An Integrative Transcriptomic Analysis Reveals EGFR Exon-19 E746-A750 Fragment Deletion Regulated miRNA, circRNA, mRNA and IncRNA Networks in Lung Carcinoma

Ling Chen*, Shenyi Li*, Weifeng Shi, Yibo Wu

The Affiliated Hospital of Jiangnan University, Wuxi, Jiangsu Province, People's Republic of China

*These authors contributed equally to this work

Correspondence: Yibo Wu, Weifeng Shi, The Affiliated Hospital of Jiangnan University, Wuxi, Jiangsu Province, People's Republic of China, Tel +86-510-68089762; +86-510-68089762, Fax +86-510-68089762, Email moliaty@aliyun.com; swfzj@hotmail.com

Introduction: Competing endogenous RNA (ceRNA) appears to be an important post-transcriptional manner that regulates gene expression through a miRNA-mediated mechanism. Mutations in exon-19 of EGFR were frequently observed in lung cancer genes, which were associated with EGFR activity and EGFR-targeted therapies.

Methods: We explored the transcriptome regulated by mutation in EGFR exon-19 E746-A750 fragment via using a network modeling strategy. We applied transcriptome sequencing to detect the deletion process of EGFR exon-19 E746-A750 fragment. Bioinformatics analyses were used to predict the gene target pairs and explain their potential roles in tumorigenesis and progression of lung cancer.

Results: We conducted an explorative lncRNA/miRNA/circRNA and mRNA expression study with two groups of lung adenocarcinoma tissues, including EGFR exon-19 E746-A750 deletion group and EGFR exon-19 wild-type group. Meanwhile, we screen out the hub genes related to the EGFR-19-D patient. Significant pathways and biological functions potentially regulated by the deregulated 128 non-coding genes were enriched.

Conclusion: Our work provides an important theoretical, experimental and clinical foundation for further research on more effective targets for the diagnosis, therapy and prognosis of lung cancer.

Keywords: microRNAs, miRNAs, mRNAs, long non-coding RNAs, lncRNAs, gene regulating network, lung cancer

Introduction

Lung adenocarcinoma is one of the most common subtypes of non-small cell lung cancer (NSCLC) and is the leading cause of cancer-related death worldwide.1 As lung cancer was prone to early metastasis, and the prognosis of lung cancer patients was generally poor, so the 5-year survival rate used to be less than 80%.2 Although the treatment has been improved over the past decades, the survival rate remains low.3

Epidermal growth factor (EGFR) is immobilized on the cell surface and binds to its specific receptor to activate cells. Somatic mutations in the EGFR gene cause them to lose their ability to bind to receptors, which activates and causes cell to divide. Some studies have shown that mutations in the EGFR signaling pathway are potential therapeutic targets. Inhibition of EGFR signaling pathway activation can improve the prognosis of cancer patients with EGFR mutations.4 There are seven types EGFR exon-19 mutation in NSCLC. Among the somatic mutations of EGFR tyrosinase domain in NSCLC, the proportion of EGFR intra-exon 19 E746-A750 (EGFR-19-D) fragment deletion is the most common.5

MicroRNAs (miRNAs) is a non-coding single-stranded RNA molecule with about 22 nucleotides encoded by an endogenous gene that inducing degradation by binding specifically to the 3'-untranslated region (UTR) mRNA sequence,6
or inhibiting of mRNA translation. Long non-coding RNAs (lncRNAs) belong to a class of non-coding ribonucleic acids with a molecular weight greater than 200 nucleotides, which regulate the expression of epigenetic genes in the transcriptional and transcriptional levels of various physiological and pathological processes. In addition, the circular RNA (circRNA) molecule has a closed ring structure, which is composed of 5'-3' polarity and a poly A tail. Without affection by RNA exonuclease, the expression is more stable and not easy to degrade.

The new regulatory mechanism between non-coding RNA and coding RNA in Senna hypothesis contributes to the understanding of cancer initiation and development. It is reported that lncRNAs, as competitive endogenous RNA (ceRNA) or microRNA sponge, inhibit the expression of microRNA response elements (MREs) by competing with common miRNAs in tumorigenic process, such as imprinting, X inactivation and development. Previous studies mainly focused on the interaction network between lncRNA and RNA regulation. Some miRNA and lncRNA molecules have important significance in the diagnosis and treatment of tumors and can be used as new molecular markers for tumor prognosis. However, EGFR 19 downstream molecules, especially miRNAs, lncRNAs, circRNAs and mRNA-related regulatory networks, have not been thoroughly investigated in NSCLC.

In this study, we conduct exploratory research on lncRNA/miRNA/circRNA and mRNA expression with two groups of lung adenocarcinoma tissues, including EGFR exon-19 E746-A750 deletion group and EGFR exon-19 wild-type group, using the transcript microarray technology. Meanwhile, we screen out the hub genes related to the EGFR-19-D patient. Accordingly we discuss the association of hub genes function with EGFR exon-19 deletion, tumorigenesis and metastasis.

Materials and Methods
Clinical Specimens
All patients were collected from the Affiliated Hospital of Jiangnan University (Table 1). The medical ethics committee of the affiliated hospital of Jiangnan University reviewed this study. This study complies with the Declaration of Helsinki. They sanctioned the collection of the specimens, and the patients gave written informed consent. The clinical specimens included three cases of EGFR19 deletion and three cases of EGFR19 non-deletion lung cancer. None of lung cancer patient received any treatment before surgery, including radiotherapy and chemotherapy. All specimens of lung cancer were obtained by the chief pathologist after operation. They were immediately frozen in liquid nitrogen and stored at −80°C. All the tissues were confirmed histologically.

RNA Isolation, cDNA Construction and Illumina Deep Sequencing
Total RNAs of all specimens was extracted using the Ribo-Zero Reagent (Invitrogen, USA) according to the instruction manual. The TruseqTM RNA sample prep kit (Illumina, San Diego, CA, USA) was used to constructed cDNA library following the manufacture’s instruction. Briefly, oligo (dT) magnetic beads were used to the purification of poly (A) +RNA from total RNA through shearing into short fragments, and Truseq RNA sample preparation kit was used to cDNA library synthesis according to the manufacturer’s instructions (Illumina). RNA was quantified using the NanoDrop (Thermo Scientific NanoDrop 2000), and RNA integrity number (RIN) >7 were used for cDNA library construction. After quantitation, the samples were clustered (TruSeq paired-end cluster kit v3-cBot-HS; Illumina) and sequenced on the HiSeq 2000 platform (100 bp, TruSeq SBS kit v3-HS 200 cycles; Illumina).

Table 1 Information of Clinical Specimens for Microarray

<table>
<thead>
<tr>
<th>No.</th>
<th>Gender</th>
<th>AGE</th>
<th>Type</th>
<th>Volume (cm)</th>
<th>Neurological Invasion</th>
<th>Vascular Invasion</th>
<th>LNM</th>
<th>Grade</th>
<th>EGFR-19 (Deletion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D191</td>
<td>Female</td>
<td>53</td>
<td>Adenocarcinoma</td>
<td>2×2×1.5</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>1-3</td>
<td>+</td>
</tr>
<tr>
<td>D192</td>
<td>Male</td>
<td>51</td>
<td>Adenocarcinoma</td>
<td>2.5×2.5×2.4</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>1-2</td>
<td>+</td>
</tr>
<tr>
<td>D193</td>
<td>Male</td>
<td>52</td>
<td>Adenocarcinoma</td>
<td>5.3×3.5×1.5</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1-2</td>
<td>+</td>
</tr>
<tr>
<td>N191</td>
<td>Male</td>
<td>59</td>
<td>Adenocarcinoma</td>
<td>4×3×4</td>
<td>-</td>
<td>0</td>
<td>1</td>
<td>1-2</td>
<td>-</td>
</tr>
<tr>
<td>N192</td>
<td>Female</td>
<td>83</td>
<td>Adenocarcinoma</td>
<td>4×3×5</td>
<td>-</td>
<td>0</td>
<td>1</td>
<td>1-2</td>
<td>-</td>
</tr>
<tr>
<td>N193</td>
<td>Male</td>
<td>49</td>
<td>Adenocarcinoma</td>
<td>2×1.6×2</td>
<td>-</td>
<td>0</td>
<td>1</td>
<td>1-2</td>
<td>-</td>
</tr>
</tbody>
</table>
GO, KEGG Pathway Analysis
KOBAS 2.0 is an open web server for annotation and identification of enriched pathways. We performed possible GO and KEGG enrichment pathways predictions through KOBAS annotation. The significant items were screened out under the conditions of \( p<0.05 \) and FDR <0.05.

PPI Network Construction and Hub Gene Analysis
To screen out the interaction networks between the differentially expressed mRNAs, we used the StringApp plugin in Cytoscape 3.7.1 to import the protein–protein interaction (PPI) network from the STRING database. Then, the Cytohubba plugin was used to screen out the hub genes. The top four hub genes were screened out by degree score=3 in the Cytohubba plugin. Then the top four hub genes related to mRNA–miRNA pair, mRNA–circRNA pairs, and mRNA–ncRNA pairs were selected. Network visualization was achieved by Cytoscape 3.7.1.

Statistical Analysis
The screened differentially expression genes were analyzed by independent-sample \( t \)-test, and the results were expressed as the mean±standard deviation (SD). \( P \) value <0.05 were consider statistically significant, and SPSS 17.0 was used to statistical analysis. The heat map and other graphs in the article are made by R-project.

Results

The lncRNAs Expression Profiling Regulated by EGFR-19-D Exon Mutation
The lncRNA expression profiling was performed in EGFR-19-D exon mutation biopsies and non-EGFR mutation lung cancer biopsies (Supplementary 1). A total of 44 lncRNAs were significantly induced, and 36 lncRNAs were reduced with the EGFR-19 exon mutation (Figure 1A). Among these 80 lncRNAs, 17 lncRNAs were non-protein coding genes and 52 lncRNAs were no-name genes. These genes were mainly involved in signaling pathways, such as the“complement and coagulation cascades”,“Glutathione metabolism”,“Taurine and hypotaurine metabolism” (Figure 1B). Recently, Qing Tang’s study indicated that compared to normal people, complement and coagulation cascades pathway played an important role in the NSCLC patient.

The miRNAs Expression Profiling Regulated by EGFR-19-D Exon Mutation
The miRNA expression profiling was performed in lung cancer between EGFR-19-D group and EGFR wild-type group (Supplementary 2). Among these differentially expressed miRNAs, 10 miRNAs were down-regulated significantly in EGFR-19-D exon mutation lung cancer, while 23 miRNAs were up-regulated compared to EGFR wild-type group (Figure 1C). These genes were mainly involved in signaling pathways, such as the “MAPK signaling pathway”, “ErbB signaling pathway”, “EGFR tyrosine kinase inhibitor resistance”, “Endocrine resistance”, “Ras signaling pathway”, “Spliceosome” (Figure 1D). Among these, MAPK signaling pathway and Ras signaling pathway were top2 mainly gene enrichment pathway. KEGG (Kyoto Encyclopedia of Genes and Genomes) database analysis shows that EGFR mutations mainly activate signaling pathway including RAS-ERK signaling pathway, PI3K-Akt signaling pathway and PLCG-ERK signaling pathway.

The circRNAs Expression Profile Regulated by EGFR-19-D Exon Mutation
Seven circRNAs expression were up-regulated and 10 circRNAs were down-regulated significantly in the EGFR-19-D exon procession (Figure 2A). These differentially expressed circRNAs were enriched in the pathway such as “Cell adhesion molecules (CAMs)”, “Synaptic vesicle cycle”, “Viral myocarditis”, “Autoimmune thyroid disease”, “Type I diabetes mellitus”, “Graft–versus–host disease”, “Allograft rejection”, “Propanoate metabolism”, and “Citrate cycle (TCA cycle)” (Figure 2B). Besides, the total number of circRNAs was shown in Supplementary Figure 1. In the end, the top 10 deregulated RNAs are shown in Table 2.
The mRNAs Expression Profile Regulated by EGFR-19-D Exon Mutation

The mRNA expression profiling was performed in EGFR-19-D exon mutant biopsies and non-EGFR mutant lung cancer biopsies (Supplementary 4). A total of 38 mRNAs were significantly induced, and 35 mRNAs were reduced by EGFR-19-D exon mutation (Figure 2C). These genes were mainly involved in signaling pathways, such as the “Complement and coagulation cascades”, “Glutathione metabolism”, and “Taurine and hypotaurine metabolism” (Figure 2D). Besides, we constructed the PPI network to explore the major protein in the EGFR-19-D exon mutation
We found that CASR, FGG, HOXB9, OPRK1 were the hub genes in the network (Figure 4A). Among these genes, the gene cluster of CASR and OPRK1 interacting with crosstalk is more obvious. It indicated that there were three HOXB9-circRNA pairs, two HOXB9-lincRNA pairs and four HOXB9-miRNA pairs (Figure 4B). Six miRNAs, four circRNAs and nine lincRNAs were related to gene CASR (Figure 4C). Besides, there were 8 FGG-lincRNA pairs, 1 FGG-circRNA pair and 4 OPRK1-circRNAs, 18 OPRK1-lincRNA pairs, and 5 OPRK1-miRNA pairs (Figure 4D).
The ceRNA Network Constructed by the Screened RNAs

Total RNAs were transferred from ENSEMAL ID to SYMBOL. The RNAs which could match symbol name were screened out. In total, 1224 RNA–RNA pairs were screened out (Figure 5), including 15 circRNA, 31 lincRNAs, 34 miRNAs, and 71 mRNAs. The total number of RNA–RNA pairs was shown in Supplementary 5.

Corroboration of the 18 RNA Expressions with RT-qPCR

RT-qPCR was performed on five pairs of lung cancer tissues with EGFR-19 exon mutation and EGFR-19 exon wild-type, including five miRNAs (Figure 6A), five mRNAs (Figure 6B), four lncRNAs (Figure 6C) and four circRNAs (Figure 6D). Eighteen RNA expressions were statistically different (p<0.05), of which miR-300, CLCA2, CYP24A1, CASR, LINCO1133, LINC00473, and circITCH were significantly different (p<0.01). And the results were consistent with the sequencing results.

Discussion

In the past study, GSH metabolism was thought to be one major mechanism of chemo-resistance in the NSCLC. In addition, Pool et al. deemed that the GSK metabolism pathway might be essential for tumor response to EGFR-TKI afatinib alone or combined with Rapamycin. Presently, EGFR-TKIs are effective in the lung cancer patients who have a deletion mutation of exon 19 and L858R mutation of exon 21. Canonical ligand-dependent EGFR signaling pathway contains the Ras/MAPK pathway, the PI3K/AKT pathway, and the phospholipase C (PLC)/protein kinase C (PKC) signaling cascade.

We found four new hub genes, which were not directly contained in the canonical pathway were related to the EGFR-19 exon E746-A750 fragment deletion. In our study, CaSR was up-regulated in the EGFR-19 exon deletion group compared to the EGFR-19 exon wild-type. When we analyzed whether there was any relationship between these CaSR and EGFR-19 exon E746-A750 mutation, we found that CaSR interacted with G proteins that could regulate the EGFR downstream signaling molecular pathway. Moreover, CaSR was activated by signals from glutathione and stimulated the CaSR trans-activated expression of EGFR, causing phosphorylation of ERK and secretion of PTHrP. Besides, CaSR could induce calcium overload and subsequent mitogen-activated protein kinase (MAPK) pathways. In our study, 6 miRNAs, 4 circRNAs and 17 lncRNAs were predicted to bind to CaSR. The function of has-miR-499b-5p is unknown, but its family member miR-499a has been well studied. MiR-499a inhibits the expression of MAP2K1 by directly targeting its 3'UTR and modulates the activity of the MEK1/ERK1/2/c-Jun pathway by auto-regulating the feedback loop. Has-miR-551a is a medium for 5-FU to acquire drug resistance. Zhang data indicated that LINC-CECR7 was negatively associated with OS Hepatocellular Carcinoma. CircRNA-CEP70 acted as a novel circRNA targeting the miR-561/E2F8 that regulated SH-SY5Y cell proliferation and differentiation. Wang et al. identified that Axl-altered miR-548b played a crucial role in the gefitinib-induced apoptosis and EMT of gefitinib-resistant lung cancer cells by targeting CCNB1.
FGG belongs to the fibrinogen family, which is the critical gene in the complement and coagulation cascades pathway. High expression of fibrinogen causes tumor metastasis and recurrence in the patients with esophageal squamous cell carcinoma, especially combined with miRNA. Thus, decreased expressed FGG might attenuate the processes of pathologic thrombosis and angiogenesis associated with cancer cells. Moreover, we found that FGG might bind to circRNA-TM9SF3, LOC105376382. Yet, the function of LINC02678 is unknown.

HOXB9 is a homeobox transcription factor, which plays an important role in carcinoma development. Nguyen et al unveiled that HOXB9 acted as promoters of cancer cell metastasis and mediators of chemotactic invasion and colony outgrowth in lung adenocarcinoma. HOXB9 was down-regulated in our study. As a consequence, we speculated that...
EGFR-19 exon deletion lung cancer cell metastasis is related to HOXB9. Besides, Siu study demonstrated that down-regulated miR-203 could induce the bone metastasis and TKI resistance, which EGFR pathway activated via altered expression of EGFR ligand (EREG and TGFA) and anti-apoptotic proteins (API5 BIRC2 and TRIAP1). MiR-873 could act as a negative regulator of tumor proliferation and metastasis.

OPRK1 is the κ opioid receptor, which encodes G protein-coupled opioid receptors. The activation of EGFR transregulates the function of G protein-coupled opioid receptors through GRK2 (G protein-coupled receptor kinases 2). Figure 4 OPRK1, CASR, HOXB9 and FGG are the hub genes in the PPI network. (A) The top four hub genes OPRK1, CASR, HOXB9 and FGG. The hub genes were screened out by the cytohubba-APP in the Cytoscape. The top four hub genes degree score are all equal to 3. (B) In the network, taking HOXB9 as the hub gene, its related lincRNAs, circRNAs and miRNAs. (C) OPRK1 and CASR genes related RNAs. (D) FGG and its related RNAs. Red circles represent hub mRNAs. The green circles represent lincRNA, the blue circles represent miRNA and the yellow circles represent circRNA.
phosphorylation and activation. GRK2 acts as a mediator of cross-talk from RTK to GPCR (G protein-coupled receptor) signaling pathway. In the past study, OPRK1 was concentrated on the alcohol and drug dependence. Zhang et al reported that was down-regulated in HCC (hepatocellular carcinoma) tissues. Multiple Scancer’s study indicated that miR-136-5p was deregulated. MiRNA-136-5p was down-regulated in hepatocellular carcinoma, while it was up-regulated in lung adenocarcinoma. So far, no study has mentioned the relationship between miR-136-5p and OPRK1. Kumar et al indicated that the putative targets of miR-300 were enriched in the TGF-beta signaling pathway, moreover up-regulation of miR-300 may induce cisplatin resistance.

In our analysis, we found that LINC01750, LINC00885, LINC01331, LINC01335, LINC01332, AC011752.1, LOC442497, AL3354766.2, AC354766.2, RIMS1, METP1, PTPN13, CEP70 and CENE-1 were the cross-linked targets.

Figure 5 The ceRNA network of lincRNA-circRNA-miRNA-mRNA in EGFR-19 exon mutation lung cancer. Blue indicate circRNAs, red indicate lincRNA, yellow indicate miRNA, and green indicate mRNA. Total RNAs were mapped with known genes.
between OPRK1 and CaSR. The interaction between OPRK1 and CaSR needs further study. However, we believe that the two genes play an important role in the expression and metastasis of EGFR-19 exon mutation in lung carcinoma.

**Conclusion**

EGFR exon-19 E746-A750 fragment deletion is prone to have a longer overall survival and/or progression-free survival with treatment of gefitinib or erlotinib compared to EGFR exon 19 wild-type. These hub genes may potentially give a new strategy for EGFR-19-D treatment. The ceRNA network may provide a new exploration about EGFR-19-D, which one point may be the newly response to EGFR-19-D patients.

**Additional Information**

Supplementary information accompanies this paper in Supplementary 1–5, Supplementary Figure 1, Tables 1 and 2. Supplementary 6 listed the primers involved in this study.
Acknowledgments
This study was supported by the grants of National Natural Science Fund (81701511) and grants of Wuxi Young Medical Talents Program (QNRC016, QNRC017).

Disclosure
The authors declare no conflicts of interest.

References