of non-small cell lung cancer cells by downregulating TGF- β 1 [13]. Furthermore, Fang et al. reported that lncRNA HOTAIR affects chemoresistance by regulating HOXA1 methylation in small cell lung cancer [14].

MiRNA is an endogenous small non-coding RNA that also plays an important biological role in the development and metastasis of lung cancer [15]. Recently, L Salmena et al. proposed the competitive endogenous RNA (ceRNA) hypothesis in which lncRNAs are able to regulate mRNAs expression as “miRNA sponges” by preferentially occupying the miRNAs response elements [16]. Therefore, construction of a ceRNA network could provide new perspectives for evaluating cancer regulatory networks.

In this study, we analyzed genomic data along with clinical information from The Cancer Genome Atlas (TCGA). Next, we used bioinformatics tools to construct a metastasis-associated lncRNAs-miRNAs-mRNAs ceRNA network in lung cancer. Using multivariate Cox regression analysis, we discovered that a signature based on 6 lncRNAs may serve as an independent prognostic factor in lung cancer. Furthermore, we validated LINC01010 as a tumor suppressor lncRNA and might be involved in lung cancer ceRNA by competing with miR-372. This study reveals a ceRNA network in metastatic lung cancer, which may provide a useful basis for formulating early diagnosis and individualized treatments.

Materials

Data collection and differential expression analysis

Patient sample data sets and clinical information were extracted from TCGA database using the Genomic Data Commons (GDC) data transfer tool, including 32 lung cancer metastasis (M1) cases and 741 lung cancer nonmetastatic (M0) cases. We then used the “edgeR” package in R software to identify differentially expressed (DE) mRNAs, lncRNAs and miRNAs between M1 and M0 groups with thresholds of $|\log_{2}FC| > 1$ and $FDR < 0.05$ [17]. All analyses were performed using R version 3.3.2.

Functional enrichment analysis

To better understand the function of DE genes, we divided 1249 DE mRNAs into upregulated and downregulated groups for GO (Gene Ontology) analysis and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis. We analyzed GO biological processes using DAVID (<https://david.ncifcrf.gov/>) and KEGG pathways using KOBAS (<http://kobas.cbi.pku.edu.cn/>). The top 15 pathways of GO and KEGG analysis were visualized using the R package ggplot2.

To explore the function of LINC01010, we searched for mRNAs co-expressed with LINC01010 using the Multi-Experiment Matrix (MEM) resource (<https://biit.cs.ut.ee/mem/>). Then, the top 200 mRNAs co-expressed with LINC01010 were selected for GO and KEGG pathway enrichment analysis.

To investigate the function of has-miR-372 and LINC01010, lung cancer gene sets were downloaded from TCGA database and used to perform gene set enrichment analysis (GSEA). The phenotype label comprised groups with high-levels of has-miR-372 or LINC01010 versus those with low levels of has-miR-372 or LINC01010. *P*-values were used to estimate the statistical significance of enrichment scores.

Constructing the competitive endogenous RNA network

We constructed a ceRNA network containing DE mRNAs, miRNAs, and lncRNAs using bioinformatics tools. TargetScan (<http://www.targetscan.org/>) and miRDB (<http://www.mirdb.org/>) were used to predict mRNA-miRNA interactions. Then, DIANA was used to predict lncRNA-miRNA interactions. The predicted intersection results were retained to construct the ceRNA network (Fig. 1), which was visualized using Cytoscape v3.6.1.

Prognosis and survival curve analysis

Multivariate COX proportional hazards regression analysis was performed to identify a prognostic model of the lncRNAs in the ceRNA network. Mathematical models were established based on the Akaike Information Criterion (AIC) using the "Survival" R package [18]. We calculated the prognostic risk score as follows:

$$\text{Risk Score} = (0.0662) * \text{LINC01287} + (0.0978) * \text{SNAP25-AS1} + (0.0567) * \text{LINC00470} + (-0.0745) * \text{AC104809.2} + (-0.2018) * \text{LINC00645} + (-0.0726) * \text{LINC01010}.$$

All lung cancer patients were divided into low-risk and high-risk groups based on the median risk score. Overall survival (OS) curves of the two groups were generated using Kaplan-Meier analysis. The predictive value of the prognostic model over 5 years was evaluated by time-dependent receiver operating characteristic (ROC) curve analysis using the "survival ROC" R package [19].

Cell culture and siRNA transfection

Human lung cancer cell lines (SPC-A-1 and A549) were purchased from American Type Culture Collection (ATCC). SPC-A-1 and A549 cells were cultured in Roswell Park Memorial Institute 1640 medium and Dulbecco's Modified Eagle's Medium (HyClone), respectively, supplemented with 10% fetal bovine serum (Gibco, Rockville, MD) at 37°C and 5% CO₂. A small interfering RNA for LNC01010 (siLNC01010) and a negative control (siRNA-NC) were used in knockdown function experiments. 50 nM siRNA-NC and 50 nM siLNC01010 (GenePharma, China) were transfected into A549 and SPC-A-1 cells using X-treme GENE siRNA Transfection Reagent (Roche, USA). Forty-eight hours after transfection, cells were harvested for RNA extraction or other functional experiments. SiRNA sequences are as follows: siLINC01010-1, 5'-GCUGUUUGCUGGCAACAAATT-3'; siLINC01010-2, 5'-GCUGUUUGCUGGCAACAAATT-3'; siLINC01010-3, 5'-GCAGCAA AUGUAGAAACAUTT-3'.

RNA isolation and quantitative real time-PCR

Total RNA from cell samples was extracted using the TRIzol (Invitrogen, Carlsbad, CA) Pyrolysis method. Total RNA was reverse transcribed into complementary DNA using a Prime Script RT reagent kit (Takara, Tokyo, Japan). Bestar™ qPCR MasterMix (DBI Bioscience) was used to perform qRT-PCR on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad, Hercules, CA). The following primers were used in qRT-PCR: LNC01010 forward, 5'-AATGATGCGGCTGAACAA-3', and reverse, 5'-CCTTGGCTTGCCTATTACC-3'; GAPDH forward, 5'-TGCAAATCCCATCACCATCT-3', and reverse 5'-TGGACTCCACGACGTACTCA-3'.

GAPDH was used as an endogenous control. qRT-PCR conditions were as follows: 95°C for 3 minutes, followed by 40 cycles of 95°C for 10 seconds and 60°C for 34 seconds with a melt curve from 60 to 95°C.

Cell migration and invasion assays

Cell migration and invasion abilities were evaluated using Transwell assays (Corning, MA, USA). 5×10^4 cells were suspended in the upper chamber, and 700 μ l medium containing 10% FBS was placed into the lower chamber. After 24 h incubation, A549 and SPC-A-1 cells that had migrated through the membrane were fixed with methanol for 15 min and stained with 10% crystal violet. For invasion assays, Matrigel (BD Biosciences, Bedford, MA, USA) was added into the upper chambers 12 hours before the experiments. The number of cells were quantified in six random fields, and three independent experiments were performed.

Statistical analysis

Unpaired t-test (two-tailed) was used to analyze different groups of cellular experiments in GraphPad Prism 6 (GraphPad Software, Inc., San Diego, CA). Pearson's test was used to measure the correlation between expression of LINC01010 and miR-372. $P < 0.05$ was considered statistically significant.

Results

Differentially expressed lncRNAs, miRNAs and mRNAs

To construct the metastasis-associated ceRNA network in lung cancer, we downloaded RNA expression profiles from TCGA database, which included 773 lung cancer samples. The differentially expressed (DE) mRNAs, lncRNAs and miRNAs between 32 metastatic tumor tissues and 741 nonmetastatic tumor tissues were explored using the "edgeR" package in R software. With the criteria of $|\log FC| > 1$ and $FDR < 0.05$, we identified 1249 DE mRNAs (569 upregulated and 680 downregulated), 440 DE lncRNAs (221 upregulated and 219 downregulated) and 26 miRNAs (21 up-regulated and 5 downregulated) between nonmetastatic and metastatic tumor tissues. The results indicate that aberrant expression of these genes might be involved in lung cancer migration.

GO and KEGG pathway analysis of DE genes

To better understand the function of the identified DE genes, DE mRNAs were divided into upregulated and downregulated groups for analysis by DAVID and KOBAS bioinformatics resources. The top 15 GO biological processes and KEGG pathways of dysregulated genes are shown in Fig. 2 Among the top 15 GO and KEGG pathways of upregulated mRNAs (Fig. 2a, b), "GO:0006810~transport", "GO:0003341~cilium movement", "MAPK signaling pathway", "ECM-receptor interaction" and "focal adhesion" are reported to promote invasion and metastasis in cancer [20, 21]. Downregulated genes participated in "GO:0031424~keratinization", "GO:0008544~epidermis development" and "drug metabolism - cytochrome P450" (Fig. 2c, d), which reportedly inhibit invasion and metastasis of cancer

[22, 23]. These results suggest that these above pathways may play important roles in lung cancer metastasis.

Construction of the ceRNA network

To construct the metastasis-associated ceRNA network, we predicted interactions among DE mRNAs, miRNAs and lncRNAs using bioinformatics tools. Target mRNAs of DE miRNAs were predicted using TargetScan and miRDB, and the intersection of these two databases was selected. After we discarded target mRNAs that were not among DE mRNAs, 117 mRNAs (95 upregulated and 22 downregulated) were used to construct the ceRNA network (Table S1). Next, we utilized DIANA to predict interactions between DE lncRNAs and DE miRNAs. We found the 22 miRNAs that are predicted to interact with 23 lncRNAs (Table S2). Ultimately, 23 lncRNAs (20 downregulated and 3 upregulated) (Table S3) and 22 miRNAs (4 downregulated and 18 upregulated) (Table S4) were included in the ceRNA network.

Based on the above findings (Table S1, 2), we constructed the ceRNA network using Cytoscape 3.6. Fig. 3 shows that 4 downregulated miRNAs, 22 upregulated mRNAs and 3 upregulated lncRNAs are involved in one ceRNA network. Meanwhile, 18 upregulated miRNAs, 95 downregulated mRNAs and 20 downregulated lncRNAs were involved in another ceRNA network.

Identification of a prognostic signature of lung cancer

Increasing studies have shown that lncRNAs effectively predict overall survival (OS) in cancer patients [24, 25]. To validate the association between the lncRNAs in the ceRNA network and OS of lung cancer, we randomly divided lung cancer patients into the training set (n = 387) and testing set (n = 386), and there was no significant difference in the clinical covariates between the two sets ($P > 0.05$) (Table S5). We analyzed the 23 DE lncRNAs in the training set by multivariate Cox regression analysis. The results demonstrated that a 6 lncRNA signal was a significant prognostic factor for lung cancer (Table S6). LINC01287, SNAP25-AS1 and LINC00470 were protective, whereas AC104809.2, LINC00645 and LINC01010 were associated with increased risk (Fig. 4a). Next, lung cancer patients in the training set were divided into high-risk and low-risk groups based on the following risk score formula: Risk Score = $(0.0662) * \text{LINC01287} + (0.0978) * \text{SNAP25-AS1} + (0.0567) * \text{LINC00470} + (-0.0745) * \text{AC104809.2} + (-0.2018) * \text{LINC00645} + (-0.0726) * \text{LINC01010}$. Kaplan-Meier analysis revealed that OS in the high-risk group was significantly lower than in the low-risk group ($P=0.0016$, Fig. 4b). The overall ten-year relative survival rates of the high-risk and the low-risk groups were 14.9% and 33%, respectively. Furthermore, the area under ROC curve (AUC) at 10 years for OS was 0.641 (Fig. 4c).

In order to validate the prognostic ability of the 6- lncRNA model, we were further performed Kaplan-Meier analysis and ROC curve analysis in the testing set. Lung cancer patients in the testing set were divided into high-risk and low-risk group (Fig. 5a) with statistically significant different overall survival ($P=0.046$, Fig. 5b) and ROC curve (AUC=0.61, Fig. 5c).

Next, all lung cancer patients were stratified by gender. The six-lncRNA signature could classify 486 male patients into high-risk group and low-risk group and there was the statistically significant difference between the high-risk and the low-risk group in Kaplan-Meier analysis ($p = 0.042$, Fig. 5d). Similarly, 287 female patients were divided into high-risk group and low-risk group with statistically significant different overall survival ($p = 0.027$, Fig. 5e). These results indicate that the 6 lncRNA signal is indeed an independent prognostic factor for predicting OS in lung cancer patients.

Clinical feature analysis of LINC01010

To further analyze whether these 6 lncRNAs affect clinical features in lung cancer from TCGA database, we divided lung cancer patients into high-expression and low-expression groups based on lncRNA expression. Kaplan-Meier survival curves revealed that only LINC01010 was positively correlated with OS ($P = 0.02779$) (Fig. 6a). OS at 5 and 10 years in the LINC01010 high-expression group was 49.9% and 26.8%, respectively, compared with 40.6% and 18.2% in the low-expression group. Therefore, LINC01010 was chosen for further study. Although there was no significant difference between normal tissues and lung cancer tissues (Fig. 6b), expression levels of LINC01010 were downregulated in M1 (metastatic samples) (Fig. 6c) and in stage IV tumors (Fig. 6d) compared to M0 (nonmetastatic samples) and stage I, respectively.

To further investigate the function of LINC01010, we performed co-expression analysis of LINC01010 using the Multi-Experiment Matrix (MEM) resource and selected the top 200 mRNAs co-expressed with LINC01010 for GO and KEGG pathway enrichment analysis (Table S7). The top 15 biological processes identified in GO and KEGG pathways revealed that mRNAs positively co-expressed with LINC01010 participated primarily in “integrin-mediated signaling pathway”, “extracellular matrix organization”, “leukocyte migration”, “regulation of actin cytoskeleton” and “proteoglycans in cancer” (Fig. 6e, f), suggesting that LINC01010 may regulate tumorigenesis and metastasis in lung cancer [26, 27].

LINC01010 represses lung cancer cell migration

To further determine whether LINC01010 affects lung cancer cell migration ability, we knocked down its expression by transfecting siRNA for LINC01010 into A549 or SPC-A-1 cells. qRT-PCR results showed that expression levels of LINC01010 were significantly decreased in three siRNA-transfected cells compared with controls (Fig. 7a, b). SiLINC01010-3 was selected for subsequent functional experiments. Transwell assays showed that suppression of LINC01010 promoted both A549 (Fig. 7c, e) and SPC-A-1 cell (Fig. 7d, f) migration and invasion compared to control cells, indicating that LINC01010 might play an important role in lung cancer metastasis.

LINC01010 associated ceRNA network

The ceRNA networks illustrated that LINC01010 may target hsa-mir-372, hsa-mir-373, hsa-mir-488 and hsa-mir-541. By Pearson correlation analysis, we found that LINC01010 was negatively correlated with hsa-mir-372 (Fig. 8a). By GSEA analysis, the gene set “KRAS.DF.V1_UP” ($P < 0.0001$) was significantly

enriched in high levels of hsa-mir-372 (Fig. 8b), which was also verified by the literature [28, 29]. Meanwhile, the gene set "KRAS.DF.V1_UP" ($P = 0.011$) was significantly enriched in low levels of LINC01010 (Fig. 8c) and "KRAS.LUNG_UP.V1_DN" ($P < 0.0001$) was significantly enriched in high levels of LINC01010 (Fig. 8d), which suggested that LINC01010 might inhibit the function of hsa-mir-372 in the MAPK pathway.

We further investigated whether expression of LINC01010 was positively correlated with hsa-mir-372 target mRNAs. Pearson correlation analysis showed that LINC01010 was positively correlated with GRIA2, HS3ST4, SLC5A7 and TRDN (Fig. 8e). Therefore, LINC01010 may play an important role in tumorigenesis in lung cancer by competing with miR-372.

Discussion

Lung cancer is the most prevalent cancer with the highest incidence and mortality rates [1]. Since most patients present with invasive tumors, it is crucial to understand the molecular mechanisms of lung cancer metastasis. Increasingly, studies have shown that lncRNAs play important roles in the pathogenesis of lung cancer [3-6], which may act as ceRNAs to regulate mRNAs expression [16]. A recent study constructed a ceRNA network between lung adenocarcinoma samples and adjacent nontumor lung tissues samples [30]. However, the role of ceRNA in metastatic lung cancer remains unclear.

In the present study, we constructed a lncRNAs-miRNAs-mRNAs ceRNA network between metastatic lung cancer and nonmetastatic lung cancer patients based on TCGA database. We then divided DE mRNAs into upregulated and downregulated groups and performed GO and KEGG analysis on these DE mRNAs to further investigate pathways involved in metastatic lung cancer. In the upregulated group, GO:0003341~cilium movement and "MAPK signaling pathway" were observed, which have been reported to be associated with lung cancer. Numerous lung diseases are marked by abnormalities in both cilia structure and function [31]. The primary cilium provides a spatially localized platform for signaling by Hedgehog, Notch and WNT, which promotes metastasis in lung cancer [20]. Furthermore, the MAPK pathway is activated in non-small cell lung tumors and plays a key role in migration and invasion of lung cancer [21, 32]. While downregulated mRNAs participated in "GO:0008544~epidermis development" and "drug metabolism - cytochrome P450". Epithelial-mesenchymal transition (EMT) is considered an essential process for metastasis [22]. During this process, epithelial markers, such as keratins, which are primarily clustered in "GO:0008544~epidermis development", are downregulated [23]. Several studies have shown that glutathione S-transferase (GST) polymorphisms increase the risk for various cancers, including lung cancer [33]. Moreover, low expression of GSTM3, which is involved in "drug metabolism - cytochrome P450", correlates with increased susceptibility to lung cancer [34, 35].

To our knowledge, few articles have identified lung cancer-specific lncRNAs as molecular biomarkers for diagnosis and prognosis. Therefore, we performed multivariate Cox regression analysis to analyze the 23 metastasis-associated DE lncRNAs. Result revealed a 6 lncRNA signal, including LINC01287, SNAP25-AS1, LINC00470, AC104809.2, LINC00645 and LINC01010, as an independent prognostic factor for

predicting OS in lung cancer patients. Furthermore, we noted that expression of LINC01010 was positively correlated with OS and negatively correlated with metastasis and lung cancer stage. Therefore, we postulate that LINC01010 may play a more crucial role in the pathogenesis and prognosis of lung cancer. GO and KEGG pathway analyses demonstrated that mRNAs positively co-expressed with LINC01010 participate in “integrin-mediated signaling pathway”, “extracellular matrix organization”, “leukocyte migration”, “regulation of actin cytoskeleton” and “proteoglycans in cancer”, which reportedly participate in tumorigenesis and metastasis of lung cancer [26, 27]. These in silico analysis results were confirmed by our functional experiments. Transwell assays showed that suppression of LINC01010 inhibited lung cancer cell migration and invasion.

Dysregulated miRNA expression is also reported to play crucial roles in the carcinogenesis of lung cancer [36]. In the present study, the ceRNA networks we constructed illustrate that LINC01010 may target hsa-mir-373, hsa-mir-488, hsa-mir-541, and hsa-mir-372. Huang Q et al. demonstrated that miR-373 stimulates cancer cell migration and invasion in vitro and in vivo and that certain cancer cell lines depend on endogenous miR-373 activity to efficiently migrate [37]. Voorhoeve P et al. indicated that either miR-373 or miR-372 prevents RAS-induced growth arrest in primary human cells, possibly through inhibition of expression of the tumor suppressor LATS2 [38]. Moreover, hsa-miR-372 promotes invasive ability of lung cancer cells and globally affects the regulatory circuits centered on MAPK/ERK signaling [28, 29]. By Pearson correlation analysis, we found that expression of LINC01010 negatively correlated with hsa-mir-372 and positively correlated with hsa-mir-372 target genes. By GSEA analysis, the gene set “KRAS.DF.V1_UP” ($P = 0.011$) was significantly enriched in low levels of LINC01010 and “KRAS.LUNG_UP.V1_DN” ($P < 0.0001$) was significantly enriched in high levels of LINC01010. Taken together, these results suggest that LINC01010 might be involved in regulation of the MAPK pathway in lung cancer by competing with miR-372. To gain further insight into the potential role of LINC01010, we will investigate whether LINC01010 sponges miR-372 by binding to its response elements in the future.

Conclusions

We constructed a metastasis-associated ceRNA network in lung cancer and identified a 6 lncRNA signature that is an independent prognostic factor for predicting OS in lung cancer. For the first time, we validated LINC01010 as a tumor suppressor lncRNA in lung cancer through bioinformatics and Transwell assay experiments. Our findings provide insight into the identification of potential lncRNA biomarkers for diagnosis and prognosis in lung cancer.

Abbreviations

mRNA:messenger RNA; miRNA:microRNA; lncRNA:long noncoding RNA; ceRNAs:competing endogenous RNAs; TCGA:the genomic data commons data portal and the cancer genome atlas database; GDC:the genomic data commons data portal; GO:gene ontology database; KEGG:kyoto encyclopedia of genes and genomes; MEM:the multi- experiment matrix resource; DAVID:database for annotation, visualization, and integrated discovery; KOBAS:KEGG orthology based annotation system; GSEA:gene set enrichment

analysis; OS:Overall survival; AIC:Akaike Information Criterion; ROC curve:receiver operating characteristic curve; AUC:area under curve; DE:differentially expressed; FDR:false discovery rate; siLNC01010:a small interfering RNA for LNC01010; siRNA-NC:a negative control.

Declarations

Acknowledgements

Not applicable.

Authors' contributions

LL and QC conceived and designed the experiments. ZWD, SZL, GYA and BBY analyzed data. QC performed the experiment. QC and ZWD wrote the manuscript. LL revised the manuscript. All authors read and approved the final version of this manuscript.

Funding

This work was supported by grants from the Natural Science Foundation of Shaanxi Provincial Department of Education (grant no. 18JK0796), the Shaanxi Province Natural Science Basic Research Program (grant no. 2019JQ-538) and the Opening Foundation of Key Laboratory of Resource Biology and Biotechnology in Western China (Northwest University), Ministry of Education (No.ZSK2018002).

Availability of data and materials

Authors can provide all of datasets analyzed during the study on reasonable request.

Ethics approval and consent to participate

No ethical approval nor informed consent was required in this study due to the public-availability of the data used.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

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. *J CA: a cancer journal for clinicians*. 2018; 68(6):394-424.

2. Torre LA, Siegel RL, Jemal A. Lung Cancer Statistics. *J Advances in Experimental Medicine Biology*. 2016; 893:1-19.
3. Loewen G, Jayawickramarajah J, Zhuo Y, Shan B. Functions of lncRNA HOTAIR in lung cancer. *Journal of hematology & oncology*. 2014; 7(1):90.
4. Yang Y, Zang S, Zhong C, Li Y, Zhao S, Feng X. Increased expression of the lncRNA PVT1 promotes tumorigenesis in non-small cell lung cancer. *International journal of clinical and experimental pathology*. 2014; 7(10):6929-35.
5. Yang J, Lin J, Liu T, Chen T, Pan S, Huang W, Li S. Analysis of lncRNA expression profiles in non-small cell lung cancers (NSCLC) and their clinical subtypes. *Lung cancer*. 2014; 85(2):110-5.
6. Zhang L, Li S, Choi Y, Lee J, Gong Z, Liu X, Pei Y, Jiang A, Ye M, Mao M. Systematic identification of cancer-related long noncoding RNAs and aberrant alternative splicing of quintuple-negative lung adenocarcinoma through RNA-Seq. *Lung cancer*. 2017; 109:21-7.
7. Chen Y, Yu X, Xu Y, Shen H. Identification of dysregulated lncRNAs profiling and metastasis-associated lncRNAs in colorectal cancer by genome-wide analysis. 2017; 6(10):2321-30.
8. Sarah D, Davis CA, Angelika M, Alex D, Timo L, Ali M, Andrea T, Julien L, Wei L, Felix S. Landscape of transcription in human cells. *Nature*. 2012; 489:101-8.
9. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489:57-74.
10. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nature Reviews Genetics*. 2009; 10:155-9.
11. Ponting CP, Oliver PL, Wolf R. Evolution and functions of long noncoding RNAs. *J Cell*. 2009; 136(4):629-41.
12. Igor U, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *J Cell*. 2013; 154(1):26-46.
13. Miao F, Chen J, Shi M, Song Y, Chen Z, Pang L. LncRNA HAND2-AS1 inhibits non-small cell lung cancer migration, invasion and maintains cell stemness through the interactions with TGF- β 1. *Bioscience reports*. 2019; 39(1):BSR20181525.
14. Fang S, Gao H, Tong Y, Yang J, Tang R, Niu Y, Li M, Guo L. Long noncoding RNA-HOTAIR affects chemoresistance by regulating HOXA1 methylation in small cell lung cancer cells. *Laboratory Investigation*. 2016; 96(1):60-8.
15. Yu S-L, Chen H-Y, Chang G-C, Chen C-Y, Chen H-W, Singh S, Cheng C-L, Yu C-J, Lee Y-C, Chen H-S. MicroRNA signature predicts survival and relapse in lung cancer. *J Cancer cell*. 2008; 13(1):48-57.
16. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*. 2011; 146(3):353-8.
17. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *J Bioinformatics*. 2010; 26(1):139-40.
18. Therneau T, Lumley T. Survival: Survival analysis, including penalised likelihood. R package version 2.35-7. J Vienna, Austria: R Foundation for Statistical Computing. 2009.

19. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *J Biometrics*. 2000; 56(2):337-44.
20. Liu H, Kiseleva AA, Golemis EA. Ciliary signalling in cancer. *J Nature Reviews Cancer*. 2018; 18:511-24.
21. Breindel JL, Haskins JW, Cowell EP, Zhao M, Nguyen DX, Stern DF. EGF Receptor Activates MET through MAPK to Enhance Non-Small Cell Lung Carcinoma Invasion and Brain Metastasis. *Cancer research*. 2013; 73(16):5053-65.
22. Thiery JP, Acloque H, Huang RY, Nieto MA. Epithelial-mesenchymal transitions in development and disease. *cell*. 2009; 139(5):871-90.
23. Willipinski-Stapelfeldt B, Riethdorf S, Assmann V, Woelfle U, Rau T, Sauter G, Heukeshoven J, Pantel K. Changes in cytoskeletal protein composition indicative of an epithelial-mesenchymal transition in human micrometastatic and primary breast carcinoma cells. *Clinical Cancer Research*. 2005; 11(22):8006-14.
24. Wang P, Lu S, Mao H, Bai Y, Ma T, Cheng Z, Zhang H, Jin Q, Zhao J, Mao H. Identification of biomarkers for the detection of early stage lung adenocarcinoma by microarray profiling of long noncoding RNAs. *Lung Cancer*. 2015; 88(2):147-53.
25. Xia L, Wang Y, Meng Q, Su X, Shen J, Wang J, He H, Wen B, Zhang C, Xu M. Integrated bioinformatic analysis of a competing endogenous RNA network reveals a prognostic signature in endometrial cancer. *Frontiers in oncology*. 2019; 9:448.
26. Meng X, Jin Y, Yu Y, Bai J, Liu G, Zhu J, Zhao Y, Wang Z, Chen F, Lee K. Characterisation of fibronectin-mediated FAK signalling pathways in lung cancer cell migration and invasion. *British journal of cancer*. 2009; 101(2):327-34.
27. Yamaguchi H, Condeelis J. Regulation of the actin cytoskeleton in cancer cell migration and invasion. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*. 2007; 1773(5):642-52.
28. Yu S-L, Chen H-Y, Chang G-C, Chen C-Y, Chen H-W, Singh S, Cheng C-L, Yu C-J, Lee Y-C, Chen H-S. MicroRNA signature predicts survival and relapse in lung cancer. *Cancer cell*. 2008; 13(1):48-57.
29. Ragusa M, Statello L, Maugeri M, Majorana A, Barbagallo D, Salito L, Sammito M, Santonocito M, Angelica R, Cavallaro A. Specific alterations of the microRNA transcriptome and global network structure in colorectal cancer after treatment with MAPK/ERK inhibitors. *Journal of molecular medicine*. 2012; 90(12):1421-38.
30. Sui J, Li Y, Zhang Y, Li C, Shen X, Yao W, Peng H, Hong W, Yin L, Pu Y. Integrated analysis of long non-coding RNA-associated ceRNA network reveals potential lncRNA biomarkers in human lung adenocarcinoma. *International journal of oncology*. 2016; 49(5):2023-36.
31. Tilley AE, Walters MS, Shaykhiev R, Crystal RG. Cilia dysfunction in lung disease. *Annual review of physiology*. 2015; 77:379-406.
32. Greenberg AK, Basu S, Hu J, Yie T-a, Tchou-Wong KM, Rom WN, Lee TC. Selective p38 activation in human non-small cell lung cancer. *Am J Respir Cell Mol Biol*. 2002; 26(5):558-64.

33. Balendiran GK, Dabur R, Fraser D. The role of glutathione in cancer. *Cell Biochemistry and Function: Cellular biochemistry and its modulation by active agents or disease*. 2004; 22(6):343-52.
34. Nakajima T, Elovaara E, Anttila S, Hirvonen A, Camus A-M, Hayes JD, Ketterer B, Vainio H. Expression and polymorphism of glutathione S-transferase in human lungs: risk factors in smoking-related lung cancer. *Carcinogenesis*. 1995; 16(4):707-11.
35. Inskip A, Elexperu-Camiruaga J, Buxton N, Dias PS, MacIntosh J, Campbell D, Jones P, Yengi L, Talbot J, Strange R. Identification of polymorphism at the glutathione S-transferase, GSTM3 locus: evidence for linkage with GSTM1* A. *Biochem J*. 1995; 312(3):713-6.
36. Xie Y, Todd NW, Liu Z, Zhan M, Fang H, Peng H, Alattar M, Deepak J, Stass SA, Jiang F. Altered miRNA expression in sputum for diagnosis of non-small cell lung cancer. *Lung cancer*. 2010; 67(2):170-6.
37. Huang Q, Gumireddy K, Schrier M, Le Sage C, Nagel R, Nair S, Egan DA, Li A, Huang G, Klein-Szanto AJ. The microRNAs miR-373 and miR-520c promote tumour invasion and metastasis. *J Nature cell biology*. 2008; 10(2):202-10.
38. Voorhoeve PM, Le Sage C, Schrier M, Gillis AJ, Stoop H, Nagel R, Liu Y-P, Van Duijse J, Drost J, Griekspoor A. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell*. 2006; 124(6):1169-81.

Figures

Figure 1

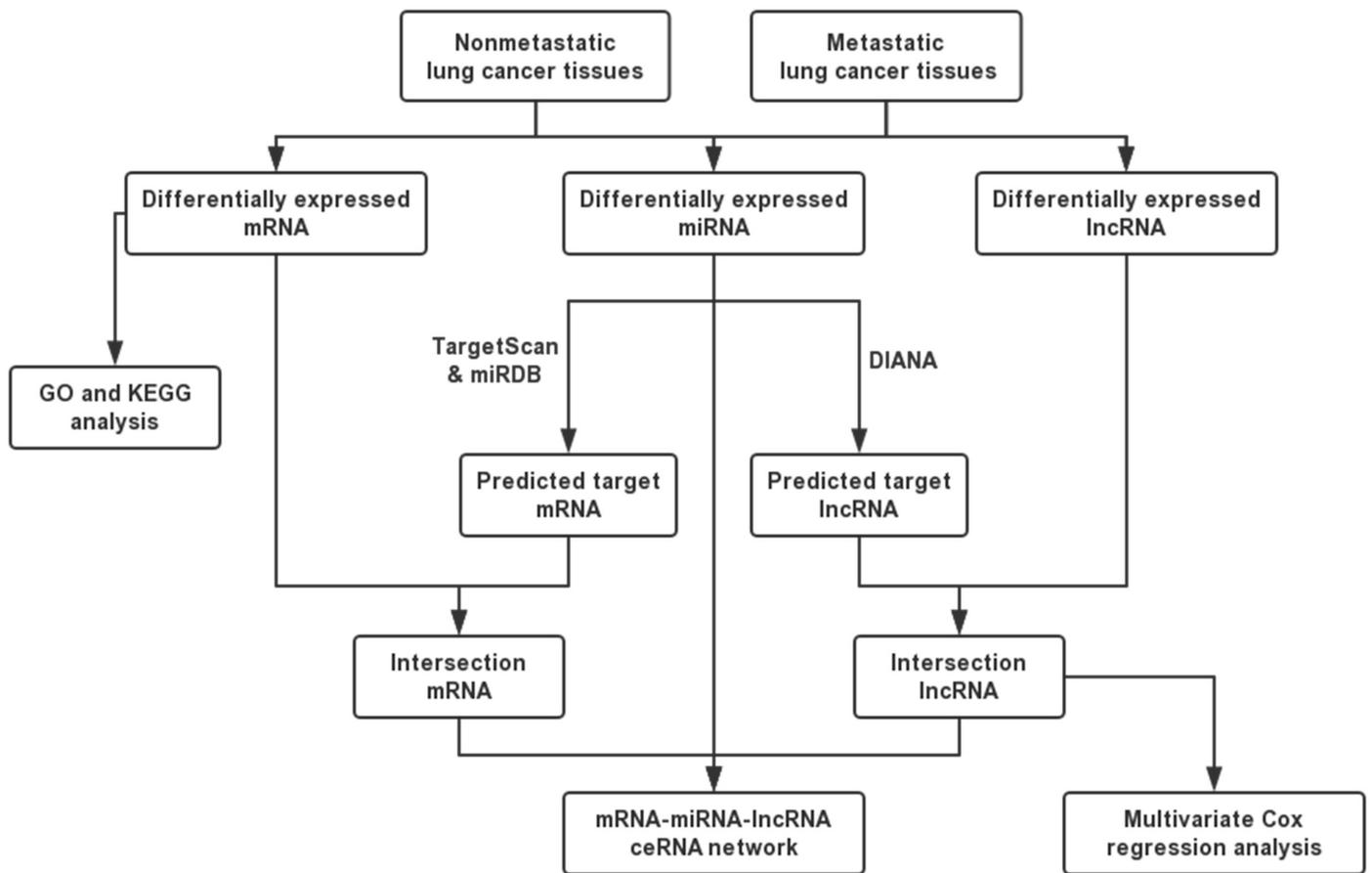


Figure 1

Flow chart of construction of the metastasis-associated ceRNA network in lung cancer.

Figure 2

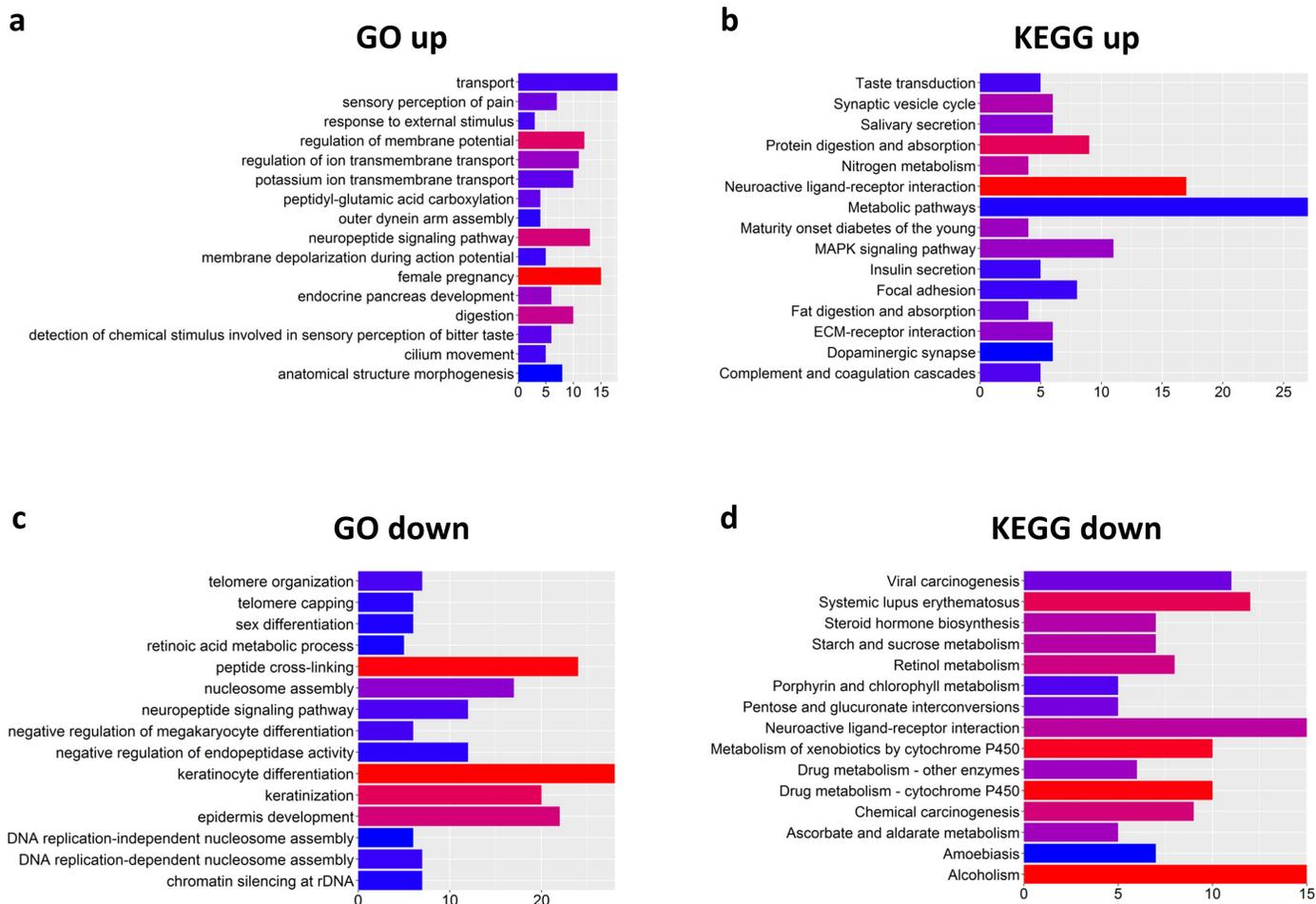


Figure 2

The functions of DE mRNAs between nonmetastatic and metastatic lung cancer tissues. a, b Top 15 GO and KEGG pathways of upregulated mRNA, respectively. c, d Top 15 GO and KEGG pathways of downregulated mRNA, respectively. Color from blue to red indicates $-\log_{10}(P\text{-value})$ from low to high. X-axes represent the number of genes involved in each pathway.

Figure 4

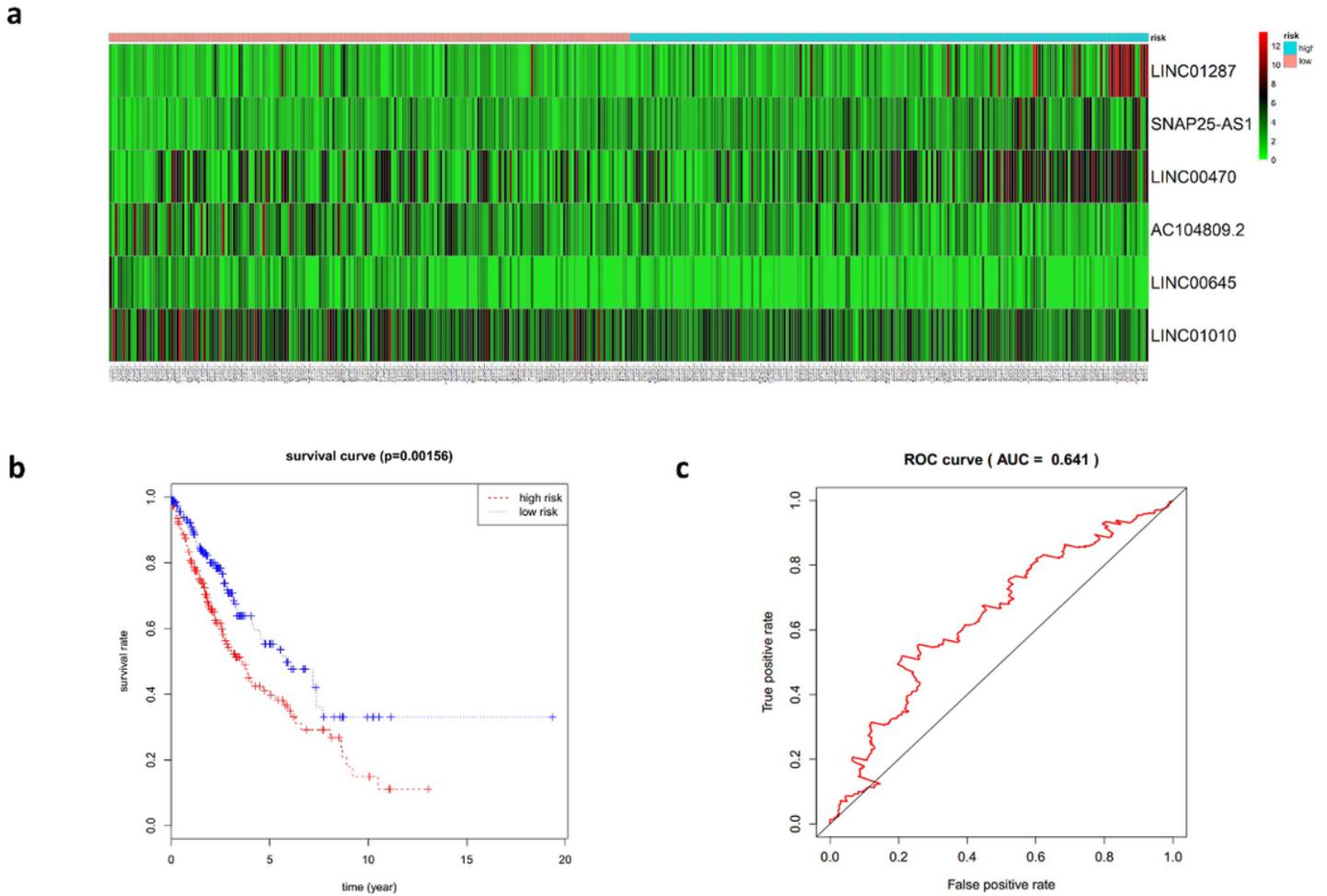


Figure 4

Evaluation of the 6 lncRNA signature in the training set. a Lung cancer patients in the training set were divided into high- and low-risk groups based on their risk scores generated from the 6 lncRNA signature. The expression heat map shows the expression profiles of the 6 lncRNAs in lung cancer patients. b Kaplan-Meier survival curve analysis for overall survival of lung cancer patients using the risk scores generated from the 6 lncRNA signature. Differences between high-risk (n= 194) and low-risk (n= 193) groups were determined by log-rank test (P = 0.016). c Validation of the 6 lncRNA signature by ROC curve for predicting 10-year survival in lung cancer patients based on risk scores.

Figure 5

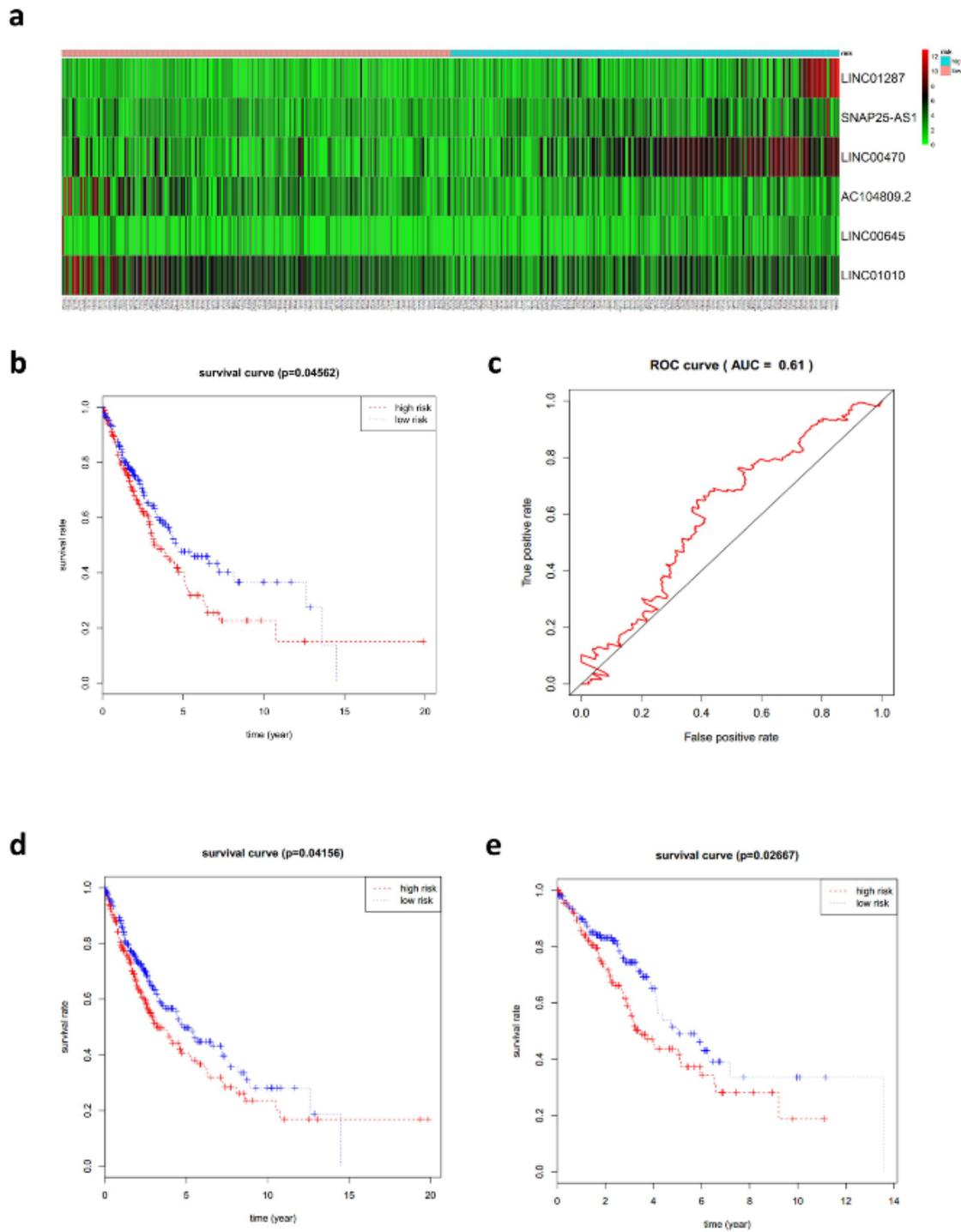


Figure 5

Prognostic evaluation of the 6 lncRNA signature in the testing set and stratification analysis by gender. a Lung cancer patients in the testing set were divided into high- and low-risk groups based on their risk scores generated from the 6 lncRNA signature. The expression heat map shows the expression profiles of the 6 lncRNAs in lung cancer patients. b Kaplan-Meier survival curve analysis for overall survival of lung cancer patients using the risk scores generated from the 6 lncRNA signature. Differences between high-

risk (n= 193) and low-risk (n= 193) groups were determined by log-rank test (P = 0.046). c Validation of the 6 lncRNA signature by ROC curve for predicting 10-year survival in lung cancer patients based on risk scores; Kaplan-Meier survival curve analysis of overall survival in high-risk and low-risk groups for male patients d and female patients e.

Figure 6

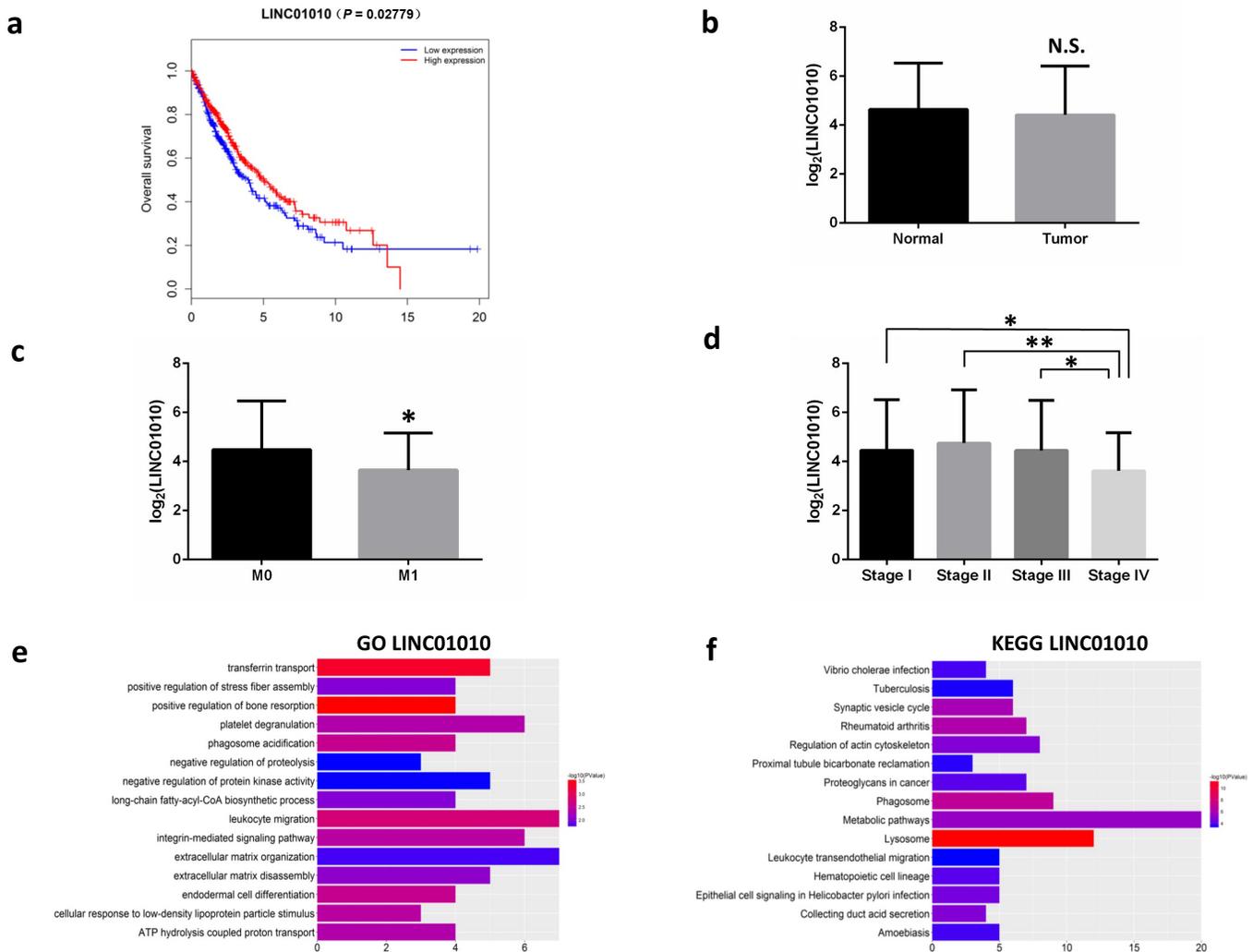


Figure 6

Clinical feature analysis of LINC01010. a Kaplan-Meier analysis of lung cancer overall survival according to expression of LINC01010. b LINC01010 expression analysis in lung cancer tissues and normal tissues. c LINC01010 expression analysis in metastatic lung cancer tissues (M1) and nonmetastatic tissues (M0). d LINC01010 expression analysis in stage I-IV lung cancer. e,f Top 15 GO and KEGG pathways of genes co-expressed with LINC01010. Color from blue to red indicates $-\log_{10}(P\text{-value})$ from low to high. X-axes represent the number of genes involved in each pathway. N.S. not significant, * P < 0.05, ** P < 0.01.

Figure 7

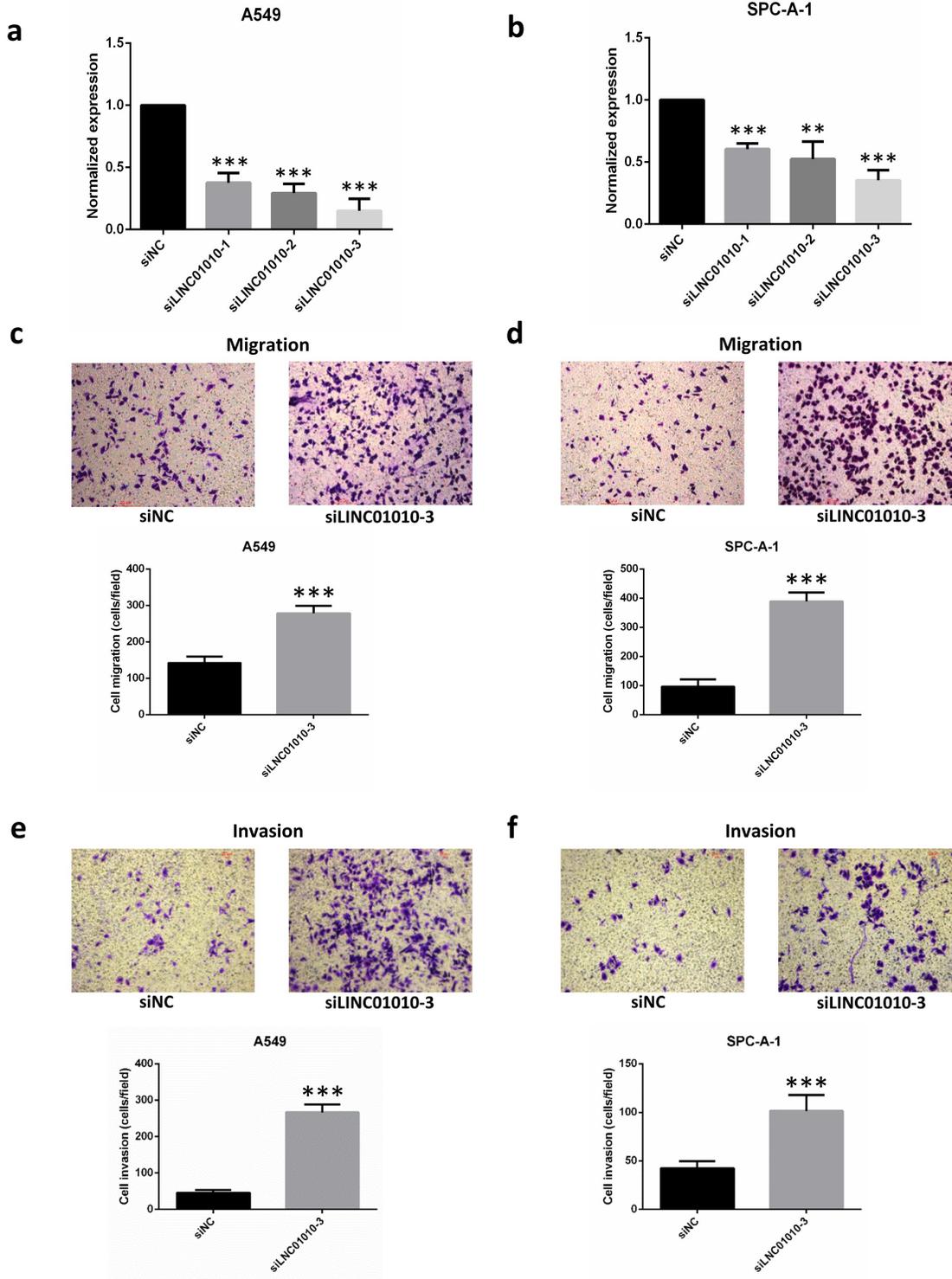


Figure 7

LINC01010 represses lung cancer cell migration. LINC01010 expression levels were determined in siNC and siLINC01010 transfected A549 a or SPC-A-1 b cells by qRT-PCR. Expression levels of LINC01010 were normalized to GAPDH. Transwell assays were used to determine the migration and invasive abilities of siNC and siLINC01010 transfected A549 c,e and SPC-A-1 d,f cells. All data are represented as the mean \pm SD (n=3), ** P < 0.01, *** P < 0.001.

Figure 8

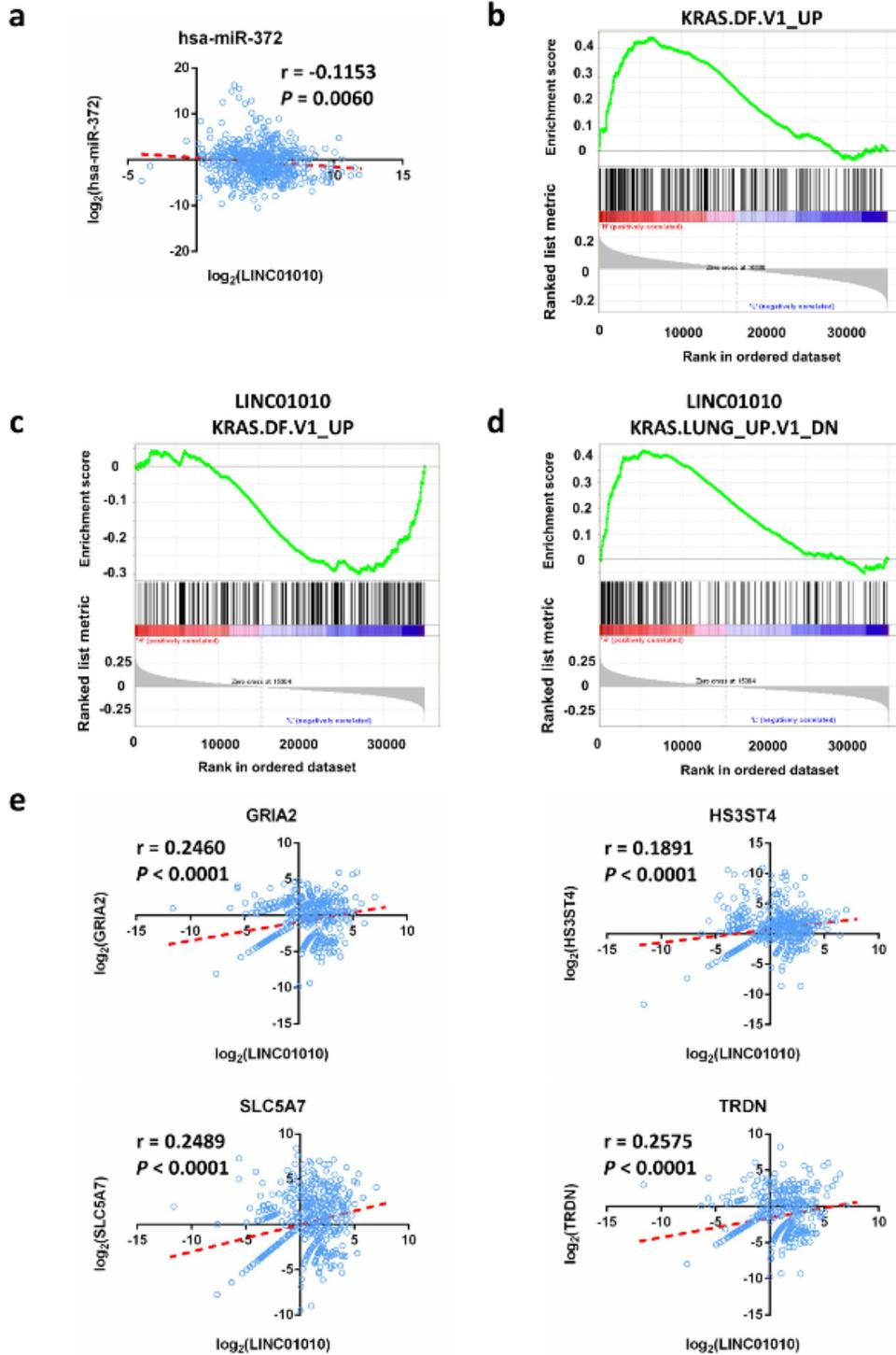


Figure 8

LINC01010 is negatively correlated with miR-372. a Pearson correlation analysis shows that LINC01010 is negatively correlated with hsa-mir-372 ($r = -0.1153$, $P = 0.0060$). b GSEA results showed a correlation between hsa-mir-372 levels and the MAPK signaling pathway in lung cancer. The gene set “KRAS.DF.V1_UP” ($P < 0.0001$) was significantly enriched in high levels of hsa-mir-372. c the gene set “KRAS.DF.V1_UP” ($P = 0.011$) was significantly enriched in low levels of LINC01010. d

"KRAS.LUNG_UP.V1_DN" ($P < 0.0001$) was significantly enriched in high levels of LINC01010. e Pearson correlation analysis showed that LINC01010 is positively correlated with GRIA2 ($r = 0.246$, $P < 0.0001$), HS3ST4 ($r = 0.1891$, $P < 0.0001$), SLC5A7 ($r = 0.2489$, $P < 0.0001$) and TRDN ($r = 0.2575$, $P < 0.0001$).

Supplementary Files

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

- [supptable.docx](#)