Molecular and Cellular Probes 30 (2016) 146–152

Contents lists available at ScienceDirect

Molecular and Cellular Probes

journal homepage: www.elsevier.com/locate/ymcpr

Identification of crucial regulatory relationships between long non-coding RNAs and protein-coding genes in lung squamous cell carcinoma

Xiaofen Wu [a], Lei Ruan [a], Yi Yang [a], Qi Mei [b, *]

[a] Department of Gerontology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China
[b] Department of Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China

ARTICLE INFO

Article history:
Received 11 September 2015
Received in revised form
22 January 2016
Accepted 19 February 2016
Available online 27 February 2016

Keywords:
Lung squamous cell carcinoma
Differentially expressed protein-coding gene
Long non-coding RNA
Regulatory network

ABSTRACT

Purpose: This study aimed to analyze the relationships of long non-coding RNAs (lncRNAs) and protein-coding genes in lung squamous cell carcinoma (LUSC).
Methods: RNA-seq data of LUSC deposited in the TCGA database were used to identify differentially expressed protein-coding genes (DECGs) and differentially expressed lncRNA genes (DE-lncRNAs) between LUSC samples and normal samples. Functional enrichment analysis of DECGs was then performed. Subsequently, the target genes and regulators of DE-lncRNAs were predicted from the DECGs. Additionally, expression levels of target genes of DE-lncRNAs were validated by RT-qPCR after the silence of DE-lncRNAs.
Results: In total, 5162 differentially expressed genes (DEGs) were screened from the LUSC samples, and there were seven upregulated lncRNA genes in the DEGs. The upregulated DECGs were enriched in GO terms like RNA binding and metabolic process. Meanwhile, the downregulated DECGs were enriched in GO terms like cell cycle. Furthermore, the lncRNAs *PVT1* and *TERC* targeted multiple DECGs. *PVT1* targeted genes related to cell cycle (e.g. *POLA2*, *POLD1*, *MCM4*, *MCM5* and *MCM6*), and reduced expression of *PVT1* decreased expression of the genes. *TERC* regulated several genes (e.g. *NDUFAB1*, *NDUFA11* and *NDUFB5*), and reduced expression of *TERC* increased expression of the genes. Additionally, *PVT1* was regulated by multiple transcription factors (TFs) identified from DECGs, such as *HSF1*; and *TERC* was modulated by TFs, such as *PIR*.
Conclusion: A set of regulatory relationships between *PVT1* and its targets and regulators, as well as *TERC* and its targets and regulators, may play crucial roles in the progress of LUSC.

1. Introduction

Lung squamous cell carcinoma (LUSC) is a common type of non small-cell lung cancer and the second leading cause of death resulting from lung cancer, causing about 400,000 deaths per year worldwide [1,2]. It is urgent to reveal the molecular mechanisms underlying LUSC oncogenesis.

In the past years, various molecular players have been demonstrated to be correlated with LUSC, such as protein-coding genes and lncRNAs. For instance, it has been demonstrated that in the mouse lung, biallelic inactivation of *Lkb1* and *Pten* that causes elevated *PD-L1* expression, leads to LUSC [3]. *PRKCI* and *SOX2* oncogenes cooperate to stimulate activation of Hedgehog signaling and drive a stem-like phenotype in LUSC [4]. Besides, high expression of *FGFR1* was found in 16% of a clinical cohort of LUSC patients [5]. Furthermore, high expressions of the lncRNAs *MALAT1* and *HOTAIR* are associated with poor prognosis of LUSC [6,7]. A recent study has been reported that the lncRNA LINC01133 is highly expressed in LUSC and it predicts survival time [8]. However, the interactions between differentially expressed protein-coding genes (DECGs) and differentially expressed lncRNA genes (DE-lncRNAs) are still elusive.

In this study, RNA-seq data of LUSC deposited in The Cancer Genome Atlas (TCGA) database were used to identify DECGs and DE-lncRNAs. Following the functional analysis of DECGs, the

* Corresponding author. Department of Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Jiefang Avenue 1095, Wuhan 430030, China.
E-mail address: borismq@163.com (Q. Mei).

http://dx.doi.org/10.1016/j.mcp.2016.02.009
0890-8508/

regulatory relationships between DECGs and DE-lncRNAs were analyzed. Additionally, expression levels of target genes of DE-lncRNAs were validated after the silence of DE-lncRNAs. These results were expected to provide novel information for the understanding of molecular mechanisms of LUSC tumorigenesis.

2. Materials and methods

2.1. RNA sequencing data

The preprocessed RNA-seq level 3 data of LUSC were downloaded from TCGA database (http://cancergenome.nih.gov/), which provides a platform for researchers to download the data sets of genomic abnormalities driving tumorigenesis. In total, the dataset contains 17 pairs of matched primary solid tumor samples and adjacent normal tissue samples, and RPKM (Reads Per Kilobase of exon model per Million mapped reads) values of each gene in these samples were included.

2.2. DEGs and DE-lncRNAs screening

The paired-sample T test was used to identify the DEGs (including DECGs and non-protein coding genes) between the two matched samples, and p-value was adjusted by the Benjamin and Hochberg (BH) method [9]. Only the genes with adjusted p-value < 0.05 and FC (fold change) > 2 were chosen as DEGs.

Based on annotation information in the GENCODE database (http://www.gencodegenes.org/) [10], the DE-lncRNAs were screened from the identified DEGs.

2.3. Enrichment analysis of DECGs

To further reveal the functions of DECGs, the Gene Ontology (GO) functional and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of DEGs were performed using the clusterProfiler package of R [11]. The GO terms and pathway terms with adjusted p-value < 0.05 were selected as the significant terms.

2.4. Prediction of DE-lncRNA target genes and their functions

Based on the coexpression relationships of lncRNA and protein-coding genes, the target genes of DE-lncRNAs were predicted from the screened DECGs, and only genes with p-value < 0.05 and expression similarity >0.8 were chosen for subsequent analyses. The expression similarity of DECGs and lncRNAs was measured as previous researchers did [12,13]. Regulatory networks of lncRNAs and their target genes were then visualized by Cytoscape (http://cytoscape.org/) [14]. Additionally, the GO functional enrichment analysis of target genes was conducted using the clusterProfiler package of R.

2.5. Prediction of upstream transcription factors (TFs) of DE-lncRNAs

To further analyze the differential expression mechanisms of DE-lncRNAs, based on the TF information in TRANSFAC database [15], we used MotifDb package of R [16] to predict the potential TFs regulating DE-lncRNAs from DECGs. This package is able to search the binding motif in TF of the promoter sequence, based on the gene promoter sequences of lncRNAs provided by users. The matching degree of motif sequence and promoter sequence was calculated by pulse-width modulation (PWM) algorithm [17], and TFs with matching degree $\geq$ 85% were considered as the potential upstream TFs regulating DE-lncRNAs. The TF-DE-lncRNA regulatory network was visualized by Cytoscape.

2.6. Validation of downstream gene expression of significant lncRNAs

Based on the gene sequences of lncRNAs in the National Center For Biotechnology Information (NCBI) database, three small interfering RNA (siRNA) sequences and one siRNA-scramble sequence for each lncRNAs (Table 6) were designed using BLOCK-iT™ RNAi Designer (www.invitrogen.com/rnai). Besides, based on the sequences of downstream genes of lncRNAs in NCBI, gene primers were designed using the Primer5 software (PRIMER-E Ltd, Plymouth, UK). Primer sequences of genes were listed in Table 6.

Human LUSC cell line SK-MES-1 purchased from the Cell Bank at the Chinese Academy of Sciences were placed in a 12 well plate and transfected with siRNA sequences and siRNA-scramble sequence using Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA), respectively. After transfection of 24 h, total RNA for mRNA expression detection was extracted from SK-MES-1 cells by TriZol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. Afterwards, total RNA was reverse transcribed using PrimeScript RT Master Mix (Takara, RR036A, Shiga, Japan). After cDNA synthesis, mRNA expression levels were tested using Power SYBR Green PCR Master Mix (#4367659, Applied Biosystems, Foster City, CA, USA), and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as the reference gene.

2.7. Statistical methods

Statistical analysis was performed using the SPSS 19.0 software (SPSS, Chicago, USA). One-way analysis of variance (ANOVA) followed by Duncan test was used to compare group means. Difference with P < 0.05 was considered significant.

3. Results

3.1. Predicted DEGs and DE-lncRNAs

In total, 5162 DEGs (4489 upregulated ones and 673 downregulated ones) were identified from the LUSC samples, comparing the normal samples. Among these DEGs, there were 7 upregulated lncRNA genes (Table 1).

3.2. Functions of DECGs

According to GO functional enrichment analysis of DECGs in the DEGs, the upregulated DECGs were significantly enriched in a series of GO terms, such as RNA binding (e.g. *CEBPZ*, *TRIM28* and *HNRNPR*), membrane-enclosed lumen (e.g. *MRPL52* and *COL1A1*) and metabolic process (e.g. *CEBPG* and *CDH3*) (Table 2). Meanwhile, the downregulated DEGs were distinctly enriched in multiple GO terms, such as ion binding (e.g. *KIF20A* and *TRAIP*), microtubule cytoskeleton (e.g. *KIF18B* and *CEP152*) and cell cycle (e.g. *KIF20A* and *CENPA*) (Table 3).

Table 1
The details of the identified differentially expressed lncRNAs.

LncRNA	Chr	Start	End	Strand
DIO3OS	chr14	101,552,221	101,560,431	–
HAR1A	chr20	63,102,205	63,104,386	+
FAM138B	chr2	113,577,382	113,578,852	+
BCYRN1	chr2	47,331,060	47,344,517	+
TERC	chr3	169,764,520	169,765,060	–
HCG9	chr6	29,975,112	29978410	+
PVT1	chr8	127,794,533	128,101,253	+

Chr, chromosome; the columns of "Start" and "End" represent the start site and end site of the gene sequence of lncRNA, respectively. "+" and "-" represent positive-sense strand and antisense strand, respectively.