Received: 2019.12.19
Accepted: 2020.04.01 Available online: 2020.05 .18

Published: 2020.07.13

## Prediction of Specific Subtypes and Common Markers of Non-Small Cell Lung Cancer Based on Competing Endogenous RNA Network

Authors' Contribution: Study Design A Data Collection B Statistical Analysis C Data Interpretation D Manuscript Preparation E Literature Search F Funds Collection G

Abcef 1 Yao Liu*
cdef 2 Hao Wang*
afg 3 Wenhan Yang
afg 1 Youhui Qian

1 Department of Thoracic Surgery, Shenzhen Second People's Hospital, Shenzhen, Guangdong, P.R. China
2 Department of Thoracic Surgery, Mingzhou Hospital, Ningbo, Zhejiang, P.R. China 3 School of Basic Medicine, Shenzhen University Health Science Center, Shenzhen, Guangdong, P.R. China

## Corresponding Author:

Source of support:

* Yao Liu and Hao Wang contributed equally to this work as co-first authors

You-hui Qian, e-mail: 13602585160@163.com
This work was supported by the Shenzhen Science, Technology, and Innovation Commission (No. KQCX2015032416521307/ JCYJ20170818100842319)

Background: There are various pathological types of lung cancer, including squamous cell carcinoma and adenocarcinoma. Although both of them are lung cancers, there are significant differences in diagnosis, pathogenesis, location, imaging, metastasis, and treatment. According to the competing endogenous RNA (ceRNA) theory, long noncoding RNAs (IncRNAs) compete with encoding protein genes (mRNAs) to connect with miRNAs, thus affecting the level of mRNA.

Material/Methods:

Conclusions:

## MeSH Keywords:

Full-text PDF:

First, using the $t$ test, we identified mRNAs and IncRNAs that have different expressions (fold change $>2, \mathrm{P}<0.01$ ) in normal samples and in tumor samples. We calculated the significance of the shared miRNAs for mRNAs and IncRNAs by hypergeometric test ( $\mathrm{P}<0.01$ ). Further, mRNA and IncRNA pairs with co-expression relationships in cancer samples were used to establish ceRNA networks. Then, the random walk algorithm was used to optimize the specific ceRNA networks and identify potential prognostic markers of survival. Finally, we built a common ceRNA network to find markers of non-small-cell lung cancer.
Results: We identified some potential key markers, such as PVT1, LINC00472, CDKN2A, and FAM83B, in LUSC and HOXA11-AS, HNF1A-AS1, LINC00511, and HOTAIR in LUAD by analyzing the ceRNA networks. Moreover, a number of common ceRNA pairs, such as CDC25C/CDK1/RRM2-LINC00355, have been found, and they are also significant markers for tumor survival and prognosis.

In summary, the present study provides a comparative analysis in 2 kinds of lung cancer ceRNA networks. Some specific and common markers we predicted that may be of great importance for clinical diagnosis and treatment.

Adenocarcinoma•Lung Neoplasms•Neoplasms, Squamous Cell•RNA Stability• RNA, Long Noncoding • Tumor Markers, Biological
https://www.medscimonit.com/abstract/index/idArt/922280


## Background

Lung cancer is a common cancer worldwide, with high morbidity and mortality [1]; $85 \%$ of lung cancer is non-small-cell lung cancer (NSCLC) [2,3]. Although significant progress has been made in treatment, such as chemotherapy and radiation therapy, the 5 -year survival rate for NSCLC is still $15 \%$ [4]. Therefore, it is very important to understand the molecular mechanism of non-small cell lung cancer and to seek new therapeutic strategies. In addition, regional or remote transfers result in higher mortality [5]. Long non-coding RNAs (IncRNA) are a class of single-stranded RNA molecules with a length of more than 200 nt , located in the nucleus or cytoplasm and lacking pro-tein-coding functions [6,7]. The expression of genes is regulated at many levels in the form of RNA (e.g., epigenetics, transcriptional regulation, and post-transcriptional regulation) [8,9]. IncRNA is only expressed at specific developmental stages, and it has tissue- or cell-specificity. Research on CeRNA (competing endogenous RNAs) and internal competitive RNAs reveals a new mechanism of interaction between RNAs. It is known that microRNAs can cause gene silencing by binding mRNAs, while ceRNAs can regulate gene expression by competitively competitive binding to microRNAs [5]. Recent studies have reported that IncRNAs are associated with other RNA transcripts by competing for internal RNAs (ceRNAs) mechanisms [10-12]. mRNAs compete with IncRNAs for miRNA molecules to alleviate the inhibition of miRNA targets [13-15]. This type of In-cRNA-related competitive triplet is a subclass of ceRNA and occurs widely in humans and several other species [16-18]. Therefore, this study constructed and analyzed the specific ceRNA network of lung adenocarcinoma and squamous cell carcinoma. By constructing the specific ceRNA network and universal network of lung cancer, we found some important markers of cancer that could be used as a guide and reference for the further study of lung cancer.

## Material and Methods

## Identification of differential expression genes

We downloaded the gene expression data (FPKM) of LUAD and LUSC from TCGA project. First, we used PCA method to extract the classification feature information of 2 kinds of cancer expression profile data, and greatly reduced the dimension of expression profile data and visualized the classification features while maintaining high classification accuracy. Differentially expressed genes of LUAD and LUSC were obtained by unpaired $t$ test and we kept the genes satisfying |Log 2 FoldChange| $>1$ and $P<0.01$. Finally, we input the list of upregulated and downregulated genes into the R package "clusterProfiler" for Go term and pathway enrichment analysis.

## Building the specific DE-ceRNA network

The miRNA targets were obtained from StarBase2.0, TargetScan, mirTarbase, LncBase, and MiRcode databases. mRNA and IncRNA pairs with more than zero shared miRNAs and a significant hypergeometric test $P$ value ( $P<0.01$ ) were retained. In total, we integrated more than human 800000 IncRNA-PCG interactions. Further, we used expression profiles of lung cancer to calculate the expression correlations of the differential mRNA-IncRNA pairs by Pearson correlation analysis. Finally, we defined the consistent expression pairs as potential ceRNA relationships (cor >0; P<0.01). We used Cytoscape software to analyze the topological properties of the network.

## Optimizing the DE-ceRNA network and building the sub-net

The random walk (RW) algorithm, which is based on a network, was used to represent an irregular form of change and some of the interesting nodes in the network were weighted as seed nodes. A weight value is distributed to other neighbor nodes along the network structure from the seed node. At the same time, the distribution of the weight of other nodes is also obtained in the walking process. Finally, a node in close contact with the seed node tends to have a higher weight. Here is a restart-type random walk with restart, with the formula is defined as:

$$
P^{t+1}=(1-r) W p^{t}+r p^{0}
$$

where $r$ represents the weight of a node in each migration with the probability assigned to the neighbor node, where the default value of 0.7 , $w$ is used as a standardized adjacency matrix representing the network, and $p 0$ means the initial weight of the node. The weight vector, $\mathrm{p} t$, represents the new weight of the node after $t$-time migration of the network. We obtained gene-disease associations from DisGeNET, Genetic Association Database, Lnc2Cancer, and LncRNADisease. Then, we mapped these known pathogenic genes of lung cancer to the ceRNA network as seed nodes. We used the random walk algorithm according to the topological structure of the ceRNA network to score each node. Finally, we selected the genes with the top random walk score ( 10 times the number of seed nodes for each cancer) as the nodes to extract the sub-net from the ceRNA network. We used Cytoscape software to visualize the sub-network.

## Building a common CeRNA network between LUAD and LUSC

To obtain the common markers of lung cancer, we extracted the same ceRNA relationship between squamous cell carcinoma and adenocarcinoma and we defined this as a common ceRNA network. We extracted the simultaneous IncRNA-mRNA


Figure 1. Classification of samples by lung cancer expression profile. (A) LUSC mRNA expression profile. (B) LUSC IncRNA expression profile. (C) LUAD mRNA expression profile. (D) LUAD IncRNA expression profile.
pairs in the 2 cancers to build a common CeRNA network. Finally, we input the genes list of each module into the DAVID (https://david.ncifcrf.gov/) website for Go term and pathway enrichment analysis.

## Survival analysis

We downloaded gene expression and clinical data from the TCGA project. The relationship between gene expression and prognosis of patients was assessed by Kaplan-Meier analysis. The mean gene expression was used as the cut-off to classify patients into high- and low-risk groups. For a gene in the ceRNA
network, we calculated an average expression to evaluate the survival relation of the gene using the log-rank test ( $\mathrm{P}<0.05$ ). We used R package ("Survival") to evaluate the differences between different groups in the survival analysis experiments.

## Results

## The functions of differentially expressed genes

We used PCA to reduce the expression profile of the 2 cancers (LUAD, LUSC) and showed the classification characteristics,


Figure 2. Enrichment analysis of differentially expressed genes. (A) Differentially expressed mRNAs for LUSC. (B) Differentially expressed IncRNAs for LUSC. (C) Differentially expressed mRNAs for LUAD. (D) Differentially expressed IncRNAs for LUAD. (E) Differentially expressed mRNAs enrichment analysis for LUSC. (F) Differentially expressed mRNAs enrichment analysis for LUAD.
finding that both sets of data maintain high classification accuracy at the same time (Figure 1). After the data quality evaluation, we used the non-paired $t$ test method to find genes that are differentially expressed (|Log 2 FoldChange| $>1$ and $\mathrm{P}<0.01$ ) between normal and cancer samples (Figure 2A-2D). The statistical significance P_value from low to high sequencing of 1000 differentially expressed genes have been annotated on some important functional GO terms cell multiplication (Figure 2E, 2F) such as "cell division", "DNA replication", "G1/S transition of mitotic cell cycle", "regulation of cell cycle", "sister chromatid cohesion", "spindle organization", "mitotic sister chromatid segregation", "G2/M transition of mitotic cell cycle", and pathways associated with cancer such as "cell cycle". These differences in mRNAs and IncRNAs were used in the following analyses.

## Lung cancer-specific DE-ceRNA network analysis

We used the shared miRNA and expression correlation between mRNAs and IncRNAs to build the specific DE-ceRNA network of 2 kinds of lung cancer (see Materials and Methods section above). To retain the highly expressed genes, we deleted the nodes in the network with an average expression of less than 0.1 in all samples. The LUSC DE-ceRNA network contains 16436 edges, 1628 IncRNAs nodes, and 1580 mRNAs nodes. The LUAD DE-ceRNA network contains 9953 edges, 1475 IncRNAs nodes, and 1011 mRNAs nodes. By analyzing the topological properties of the network with Cytoscape software, we found that only 1 frequency of shared neighbor node was the highest, with increasing frequency of shared neighbors (Figure 3A, 3E). The distribution of the node degrees of the 2 networks follow the power law distribution (Figure 3B, 3F). Most of the nodes have a small number of connections,


Figure 3. Topology attribute of specific DE-ceRNA network. (A-D) Attributes of DE-ceRNA network for LUSC. (E-H) Attributes of DE-ceRNA network for LUAD.


Figure 4. Visualization of optimized DE-ceRNA Network. (A) Optimized DE-ceRNA Network for LUSC. (B) Optimized DE-ceRNA Network for LUAD. The border colour of the seed nodes is red.


Figure 5. Visualization of sub-DE-ceRNA Network. (A) Sub-DE-ceRNA Network for LUSC. (B) Sub-DE-ceRNA Network for LUAD. The border colour of the seed nodes is red.


Figure 6. KM survival curve of marker genes. (A) KM survival curves of IncRNAs and mRNAs for LUSC. (B) KM survival curves of IncRNAs and mRNAs for LUAD.
and very few nodes have a very large number of connections, indicating that the network construction is a non-random process. The closeness centrality and betweenness centrality of the network show a positive correlation with the number of shared neighbors (Figure 3C, 3D, 3G, 3H). Through the topology analysis of the network, we determined that the network has strong connectivity, and since the size of the network is large, the next step is to optimize the network.

## Optimizing the DE-ceRNA network and finding marker genes

To optimize the network, we defined known pathogenic genes obtained from 4 gene-associated disease databases as seed
nodes to score all nodes in the network by random walk [19]. Then, we reinvested the top genes with random walk scores from high to low ranking into the DE-ceRNA network to get edge information and build a sub-network. Each sub-network is automatically divided into up (orange color) and down (blue color) parts (Figure 4A, 4B). There are nodes in the network that are significantly higher than other nodes, and the contribution of these nodes to the network may be greater. Therefore, we extracted 1 step of neighbor nodes of 2 networks with degree of nodes greater than 20 and maintained the original connection. These nodes that participate in multiple connections in the network are defined as markers such as oncogene PVT1 [20,21], LINC00472 [22], multiple tumor suppressor CDKN2A [23], prognostic Biomarkers in NSCLC


Figure 7. Common ceRNA network for non-small cell lung cancer. (A) The intersection of the ceRNA pairs, IncRNAs, or mRNAs for 2 kinds of non-small cell lung cancer. (B) Visualization of common ceRNA network.

FAM83B [24], cancer susceptibility gene BRCA2 in LUSC and HOXA11-AS [25], HNF1A-AS1 [26], LINC00511 [27,28], and HOTAIR [29-32] in LUAD (Figure 4A, 4B). There are many genes in this network that are significantly associated with survival of patients with 2 types of cancer (Figure 5). The expression of the genes PVT1, KC877982.1, LINC01133, MAGI2-AS3, DSCR9, SFTA1P, LINC00968, and SOX2-OTcan significantly distinguish the survival rate of LUSC patients (Figure 6A), and the expression of the genes AC064807.2, LINC00987, LINC01238, TMPOAS1, OGFRP1, DRAIC, AC023421.1, and RNF144A-AS1 can significantly distinguish the survival rate of LUAD patients (Figure 6B).

## Common ceRNA network analysis

To find common markers, we obtained the intersection of IncRNA-mRNA pairs between 2 cancers to build a common ceRNA network. The network contains 2680 IncRNA-mRNA pairs, 633 IncRNAs and 596 mRNAs (Figure 7A, 7B). Genes that occur at high frequencies in 2 types of cancer are thought to be
more important in the treatment and detection of lung cancer. In the common network, mRNAs are also enriched in important pathways and functional terms of cancer development, such as "hsa04110: Cell cycle", "hsa04115: p53 signaling pathway", "GO: 0008283~cell proliferation", "GO: 0051301~cell division", "h_rbPathway: RB Tumor Suppressor/Checkpoint Signaling in response to DNA damage", and "GAD_DISEASE: lung cancer" (Figure 8A). We selected cell cycle and p53 pathway modules to display in this common ceRNA network (Figure 8B). For the cell cycle pathway module, we speculated that IncRNAs connected by the coding genes (CDC6, CDK1, E2F2, PKMYT1, TTK, CHEK1, ESPL1, PTTG1, CDC25C, CDC25A, CCNB1, CCNE2, MAD2L1, CCNB2, PLK1, BUB1, BUB1B, ORC6, ORC1, CCNA2) in the pathway may be an important molecule that affects the cell cycle and the increased number of cancer cells, and they exist in an endogenous competitive relationship (CDC25CLINC00355; ORC6-DDX11-AS1; BUB1/CCNA2/PLK1-VPS9D1AS1; ORC1/ORC6-LINC00337). In the same way, IncRNA connected by the coding genes (CCNB1, CCNE2, CDK1, CCNB2,


Figure 8. Common markers for non-small cell lung cancer. (A) Enrichment analysis of mRNAs in common ceRNA network. (B) Network modules related to cancer-related pathway.

SERPINB5, RRM2, CHEK1, GTSE1) may also be an important molecule in the P53 cancer pathway. These close regulatory relationships are more likely to be pathogenic factors of NSCLC.

## Discussion

The CeRNA mechanism plays an important role in the development of lung cancer. For example, IncRNA PVT1 regulates expression of HIF1 $\alpha$ via functioning as ceRNA for miR-199a-5p in non-small cell lung cancer under hypoxia. We identified some cancer markers by building and analyzing specific or common ceRNA networks [9,33,34]. The specific treatment of 2 subtypes of non-small cell lung cancer is a problem that
both researchers and clinicians want to solve, but so far there is no good explanation or evidence [35-37]. We constructed 2 specific networks with mRNA and IncRNA using the specificity of IncRNA to find markers and survival-related genes of 2 different cancers. The representative markers we found in the LUSC network and LUAD network are PVT1, HOTAIR, FAM83B, LINC00472, HOXA11-AS, and CDKN2A, which have also been reported as potential risk factors for malignant tumors such as gastric cancer, glioma, and cervical cancer [20-22,24,25,28-31]. In addition, we analyzed the influence of gene expression on patient survival by combining the clinical information of lung cancer patients. We screened a few genes related to lung cancer patient survival and prognosis, such as PVT1, LINC01133, TMPO-AS1, and DRAIC.

To predict some of the markers in both cancers, we also built a common ceRNA network, and we selected 2 important cancer pathways - Cell cycle and P53 - to analyze the subnetworks associated with them. We found that the CDC25C/CDK1/RRM2LINC00355 ceRNA relationship can affect non-small cell lung cancer. It is reported in recent studies that exosomes-mediated LINC00355 regulates bladder cancer cell proliferation and invasion [38]. Identification of these markers will further improve understanding of IncRNA-mediated pathway regulation and ceRNA functions. Our study validated the conclusions of some previous experiments and also produced novel and innovative results awaiting verification by biologists. Continued investigation of the ceRNA networks analyzed here will aid in the development of targeted therapy for lung cancer. We will plan to collect more cancer expression data and clinical information to identify molecular markers more precisely using our network-based approach in future studies.

## Conclusions

Accumulating evidence suggests that IncRNAs play important roles in lung cancer. However, only a few specific biomarkers have been identified in subtypes of non-small cell lung cancers so far. In this study, the network analysis and functional genes prediction of non-small cell lung cancer were assessed from 2 aspects - specificity and universality. These results suggest that our method identifies different cancer subtypes with remarkable molecular and cancer-related pathways, which would help to improve personalized cancer management. Our results provide a basis and evidence for future research and clinical treatment, and our experimental data will be of value to researchers.

## Availability of data and materials

The datasets used and analyzed in this study are available from the corresponding author on reasonable request.

## Conflict of interests

None.
14. Zhang L, Fang F, He X: Long noncoding RNA TP73-AS1 promotes non-small cell lung cancer progression by competitively sponging miR-449a/EZH2. Biomed Pharmacother, 2018; 104: 705-11
15. Ning P, Wu Z, Hu A et al: Integrated genomic analyses of lung squamous cell carcinoma for identification of a possible competitive endogenous RNA network by means of TCGA datasets. Peer J, 2018; 6: e4254
16. Salmena L, Poliseno L, Tay Y et al: A ceRNA hypothesis: The Rosetta Stone of a hidden RNA language? Cell, 2011; 146: 353-58
17. Panzitt K, Tschernatsch MM, Guelly C et al: Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. Gastroenterology, 2007; 132: 330-42
18. Wang J, Liu $X$, Wu H et al: CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Res, 2010; 38: 5366-83
19. Abdullahi A, Shohaimi S, Kilicman A, Ibrahim MH: Stochastic models in seed dispersals: random walks and birth-death processes. J Biol Dyn, 2019; 13: 345-61
20. Wang Y, Chen W, Lian J et al: The IncRNA PVT1 regulates nasopharyngeal carcinoma cell proliferation via activating the KAT2A acetyltransferase and stabilizing HIF-1alpha. Cell Death Differ, 2020; 27(2): 695-710
21. Wang C, Han C, Zhang Y, Liu F: LncRNA PVT1 regulate expression of HIF1alpha via functioning as ceRNA for miR199a5p in nonsmall cell lung cancer under hypoxia. Mol Med Rep, 2018; 17: 1105-10
22. Chen L, Zhang W, Li DY et al: Regulatory network analysis of LINC00472, a long noncoding RNA downregulated by DNA hypermethylation in colorectal cancer. Clin Genet, 2018; 93: 1189-98
23. Zeng H, Jorapur A, Shain AH et al: Bi-allelic Loss of CDKN2A initiates melanoma invasion via BRN2 activation. Cancer Cell, 2018; 34: 56-68.e9
24. Richtmann S, Wilkens D, Warth A et al: FAM83A and FAM83B as prognostic biomarkers and potential new therapeutic targets in NSCLC. Cancers (Basel), 2019; 11(5): pii: E652
25. Sun M, Nie F, Wang Y et al: LncRNA HOXA11-AS promotes proliferation and invasion of gastric cancer by scaffolding the chromatin modification factors PRC2, LSD1, and DNMT1. Cancer Res, 2016; 76: 6299-310
26. Liu L, Chen Y, Li Q, Duan P: IncRNA HNF1A-AS1 modulates non-small cell lung cancer progression by targeting miR-149-5p/Cdk6. J Cell Biochem, 2019; 120: 18736-50
27. Ding J, Yang C, Yang S: LINC00511 interacts with miR-765 and modulates tongue squamous cell carcinoma progression by targeting LAMC2. J Oral Pathol Med, 2018; 47: 468-76
28. Yan L, Wu X, Liu Y, Xian W: LncRNA Linc00511 promotes osteosarcoma cell proliferation and migration through sponging miR-765. J Cell Biochem, 2018 [Epub ahead of print]
29. Li J, Li S, Chen Z et al: miR-326 reverses chemoresistance in human lung adenocarcinoma cells by targeting specificity protein 1 . Tumour Biol, 2016; 37: 13287-94
30. Loewen G, Jayawickramarajah J, Zhuo Y, Shan B: Functions of IncRNA HOTAIR in lung cancer. J Hematol Oncol, 2014; 7: 90
31. Malissovas N, Ninou E, Michail A, Politis PK: Targeting long non-coding RNAs in nervous system cancers: New insights in prognosis, diagnosis and therapy. Curr Med Chem, 2019; 26: 5649-63
32. Zhuang Y, Wang X, Nguyen HT et al: Induction of long intergenic non-coding RNA HOTAIR in lung cancer cells by type I collagen. J Hematol Oncol, 2013; 6: 35
33. Shao T, Wu A, Chen J et al: Identification of module biomarkers from the dysregulated ceRNA-ceRNA interaction network in lung adenocarcinoma. Mol Biosyst, 2015; 11: 3048-58
34. Sui J, Xu SY, Han J et al: Integrated analysis of competing endogenous RNA network revealing IncRNAs as potential prognostic biomarkers in human lung squamous cell carcinoma. Oncotarget, 2017; 8: 65997-6018
35. Zhu TG, Xiao X, Wei $Q$ et al: Revealing potential long non-coding RNA biomarkers in lung adenocarcinoma using long non-coding RNA-mediated competitive endogenous RNA network. Braz J Med Biol Res, 2017; 50: e6297
36. Li J, Tian H, Pan J et al: Pecanex functions as a competitive endogenous RNA of S-phase kinase associated protein 2 in lung cancer. Cancer Lett, 2017; 406: 36-46
37. Sui J, Li YH, Zhang YQ et al: Integrated analysis of long non-coding RNAassociated ceRNA network reveals potential IncRNA biomarkers in human lung adenocarcinoma. Int J Oncol, 2016; 49: 2023-36
38. Yan L, Wang P, Fang W, Liang C: Cancer-associated fibroblasts-derived exo-somes-mediated transfer of LINC00355 regulates bladder cancer cell proliferation and invasion. Cell Biochem Funct, 2020; 38(3): 257-65

