The construction and analysis of the aberrant lncRNA-miRNA-mRNA network in non-small cell lung cancer

Xiaoyuan Wang¹#, Hang Yin²#, Luquan Zhang¹#, Dayong Zheng¹, Yingnan Yang¹, Jinfeng Zhang¹, Hao Jiang¹, Xiaodong Ling¹, Yanzhong Xin¹, Hao Liang¹, Chengyuan Fang¹, Jianqun Ma¹, Jinhong Zhu³

¹Department of Thoracic Surgery, ²Department of Radiotherapy Oncology, ³Department of Clinical Laboratory, Molecular Epidemiology Laboratory, Harbin Medical University Cancer Hospital, Harbin 150040, China

Contributions: (I) Conception and design: X Wang, H Yin, J Ma; (II) Administrative support: X Wang, J Ma; (III) Provision of study materials or patients: D Zheng, L Zhang; (IV) Collection and assembly of data: Y Yang, J Zhang, H Jiang, Y Xin; (V) Data analysis and interpretation: H Liang, C Fang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

#These authors contributed equally to this work.

Correspondence to: Jianqun Ma. Department of Thoracic Surgery, Molecular Epidemiology Laboratory, Harbin Medical University Cancer Hospital, Harbin 150040, China. Email: jianqunma@aliyun.com; Jinhong Zhu. Department of Clinical Laboratory, Molecular Epidemiology Laboratory, Harbin Medical University Cancer Hospital, Harbin 150040, China. Email: jinhongzhu625@gmail.com.

Background: Non-small cell lung cancer (NSCLC) is the most common cancer and the pathogenesis remain unclear. According to the competing endogenous RNA (ceRNA) theory, long noncoding RNA (lncRNA) have a competition with mRNAs for the connecting with miRNAs that affecting the level of mRNA. In this work, the ceRNA network and the important genes to predict the survival prognosis were explored.

Methods: In the study, we recognized differently expressed genes (mRNAs, lncRNAs and miRNAs) between NSCLC and normal tissues from The Cancer Genome Atlas database (fold change >2, P<0.01) using edgeR. Then, the interaction between lncRNA and miRNA or mRNA and miRNA was explored by miRcode, miRDB, TargetScan, and miRanda. Furthermore, the functions and KEGG pathway were analyzed with DAVID and KOBAS. The connections of these mRNAs were explored by STRING online database. The relation between genes in the network and survival time were further explored by survival package in R.

Results: By bioinformatics tools, we explored 155 lncRNAs, 30 miRNAs and 68 mRNAs and constructed ceRNA network. The functions and KEGG pathway of 68 mRNAs were further analyzed. AQP2, EGF, SLC12A1, TRPV5 and AVPR2 was in the center of network and may play key roles in the development of NSCLC. And mRNA (CCNB1, COL1A1, E2F7, EGLN3, FOXG1 and PFKP), miRNA (miR-31, miR-144 and miR-192) and lncRNA (AC080129.1, AC100791.1, AL163952.1, AP000525.1, AP003064.2, C2orf48, C10orf91, FGF12-AS2, HOTAIR, LINCO0518, LINX1-AS1, MED4-AS1, MIG31HG, MUC2, TTTY16 and UCA1) were closely related with overall survival (OS).

Conclusions: In summary, the present study provides a deeper understanding of the lncRNA-related ceRNA network in NSCLC and some genes may be new target to treat for NSCLC patients.

Keywords: Long noncoding RNAs (lncRNAs); non-small cell lung cancer (NSCLC); The Cancer Genome Atlas; competing endogenous RNAs network; STRING
Introduction

Lung cancer is one of the most common cancer worldwide with the highest morbidity and mortality (1). Eighty to eighty-five percent of lung cancer is non-small cell lung cancer (NSCLC) (2). Though there is a great progress in surgery, chemotherapy and radiation therapy, the 5-year survival rate of NSCLC was still 15% (3). Hence, it is significantly important to have a deeper understanding of the molecular mechanism underlying NSCLC tumorigenesis and seek new therapeutic strategies for NSCLC. Moreover, the regional or distant metastasis will lead to high percent of mortality (4). Therefore, to better study the disease and find the new targets to cure has become extremely importantly. Long noncoding RNAs (lncRNAs) is a type of non-coding RNAs (ncRNAs) with 200 nucleotides to 100 kb in length (5). LncRNAs are with the function of regulating the target gene expression in the process of transcriptional and post-transcriptional but no protein-coding function (6). The research showed the manifestations of lncRNAs in tumors and normal tissues was different. Some lncRNAs played important roles in cancer progression and may be biomarkers in diagnosis, treatment and conditions of prognosis for its stronger tissue specificity (7). However, the clear mechanism of lncRNA in cancer was still not clear. In 2011, the theory of Competing endogenous RNA (ceRNA) hypothesis which is proposed by Sardina et al. was appealing (8). The main theory was some lncRNAs could have an effect of sponge on miRNAs which weak the impact of miRNAs on mRNAs. Additionally, there are more and more research showed the network of lncRNA, miRNA and mRNA is important in the pathogenesis and progression of cancers (9,10). For example, Zhang investigated the lncRNA analysed and built a lncRNA-miRNA-mRNA network in hepatocellular carcinoma (11). Besides, a ceRNA network of esophageal cancer was constructed (12). In renal cell cancer, a ceRNA network was constructed likewise (13). Nevertheless, the study of large-scale samples in NSCLC is uncommon. And the nexuses between prognosis and lncRNAs are undetermined. Therefore, building ceRNA network is very important for prognosis prediction, therapeutic decision and improving the overall survival (OS) for NSCLC sufferers.

At present we studied the expression of genes (lncRNA, miRNA and mRNA) in NSCLC from the database of TCGA. In addition, we built a ceRNA network in NSCLC by bioinformatics means which can find the novel targets and pathway to improve survival for patients.

Methods

Explore these genes with differentially expressed

We obtained the data of RNA sequencing from TCGA. One thousand and thirty-seven NSCLC tumor and 108 normal tissues of mRNA and lncRNA, 999 NSCLC tumor and 91 normal tissues of miRNA were obtained. The edgeR was used to detect the remarkable differentially expressed miRNAs (DEmiRNAs), lncRNAs (DElncRNAs) and mRNAs (DEMmRNAs) (14). To maximize the reliability of these data, these lncRNAs which were not recoded in the database of GENCODE were excluded (15). |Log₂FC| >2 and FDR <0.01 was set as the cut-off value (FC, fold change; FDR, false discovery rate) (16).

Explore the ceRNA network

The miRcode (http://www.mircode.org/) was used to predict lncRNA and miRNA interaction. Then to predict the miRNA and mRNA interaction, the database of miRDB, TargetScan and miRanda were utilized. At last we used the Cytoscape software to explore the lncRNA-miRNA-mRNA network with the interactions of results (17).

Function and pathway enrichment

We used the Database for Annotation Visualization and Integrated Discovery 6.8 (DAVID) to explore the function (18). Then the KOBAS was used to enrich the pathway of Kyoto Encyclopedia of Genes and Genomes (KEGG) (19). The connection between them was explored by STRING (20).

Statistic analysis

We used the unpaired t-test to explore the DEMiRNAs, DElncRNAs and DEMmRNAs. To further understand these genes, the clinical data of NSCLC patients were explored as well. The R package of survival was utilized to plot the survivorship curve.

Results

Differentially expressed lncRNAs, miRNAs and mRNAs

First, we explored the differentially expression of mRNA, miRNA and lncRNA. Two thousand and nine hundred sixty-five mRNAs (2,295 elevations and 670 downregulations, Figure 1A), 1,817 lncRNAs (1,605 elevations and 212
downregulations, Figure 1B) and 135 miRNAs between normal and cancerous tissues (117 elevated and 18 downregulated, Figure 1C) were explored using the above standard filtration analysis (absolute fold change >2 and FDR value <0.01). The results indicate that these genes may distinguish between normal and cancerous tissues.

The functions and KEGG pathway
In addition, the function of DE mRNA and the KEGG pathway were explored by the DAVID and KOBAS online database. The top 20 GO function according to P value were extracellular region, extracellular space, nucleosome, epidermis development, structural molecules activity, integral component of plasma membrane and so on (Figure S1A,B). In the KEGG pathway analysis, the top 5 pathways according to P value was neuroactive ligand-receptor interaction, alcoholism, systemic lupus erythematosus, metabolism of xenobiotics by cytochrome P450 and steroid hormone biosynthesis (Figure S1C,D).

The network of DEmRNAs
To explore the connections of these mRNAs whose expression is different between NSCLC and normal lung tissues, we used the STRING online database. The top 200 gene according to P value was explored by SRTING. The results showed AQP2, EGF, SLC12A1, TRPV5 and AVPR2 were in the center of network and may play key roles in the development of NSCLC (Figure 2).

The construction of ceRNA network
Then, potential interactions between these genes are predicted based on the ceRNA hypothesis. It is expected that the 155 DE1ncRNA can interact with 30 DEmiRNAs
via the miRcode online tool. In addition, miRDB, miRanda and Targetscan were combined to predict 30 DEmiRNA according to DEmiRNA. Furthermore, there are 68 target DEmiRNAs involved in 2,897 different mRNAs were enrolled in the network (Figure 3). Totally, there are 155 DElncRNAs, 30 DEmiRNAs and 68 DEmiRNAs in the network. Besides, the interactions among DElncRNAs, DEmiRNAs and DEmiRNAs were shown in Figure S2.

The OS of ceRNA network

To further explore the function of the genes in the ceRNA network, we did the correlation between these genes and OS. There are 999 patients who have intact follow-up information enrolled in the study. The lncRNA, miRNA and mRNA in the ceRNA network were explored. The results showed the mRNA (CCNB1, COL1A1, E2F7, EGLN3, FOXG1 and PFKP) were closely related with OS (P<0.05, Figure 4). The miRNA (miR-31, miR-144 and miR-192, Figure 5) and lncRNA (AC080129.1, AC100791.1, AL163952.1, AP000525.1, AP003064.2, C2orf48, C10orf91, FGF12-AS2, HOTAIR, LINC00518, LNX1-AS1, MED4-AS1, MIG31HG, MUC2, TTTY16 and UCA1) were closely related with OS (P<0.05, Figure S3).

Discussion

Lung cancer is the most common cause of cancer-related death in the world. Approximately 85% of lung cancer is NSCLC. Therefore, the mechanisms of this disease are urgently needed (21). LncRNAs play a significant role in the development of the disease according to many studies especially in cancer (22). In gastric cancer, lncRNA MT1JP could regulate FBXW7 expression through competitively
Figure 4 The mRNA (CCNB1, COL1A1, E2F7, EGLN3, FOXG1 and PFKP) were closely related with overall survival.

Figure 5 The miRNA (miR-31, miR-144 and miR-192, Figure 5) were closely related with overall survival.

binding to miR-92a (23). In ovarian cancer, lncRNA ABHD11-AS1 played a role in the tumorigenesis and progression by targeting Rhoc (24). In esophageal squamous cell carcinoma, lncRNA CASC9 promotes squamous cell carcinoma growth by negatively regulating PDCD4 expression through EZH2 (25). In colorectal cancer, lncH19 acts oncogene in regulating cancer cell growth (26). In breast cancer, lncRNA-ATB promotes trastuzumab resistance and invasion (27). In prostate cancer, lncRNA HOXD-AS1 regulates proliferation and chemoresistance
by recruiting WDR5 (28). Hence, lncRNAs play important biological roles by regulating gene expression and may be the biomarkers in diagnostics and target to treatment.

In our study, we found the differential expressed lncRNAs, miRNAs and mRNAs between NSCLC and normal lung tissues. Furthermore, we put the DEMRNAs into function and KEGG pathway analysis by using DAVID and KOBAS. The protein network was also constructed by String database. The connection between lncRNA and miRNA was conducted by miRcode (http://www.mircode.org/). The connection between miRNA and mRNA was conducted by miRDB database, TargetScan database and miRanda database. Next, the network was constructed by the Cytoscape software. Finally, the genes in the ceRNA network was also explored to detect whether they are related to OS. The results showed mRNA (CCNB1, COL1A1, E2F7, EGLN3, FOXG1 and PFKP) miRNA (miR-31, miR-144 and miR-192) and lncRNA (AC080129.1, AC100791.1, AL163952.1, AP000525.1, AP003064.2, C2orf48, C10orf91, FGFI2-AS2, HOTAIR, LINC00518, LNX1-AS1, MED4-AS1, MIG31HG, MUC2, TTTY16 and UCA1) were closely related with OS (P<0.05).

But, a little of limitations still existed in this research. The major method of our research was bioinformatics technology, a useful tool for understanding the interactions, pathways and network. But, the networks and functions still need further verification.

Conclusions

In a word, the present study provides a deeper understanding of the lncRNA-related ceRNA network in NSCLC and provided new target for NSCLC patients.

Acknowledgments

Funding: This work was supported by the Fundamental Research Funds for the Provincial Universities in Heilongjiang (No. 2017LCZX94 to H Yin), Haiyan Foundation of Harbin Medical University Cancer Hospital (No. JJQN2017-10 to H Yin and No. JJQN2019-01 to X Wang), Chen Xiao-ping Foundation for the Development Science and Technology of Hubei Province (H Yin), by Hei Long Jiang Postdoctoral (X Wang) and China Postdoctoral Science Foundation (2018M641861 to X Wang), by Jieping Wu Foundation (No. 320.6799.15046 to J Ma) and Academy of Medical Sciences in Heilongjiang (No. 201606 to J Ma).

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References


Figure S1 The functions and KEGG pathway: (A,B) Top 20 GO function of DEmRNAs according to P value; (C,D) the KEGG pathway.
Figure S2 The interactions among DEIncRNAs, DEmiRNAs and DEmRNAs.
The lncRNA (AC080129.1, AC100791.1, AL163952.1, AP000525.1, AP003064.2, C2orf48, C10orf91, FGF12-AS2, HOTAIR, LINC00518, LNX1-AS1, MED4-AS1, MIG31HG, MUC2, TTTY16 and UCA1) were closely related with overall survival. lncRNA, long noncoding RNA.