CANCER BIOLOGY & THERAPY
2017, VOL. 18, NO. 2, 115–122
http://dx.doi.org/10.1080/15384047.2017.1281494

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

# Co-expression network analysis of long noncoding RNAs (lncRNAs) and cancer genes reveals*SFTA1P* and *CASC2*abnormalities in lung squamous cell carcinoma

Guang-Qing Huang[a,*], Zun-Ping Ke[b,*], Hai-Bo Hu[c*], and Biao Gu[d]

[a]Department of Intensive Care Unit, Renmin Hospital, Hubei University of Medicine, Shiyan, Hubei, P.R. China; [b]Department of Cardiology, The Fifth People's Hospital of Shanghai, Fudan University, Shanghai, China; [c]Department of Thoracic Surgery, The Affiliated Huai'an Hospital of Xuzhou Medical University and The Second People's Hospital of Huai'an, Huai'an, China; [d]Department of Thoracic Surgery, Huai'an First People's Hospital, Nanjing Medical University, Huai'an, Jiangsu, China

**ABSTRACT**

Lung squamous cell carcinoma(LSCC) is the most common and aggressive lung tumor with poor clinical outcome. Previously studies showed that deregulation of long noncoding RNAs (lncRNAs) were involved in LSCC. We intended to figure out the role of lncRNAs in the regulation process of cancer-related genes and pathways they are involved. Data of 552 samples, including 501 cancer samples and 51 normal ones, were extracted from The Cancer Genome Atlas (TCGA). Differentially expressed lncRNAs (DEIs) were screened out (FDR<0.05, |logFC|>1) and then followed by GO ontology and KEGG annotation analysis. Oncogenes from COSMIC data set and Tumor suppressor genes (TSGs) from TSGene data set were collected and analyzed by gene Set Enrichment Analysis (GSEA) . The differentially expressed oncogenes and tumor supressor gene (TSGs) were obtained and co-expression analysis was conducted to generate co-expression lncRNA-gene pairs, which can be helpful in figuring out the role of lncRNA in the regulation of oncogenes and tumor suppressor genes. A total of 31 lncRNAs with low expression levels and 37 lncRNAs with high expression levels were screened out and most of them were enriched in pathways such as meiosis, male gamete generation, defensins. Of note, *SFTA1P* and *CASC2* were found to be related with most of the oncogenes and TSGs by co-expression analysis. We suggested *SFTA1P* and *CASC2* played important role in the regulation of both oncogene and TSGs during the carcinogenesis of LSCC and have the potential to be applied in future diagnosis, prognostic process and target therapy of LSCC.

**ARTICLE HISTORY**
Received 23 September 2016
Revised 17 December 2016
Accepted 8 January 2017

**KEYWORDS**
Enrichment pathways; lncRNA; lung squamous cell carcinoma; prognostic marker

## Introduction

Lung squamous cell carcinoma is one of the most common type of cancer, accounting for 40∼50% of all primary lung cancer.[1] Currently, the prognosis of NSCC patients is still very poor. There was not much research about the prognostic and predictive markers for LSCC other than TP53 mutations, which was identified as a causative somatic aberration.[2]

Long noncoding RNA(lncRNA), with the length of 200 nt and pervasively transcribed in the mammalian genome,[3] is involved in a series of processes regulating tumor biology.[4, 5] It was reported that lncRNAs are differentially expressed in many kinds of cancer, exerting important regulations on tumor biology via regulating oncogenes or tumor suppressors.[6] Given the fact that the number and functions of all kinds of lncRNAs are not well known, differences in IncRNA levels from one specific type of lung cancer might help us better in identifying their potential role as biomarker of LSCC. Therefore, the subtypes of NSCLC and SCLC were not included in our study. We could expound role of the differentially expressed lncRNA in the occurrence and development of cancer more convincingly. Several reviews have emphasized that the role of ncRNAs in different cellular processes has been largely underestimated. It's reported that non-coding RNAs may be involved in the silence of tumor suppressor genes (TSGs) epigenetically, which can inhibit normal cellular growth in cancer.[7] LncRNAs mainly played roles in regulation of transcriptional, posttranscriptional and epigenetic mechanisms in the process of tumorigenesis.[8] Several studies confirmed the deregulation role of lncRNA in cancers, such as hepatocellular carcinoma, breast,[9] osteosarcoma,[10] colon and prostate cancers.[11, 12]

Differentially expressed IncRNAs provide essential evidence to figure out the difference between cancer samples and normal specimens, which can help us in exploring the relevant pathways they were involved. For example, in previous studies, *NOTCH1* or *NOTCH2* were found to be 2 tumor suppressor specific to epithelial malignancies, thus coming to the conclusion that targeted inhibition of the Notch pathway may affect squamous epithelial malignancies.[4]

In spite of many studies on the potential important roles of lncRNAs in various kinds of cancer, there had not been much report on lncRNAs associated with LSCC except one study which

**CONTACT** Hai-Bo Hu dr_huhaibo@aliyun.com Department of Thoracic Surgery, The Affiliated Huai'an Hospital of Xuzhou Medical University and The Second People's Hospital of Huai'an, Huai'an, China; Biao Gu prof_zh@sina.com Department of Thoracic Surgery, Huai'an First People's Hospital, Nanjing Medical University, 6 Beijing West Road, Huai'an, Jiangsu 223300, China.
*Co-first authors.

Supplemental data for this article can be accessed on the publisher's website.

indicated that lower expression of CASC2 was accompanied by poor prognosis of non small cell lung cancer(NSCLC).[13] In this study, we attempted to provide large-scale survey of lncRNAs within LSCC, the subtype of non-small cell lung cancer, and to figure out the role of lncRNA in the regulation process of cancer-related genes of LSCC patients, which may help us in further analysis of identifying pathways they are involved, and thus to identify prognostic marker and novel therapy for LSCC patients.

## Materials and methods

### Source of data and data pre-processing

A total of 552 expression data (501 LSCC samples and 51 normal ones) were extracted from TCGA (The Cancer Genome Atlas), based on the platform of IlluminaHiSeq. A total of 175 lncRNAs were screened out by annotation profiles from ensemble data set and FRKM values were extracted out and used in the formation of data matrix.[14, 15]

### Identification of differentially expressed lncRNA

Differentially expressed lncRNAs (DEIs) were extracted out from the expression matrix by the algorithm method of limma package.[16] The thresholds were FDR<0.05 and |logFC|>1.

### GO annotation and KEGG pathway analysis

We applied lncRNA2 function [17] software which include functional annotation of 9625 human lncRNA genes to annotate the DEIs. All nodes in Gene Ontology (GO) and a total of 4380 function pathways were included in this functional pathway. GO functional annotation and KEGG pathway analysis of DEIs involved was analyzed by lncRNA2 function.

### Collection of oncogenes and Tumor suppressor genes (TSGs)

Data of all the 634 oncogenes were from COSMIC data set,[17, 18] which included information of somatic mutation associated with all kinds of cancers. The whole 1452 tumor suppressor gene data were extracted from TSGene data set and TAG data set.[19] There were 1217 TSGs collected from more than 9000 studies in TSGene 2.0 (1018 protein coding genes and 199 non-coding genes) while 265 TSGs were included in TAG data set.

### Identification of differentially expressed oncogenes and TSGs

Cluster heatmap of the top 100 most significantly differentially expressed oncogenes and TSGswere painted by heatmap 2.0 of Limma package in R language with the thresholds of FC = 2 and FDR<0.05.

### Enrichment analysis of gene set

Differentially expressed oncogenes and TSGs were annotated by GSEA (Gene Set Enrichment Analysis).[20, 21] GO, KEGG and cancer gene neighborhood were all included in GSEA, which can help us in obtaining the pathways and functions of differentially expressed cancer genes.

### Co-expression analysis

Pearson correlation coefficient of each lncRNA and cancer gene, as well as confidence interval and p-value, was calculated out using the function of cor ()andcor.test() of R fuction. The formula were as follow,

$$r = \frac{1}{n=1}\sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{s_X}\right)\left(\frac{Y_i - \overline{Y}}{s_Y}\right)$$

Co-expression pair of lncRNA-gene was screened out with the threshold of |correlation coefficient|>0.5 and FDR<0.05 to analyze the regulation role of lncRNApalyed on oncogenes and TSGs.[22]

**Table 1.** The top 10 most enriched up and downregulated lncRNAs.

| Down_regulated | logFC | AveExpr | P.Value | adj.P.Val |
|---|---|---|---|---|
| C14orf132 | −1463.05 | 491.5892 | 2.95E-99 | 2.58E-97 |
| SFTA1P | −1061.46 | 143.3066 | 8.72E-153 | 1.53E-150 |
| MALAT1 | −831.21 | 2824.661 | 0.038047 | 0.097914 |
| IPW | −154.206 | 261.4601 | 2.58E-06 | 1.46E-05 |
| HCG11 | −95.327 | 295.2437 | 0.00017 | 0.000675 |
| C10orf95 | −62.8257 | 22.00978 | 2.72E-42 | 7.94E-41 |
| CASC2 | −51.757 | 34.08992 | 4.99E-40 | 1.25E-38 |
| HCG26 | −41.9228 | 63.43647 | 5.44E-07 | 3.53E-06 |
| MGC27382 | −25.3669 | 4.139212 | 2.17E-62 | 9.49E-61 |
| INE2 | −9.15558 | 7.585014 | 6.68E-10 | 6.15E-09 |
| up_regulated | logFC | AveExpr | P.Value | Adj.P.Val |
| TUG1 | 1672.372 | 4458.745 | 4.68E-09 | 3.90E-08 |
| GAS5 | 1102.298 | 2215.222 | 1.12E-05 | 5.30E-05 |
| SNHG1 | 940.4861 | 1150.743 | 1.07E-18 | 2.33E-17 |
| MIAT | 865.0085 | 1007.605 | 0.006695 | 0.020554 |
| SNHG5 | 574.2911 | 1803.889 | 0.006891 | 0.020791 |
| SNHG6 | 516.5869 | 1115.43 | 1.87E-06 | 1.13E-05 |
| TP53TG1 | 168.8525 | 387.5029 | 5.51E-06 | 2.92E-05 |
| PVT1 | 138.7953 | 159.0165 | 4.58E-16 | 8.90E-15 |
| SNHG12 | 136.07 | 307.6558 | 1.55E-05 | 6.80E-05 |
| SNHG11 | 100.5078 | 203.8896 | 3.76E-12 | 5.49E-11 |

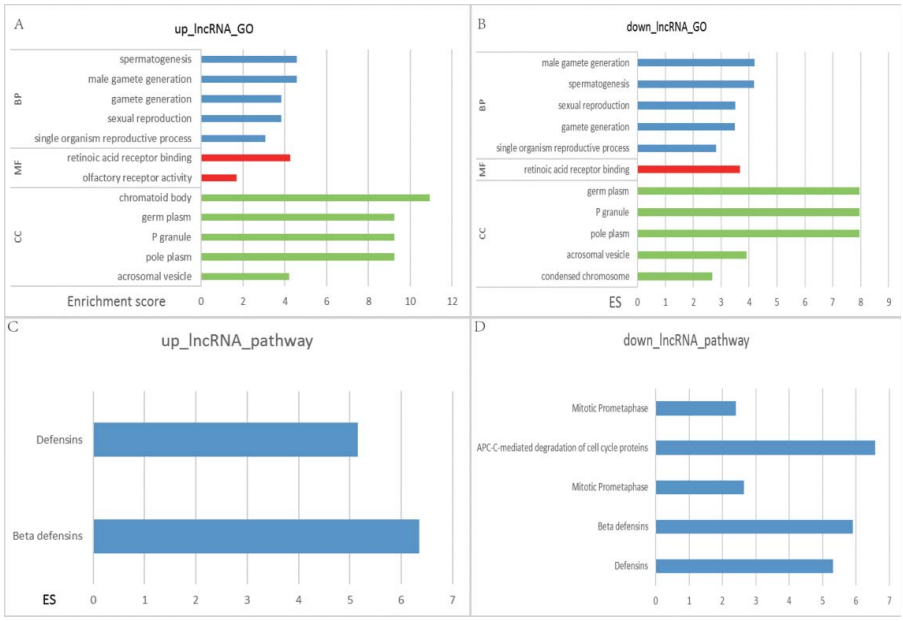


**Figure 1.** The top 5 pathways of GO and KEGG enrichment analysis in differentially expressed lncRNA. The horizontal axis included the score of enrichment while the vertical axis represented the enriched pathways. GO annotation of upregulated lncRNA was displayed in (A) while the downregulated ones were in (B). (C) was the KEGG annotation of upregulated lncRNA while (D) was that of downregulated lncRNA. The higher the enrichment score was, the more the pathways were involved in cancer regulation.

## Results

### *Identification of DEIs*

DEIs were identified by Limma package with the thresholds of FDR<0.05 and |logFC|>0. As a result, a toal of 32 downregulated lncRNA and 37 upregulated lncRNA were screened out (Table 1). Multiple of the top 10 listed lncRNAs of all the up-and-downregulated genes was extremely significant, indicating expression of these lncRNAs were quite different between cancer samples and normal ones. Full information of DEIs were in Table S1 and results of TSG difference in detail were listed in Table S2 while the difference of oncogene in detail was listed in Table S3.

### *GO and KEGG annotation of lncRNAs*

There were 45 biologic processes in mostly enriched GO annotation of upregulated lncRNA, along with 11 cellular_component, 2molecular_function and 2 KEGG annotations (Fig. 1). The most enriched GO terms were spermatogenesis while the most enriched KEGG wasdefensins. There were 58 biologic processes in the mostly enriched GO terms of lncRNA, as well

**Table 2.** The top 10 most enriched genes of TSGs.

| Up_regulated | logFC | AveExpr | P.Value | adj.P.Val |
|---|---|---|---|---|
| KRT19 | 28845.27 | 33309.9 | 2.72E-14 | 1.05E-13 |
| JUP | 17128.21 | 20583.57 | 2.68E-21 | 1.49E-20 |
| SFN | 14604.06 | 15739.99 | 2.21E-12 | 7.32E-12 |
| P4HB | 14486.82 | 28042.46 | 3.14E-14 | 1.21E-13 |
| S100A2 | 14480.6 | 13260.58 | 5.90E-07 | 1.34E-06 |
| CLDN1 | 13579.75 | 13894.82 | 9.24E-07 | 2.07E-06 |
| IGFBP3 | 11991.54 | 12730.01 | 5.97E-06 | 1.26E-05 |
| S100A11 | 11224.27 | 18430.29 | 1.23E-13 | 4.55E-13 |
| HSP90B1 | 10926.5 | 22550.73 | 5.72E-16 | 2.40E-15 |
| TP63 | 10801.41 | 10001.16 | 3.78E-16 | 1.63E-15 |
| Down_regulated | logFC | AveExpr | P.Value | adj.P.Val |
| B2M | −61028.7 | 57288.05 | 1.70E-24 | 1.07E-23 |
| EPAS1 | −43960.2 | 9017.274 | 9.54E-152 | 1.34E-149 |
| AHNAK | −39444.6 | 25397.52 | 4.28E-54 | 6.11E-53 |
| VIM | −34186.1 | 17043.41 | 2.63E-72 | 5.76E-71 |
| TIMP3 | −29060.2 | 10151.26 | 5.52E-58 | 8.41E-57 |
| EMP2 | −26385.1 | 6813.104 | 1.23E-170 | 2.88E-168 |
| CAV1 | −24996.5 | 5456.84 | 6.22E-147 | 7.93E-145 |
| GPX3 | −22076.3 | 4341.227 | 1.65E-134 | 1.44E-132 |
| SPTBN1 | −20717.6 | 8233.457 | 1.47E-161 | 2.28E-159 |
| TXNIP | −20001.2 | 9212.048 | 3.83E-71 | 7.77E-70 |

**Table 3.** The top 10 up and downregulated oncogenes.

| Up_regulated | logFC | AveExpr | P.Value | adj.P.Val |
|---|---|---|---|---|
| COL1A1 | 55834.14 | 74270.7 | 3.93E-05 | 7.82E-05 |
| HSP90AB1 | 15214.92 | 32400.67 | 8.79E-14 | 3.66E-13 |
| LAMC2 | 11714.45 | 14225.67 | 0.000332 | 0.000591 |
| TFRC | 11713.08 | 17040.72 | 1.66E-05 | 3.45E-05 |
| HSP90AA1 | 10399.82 | 27118.23 | 3.11E-10 | 1.04E-09 |
| CALR | 8826.698 | 24686.2 | 7.66E-12 | 2.77E-11 |
| GPX2 | 6511.432 | 5962.844 | 1.25E-09 | 3.85E-09 |
| GNAS | 6483.241 | 24242.39 | 2.23E-05 | 4.51E-05 |
| HMGA1 | 5950.289 | 6397.132 | 3.91E-32 | 3.87E-31 |
| ATP2A2 | 5798.927 | 10919.86 | 4.03E-20 | 2.28E-19 |
| down_regulated | logFC | AveExpr | P.Value | adj.P.Val |
| CD74 | −99484 | 49787.68 | 7.05E-62 | 1.47E-60 |
| SLC34A2 | −61364.2 | 11025.63 | 3.18E-137 | 6.17E-135 |
| FOS | −19955.7 | 6364.295 | 1.14E-52 | 2.01E-51 |
| MSN | −12400.6 | 13715.76 | 6.16E-37 | 7.33E-36 |
| SDC4 | −9841.3 | 7256.907 | 9.19E-26 | 6.96E-25 |
| ALDH2 | −9295.91 | 3688.583 | 2.70E-105 | 1.05E-103 |
| HLA-A | −8937.56 | 26264.03 | 0.000358 | 0.000634 |
| EZR | −7375.86 | 9158.441 | 9.26E-30 | 7.71E-29 |
| MYH11 | −6842.01 | 4572.208 | 1.61E-05 | 3.35E-05 |
| LMNA | −5908.48 | 10217.69 | 7.28E-13 | 2.83E-12 |

as 20 cellular_component and one molecular_function. Of all the 8 KEGG annotations, the most significant GO term was spermatogenesis while the most significant KEGG term was APC-C-mediated degradation. The 2 terms were both GO annotation of spermatogenesis, indicating that these significant lncRNAs were quite different from those normal samples, thus affecting the regulation of reproduction and inheritance process. lncRNAs in these pathways were involved in regulation of cellular cycles and pathways of defensins.

### *Identification of differentially expressed oncogene and TSGs*

A total of 518 downregulated TSGs and 503 upregulated TSGs were screened out by Limma package with the thresholds of FDR<0.05 and |logFC|>1. The top 10 most genes enriched in TSGs were list in Table 2 Similarly, there were 177 upregulated oncogenes and 234 downregulated oncogenes (Table 3). Heatmap of the top 100 differentially expressed TSGs and oncogeneswas shown in Fig. 2 and 3, respectively, indicating that there was significant difference of cancer related genes between normal samples and tumor specimens.

### *GSEA analysis of differentially expressed oncogenes*

No upregulated KEGG gene set but 3 downregulated ones were screened out with the threshold of NOM p-val <0.25. downregulated oncogene was mainly enriched in pathways as *CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION, KEGG_MAPK_SIGNALING_PATHWAY* and *KEGG_JAK_-*

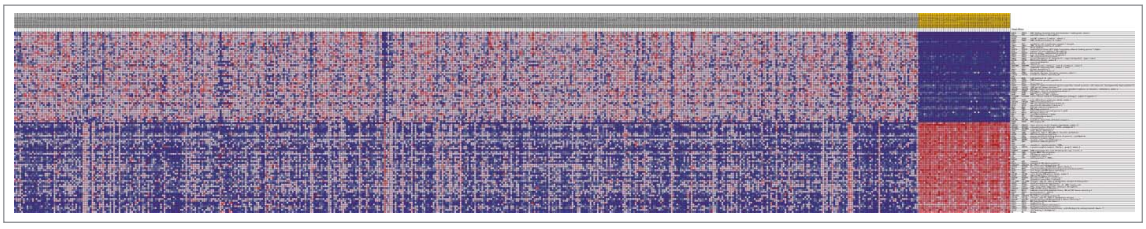

**Figure 2.** Heatmap of the top 100 differentially expressed TSGs between normal samples and tumor specimens. The horizontal axis were different samples, the gray ones were tumor samples while the yellow ones were normal specimens. The vertical axis was genes, which indicated that there was obvious difference among normal samples and tumor ones.

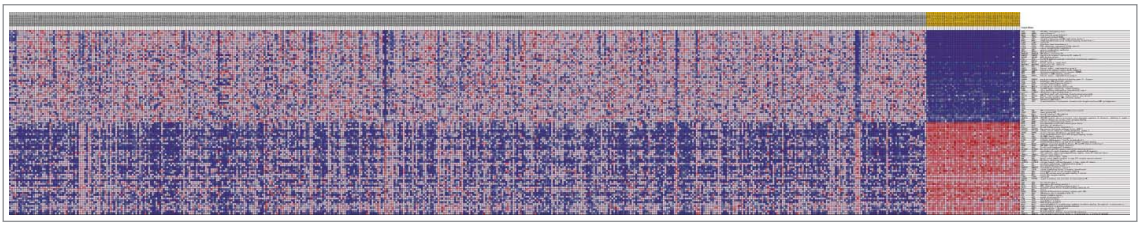

**Figure 3.** The top 100 differentially exprssed oncogenes between normal samples and tumor specimens. The horizontal axis was samples whie the gray ones were tumor samples and the yellow ones were normal specimens. The vertical axis stands for different genes.

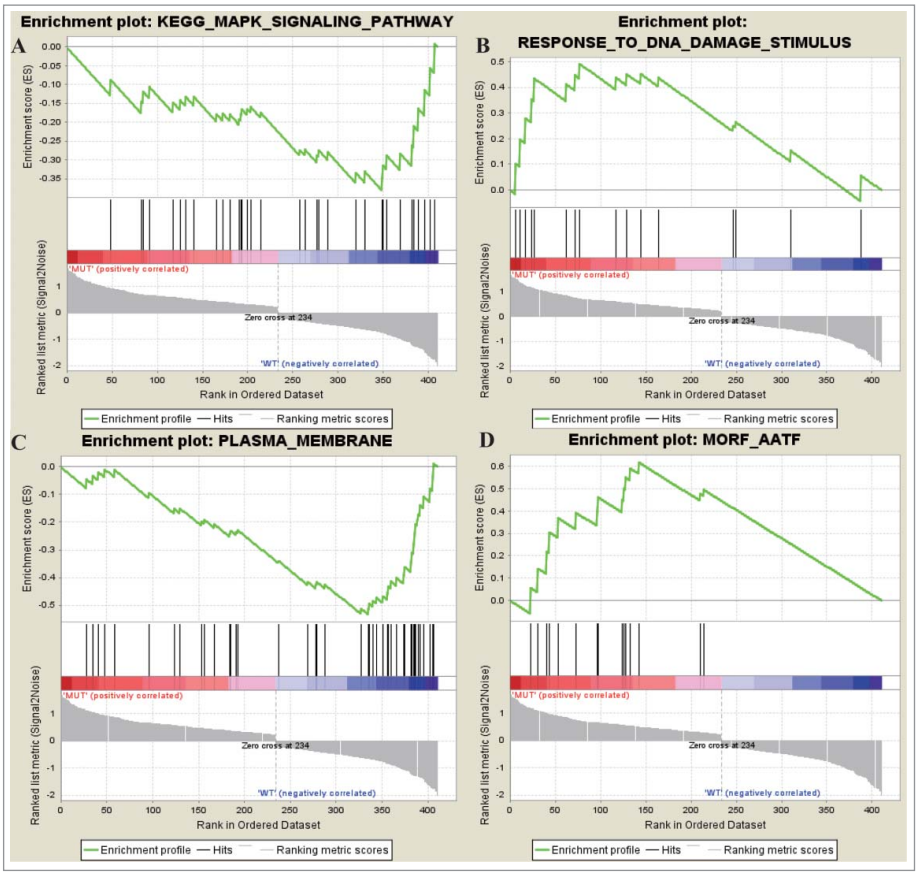


**Figure 4.** The most enriched data set of differentially expressed up and downregulated oncogenes. (A) represents the downregulated genes in KEGG_MAPK_SIGNALING_PATHWAY data set and (B) represented how many genes were upregulated in GO data set RESPONSE_TO_DNA_DAMAGE_STIMULUSj. (C) represents the downregulated genes in PLASMA_MEMBRANE data set and (D) represented how many genes were upregulated in MORF_AATF.

*STAT_SIGNALING_PATHWAY*.A total of 3 upregulated GO gene sets and 14 downregulated ones were enriched. The 3 upregulated Go gene sets were *RESPONSE_TO_DNA_DAMAGE_STIMULUS, DNA_METABOLIC_PROCESS, RESPONSE_TO_ENDOGENOUS_STIMULUS*, while the downregulated ones were *PLASMA_MEMBRANE, PLASMA_MEMBRANE_PART, MEMBRANE*. There were 14 gene sets in upregulated cancer neighborhood gene while there was no gene set in downregulated genes. Genes sets of upregulated cancer neighborhood gene were *MORF_AATF, MORF_DEK, MORF_UBE2I*. On the other hand, upregulated gene of differentially expressed oncogene were-*RESPONSE_TO_DNA_DAMAGE_STIMULUS* and *MORF_AATF* (Fig. 4). AATF was a kind of transcription factor against apoptosis process, indicating that the main cause in the occurrence of NCSC was DNA damage and cell obtained the characteristic of anti-apoptosis.

### GSEA analysis of differentially expressed TSGs

Two upregulated KEGG gene sets and 5 downregulated KEGG gene sets were enriched. The upregulated gene sets were *KEGG_CELL_CYCLE* and *KEGG_P53_SIGNALING_PATHWAY* while the downregulated ones were *KEGG_TIGHT_JUNCTION, KEGG_LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION*, and *KEGG_MAPK_SIGNALING_PATHWAY*.A total of 17 upregulated GO terms and 25 downregulated GO terms were screened out. The upregulated terms were *RESPONSE_TO_DNA_DAMAGE_STIMULUS,RESPONSE_TO_ENDOGENOUS_STIMULUS, DNA_METABOLIC_PROCESS* while the downregulated ones were *PLASMA_MEMBRANE, INTRINSIC_TO_MEMBRANE, INTEGRAL_TO_MEMBRANE*.

### Cancer neighborhood gene sets

A total of 24 upregulated cancer neighborhood gene sets and 2 downregulated gene sets were figured out. The upregulated genes sets of cancer neighborhood gene were *MORF_BUB3, MORF_EIF3S2* and *MORF_XRCC5* while the downregulated ones were *GCM_MAP4K4* and *MORF_CDC2L5*.

The most significant terms in TSG were mainly enriched in *KEGG_CELL_CYCLE*and *RESPONSE_TO_DNA_DAMAGE_STIMULUS* (Fig. 5). Genes were downregulated in *KEGG_TIGHT_JUNCTION* and *PLASMA_MEMBRANE*.

Since cells in LSCC samples appeared with the characteristic of keratosis and intercellular bridge, including spindle cell carcinoma, pathways associated with the structure of cell membrane could become abnormal. Growth and division of cells could be constant and hereditary substance became variated after malignant transformation process, which could stimulate the overexpression of suppressor genes in cells.

### Co-expression of DEIs and TSG or oncogenes

Co-efficient of up-or-downregulated lncRNA with TSGs or oncogene was calculated out respectively by R function. The

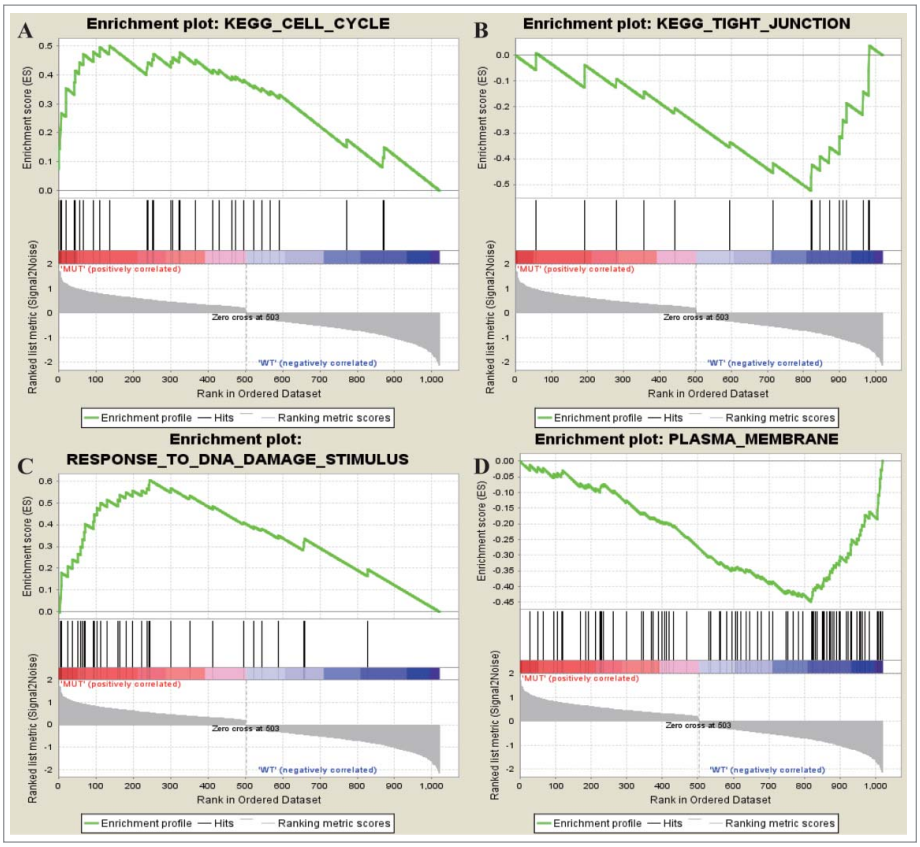


**Figure 5.** The most enriched data sets of tumor supressor gene. (A) represented upregulated genes in KEGG_CELL_CYCLE and (B) represented downregulated genes in KEGG_TIGHT_JUNCTION. (C) represented upregulated genes in RESPONSE_TO_DNA_DAMAGE_STIMULUS while (D) represented how many downregulated genes in PLASMA_MEMBRANE.

thresholds of 0.5and −0.5 was used to screen out the co-expression pair of lncRNAs-genes.*SFTA1P* appeared to be the most co-expressed gene between downregulated lncRNAs and oncogenes (Table 4). There were 7 co-expression genes, that were*PTPRB, SLC34A2, FGR,ROS1,CSF3R, CXCL2, RND1. CACNA1D*was found to be the most co-expression oncogenes-with lncRNAs, while all these co-expression oncogenes were downregulated.There were 2 co-expressed genes, *KCNH8* and *SOX2*, between upregulated lncRNAs and oncogenes and *MIAT* was the one with the most co-expression. *PABPC1L* was found to be the oncogene co-expressing with more than one lncRNAs.

There were 3 genes with most co-expression between downregulated lncRNAs and TSGs, *SFTA1P* with the 16 co-expression genes, while *C10orf95* with 8 co-expression genes and *CASC2* with 7 co-expression genes (Table 5). All the TSGs, that is,*GADD45B, CSRNP1, RHOBTB2, DOK2* were all downregulated.Four co-expression genes in *DLEU1*, 5 co-expression genes in *GAS5* and 4 co-expression genes in *SNHG12*. The 3 TSGs, *RPL15, SELT, RBX1*, were all overexpressed.

**Table 4.** Co-expression pairs of lncRNA and oncogenes.

| down_lncRNA | oncogene | cor_value | up_lncRNA | oncogene | cor_value |
|---|---|---|---|---|---|
| SFTA1P | PTPRB | 0.606947516 | MIAT | KCNH8 | 0.57164 |
| SFTA1P | SLC34A2 | 0.585553018 | MIAT | SOX2 | 0.501148 |
| SFTA1P | PRAM1 | 0.621849534 | HCG18 | TRIM27 | 0.550524 |
| SFTA1P | FGR | 0.522879132 | SNHG12 | PABPC1L | 0.529173 |
| SFTA1P | ROS1 | 0.526725187 | C9orf163 | RALGDS | 0.506619 |
| SFTA1P | CSF3R | 0.537765065 | TUG1 | PATZ1 | 0.536315 |
| SFTA1P | CXCL2 | 0.723380194 | SNHG3 | PABPC1L | 0.57944 |
| SFTA1P | RND1 | 0.563852798 | SNHG4 | KIAA1598 | 0.503864 |
| HCG26 | HLA-A | 0.506811902 | SNHG11 | PABPC1L | 0.542648 |
| HCG26 | TNFRSF14 | 0.579783753 | | | |
| HCG26 | WAS | 0.503636572 | | | |
| CASC2 | ECT2L | 0.608696582 | | | |
| RMST | ASPSCR1 | 0.527844789 | | | |
| KIAA0087 | CACNA1D | 0.628973394 | | | |
| DKFZp779M0652 | CACNA1D | 0.676408943 | | | |
| C9orf106 | CACNA1D | 0.654643601 | | | |

Table 5. Co-expression pairs between lncRNA and TSGs.

| down_lncRNA | TSG | cor_value | up_lncRNA | TSG | cor_value |
|---|---|---|---|---|---|
| SFTA1P | SCGB3A1 | 0.557159 | DLEU1 | RPL15 | 0.505824 |
| SFTA1P | LEFTY2 | 0.668367 | DLEU1 | SELT | 0.572985 |
| SFTA1P | GADD45B | 0.682329 | DLEU1 | RBX1 | 0.542194 |
| SFTA1P | CITED2 | 0.510908 | DLEU1 | NDUFA13 | 0.508297 |
| SFTA1P | BCL6B | 0.544449 | GAS5 | RPL15 | 0.544716 |
| SFTA1P | CSRNP1 | 0.662952 | GAS5 | EEF1A1 | 0.505638 |
| SFTA1P | RHOBTB2 | 0.565762 | GAS5 | HINT1 | 0.512073 |
| SFTA1P | DOK2 | 0.51381 | GAS5 | TOMM20 | 0.508488 |
| SFTA1P | COL4A3 | 0.655631 | GAS5 | DLEU1 | 0.505154 |
| SFTA1P | RPS6KA2 | 0.530706 | SNHG12 | RBM6 | 0.593483 |
| SFTA1P | ADAMTS8 | 0.750708 | SNHG12 | ARGLU1 | 0.523387 |
| SFTA1P | SUSD2 | 0.573774 | SNHG12 | RBM5 | 0.559451 |
| SFTA1P | EDNRB | 0.581291 | SNHG12 | RASSF1 | 0.516762 |
| SFTA1P | SELENBP1 | 0.716833 | | | |
| SFTA1P | CDH5 | 0.67137 | | | |
| SFTA1P | GKN2 | 0.820088 | | | |
| CASC2 | CASC1 | 0.682196 | | | |
| CASC2 | PACRG | 0.588571 | | | |
| CASC2 | KCNRG | 0.603464 | | | |
| CASC2 | SNTN | 0.594486 | | | |
| CASC2 | ZMYND10 | 0.645154 | | | |
| CASC2 | OSCP1 | 0.715601 | | | |
| CASC2 | C2orf40 | 0.590077 | | | |

## Discussion

Long non-coding RNA (lncRNA), is a kind of RNAs whose length larger than 200 bases with biologic functions. LncRNAs can suppress translation by biding to their corresponding mRNAs and that's the main reason why they have an essential role in gene regulation. It's involved in the growth and pathological process by the way of chromatin reprogramming, cis regulation at enhancers or post-trtanscriptional regulation, which was similar with way of the encodinggenes. Since the close relation between gene expression patterns and tumor subtypes was essential in understanding the molecular basis of tumorigenesis,[23, 24] the 2 differentially expressed IncRNAs between normal samples and patients with SCLC we identified in this study, SFTA1P and CASC2, may of much value in clinical management of the disease.

There were reports about abnormal expression of some IncRNA in the occurrence, invasion and development in many types of tumor tissues, indicating their association with the suppression of cancer. The expression degree of these IncRNAs may reflect the degree of prognosis process. Given their specialty in cancer tissues, these IncRNAs were regarded as prognostic biomarkers of cancer disease.

In this study, spermatogenesis was the most enriched GO terms in both up and downregulated IncRNAs. We speculated that there was much difference of differentially expressed IncRNAs between normal samples and cancer specimens, which thus affected the regulation of reproduction and inheritance process. Besides, the results of pathway analysis revealed that these IncRNAs were also involved in the pathways of cell cycle regulation and alexin.

Functional analysis of oncogenes revealed that they had large effect in signaling pathways associated with cytokine and DNA damage, indicating that there was large variation in genetic material of NSCC and activation of signaling pathways led to the deterioration of the cancer.

Two IncRNA, SFTA1P and CASC2 were found to be associated with most of tumor supressor genes and oncogenes and had an great effect on the regulation of LSCC. It was reported that SFTA1P was downregulated in NSCC and its main function was lying on epidermal growth, cell attachment and response to DNA damage.[25] The other IncRNA, CASC2, was reported to be inhibited by miR-21 in gliomas and,[26, 27] which could be the ceRNA of other oncogenes.

It was reported that lncRNAs could in cis regulate the expression of their neighboring genes,[4, 28] we suspected the effect of these 2 genes may be related with this pathway, which needed more relevant research.

In summary, we identify 2 potential biomarkers, that is, SFTA1P and CASC2, associated with the regulation and development of lung squamous cell carcinoma, which could provide more specific and accurateprognostic and predictive indicators to clinical outcome of LSCC patients,implying their application in clinical diagnosis and treatment of this disease.

## Conclusions

In summary, we identify 2 potential biomarkers, that is, SFTA1P and CASC2, associated with the regulation and development of lung squamous cell carcinoma, which could provide more specific and accurate prognostic and predictive indicators to clinical outcome of LSCC patients, implying their application in clinical diagnosis and treatment of this disease.

## Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

## Author contributions

GQ.H, ZP.K and B.G conceived and designed the experiments; GQ.H, ZP.K performed the experiments; ZP.K and B.G analyzed the data; Y.C contributed analysis tools; GQ.H, ZP.K, B.G and HB.H wrote the paper.

## References


1. Network C. Comprehensive genomic characterization of squamous cell lung cancers. Nature 2012; 489(7417): 519-25; PMID:22960745; http://dx.doi.org/10.1038/nature11404
2. Chiba I, Takahashi T, Nau MM, D'Amico D, Curiel DT, Mitsudomi T, Buchhagen DL, Carbone D, Piantadosi S, Koga H, et al. Mutations in the p53 gene are frequent in primary, resected non-small cell lung cancer. lung cancer study group. Oncogene 1990; 5(10): 1603-10; PMID:1979160.
3. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Research 2012; 22(9): 1775-1789; PMID:22955988; http://dx.doi.org/10.1101/gr.132159.111
4. Wang K, Chang H. Molecular mechanisms of long noncoding RNAs. Molecular cell 2011; 43(6): 904-14; PMID:21925379; http://dx.doi.org/10.1016/j.molcel.2011.08.018
5. Gupta R, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 2010; 464(7291): 1071-76; PMID:20393566; http://dx.doi.org/10.1038/nature08975
6. Wilusz J, Sunwoo H, Spector D. Long noncoding RNAs: functional surprises from the RNA world. Genes & Development 2009; 23(13): 1494-1504; http://dx.doi.org/10.1101/gad.1800909
7. Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, Cui H. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. Nature 2008; 451(7175): 202-206; PMID:18185590; http://dx.doi.org/10.1038/nature06468
8. Wang N, Sanborn Z, Arnett KL, Bayston LJ, Liao W, Proby CM, Leigh IM, Collisson EA, Gordon PB, Jakkula L, et al. Loss-of-function mutations in Notch receptors in cutaneous and lung squamous cell carcinoma. Proceedings of the National Academy of Sciences 2011; 108 (43): 17761-66; ; http://dx.doi.org/10.1073/pnas.1114669108
9. Guffanti A, et al. A transcriptional sketch of a primary human breast cancer by 454 deep sequencing. BMC Genomics 2009; 10(1): 1; PMID:19121221; http://dx.doi.org/10.1186/1471-2164-10-163
10. Yamada K, Kano J, Tsunoda H, Yoshikawa H, Okubo C, Ishiyama T, Noguchi M. Phenotypic characterization of endometrial stromal sarcoma of the uterus. Cancer Science 2006; 97(2): 106-12; PMID:16441420; http://dx.doi.org/10.1111/j.1349-7006.2006.00147.x
11. Lin R, Maeda S, Liu C, Karin M, Edgington TS. A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. Oncogene 2007; 26(6): 851-58; PMID:16878148; http://dx.doi.org/10.1038/sj.onc.1209846
12. Luo J, Ren B, Keryanov S, Tseng GC, Rao UN, Monga SP, Strom S, Demetris AJ, Nalesnik M, Yu YP, et al. Transcriptomic and genomic analysis of human hepatocellular carcinomas and hepatoblastomas. Hepatology 2006; 44(4): 1012-24; PMID:17006932; http://dx.doi.org/10.1002/hep.21328
13. He X, et al. Low expression of long noncoding RNA CASC2 indicates a poor prognosis and regulates cell proliferation in non-small cell lung cancer. Tumor Biology 2016; 37(7):9503-9510.
14. Du Z, et al. Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory network in prostate cancer. Nature Communications 2016; 7:10982; http://dx.doi.org/10.1038/ncomms10982
15. Di Martino M, Guzzi PH, Caracciolo D, Agnelli L, Neri A, Walker BA, Morgan GJ, Cannataro M, Tassone P, Tagliaferri P, et al. Integrated analysis of microRNAs, transcription factors and target genes expression discloses a specific molecular architecture of hyperdiploid multiple myeloma. Oncotarget 2015; 6(22): 19132; PMID:26056083; http://dx.doi.org/10.18632/oncotarget.4302
16. Ritchie M, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 2015; 43(7):e47; http://dx.doi.org/10.1093/nar/gkv007
17. Jiang Q, et al. LncRNA2Function: a comprehensive resource for functional investigation of human lncRNAs based on RNA-seq data. BMC Genomics 2015; 16(3): 1. PMID:25553907; http://dx.doi.org/10.1186/1471-2164-16-S3-S2
18. Futreal P, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. A census of human cancer genes. Nat Rev Cancer 2004; 4(3): 177-83; PMID:14993899; http://dx.doi.org/10.1038/nrc1299
19. Chen J, Hung WS, Chan HH, Tsai SJ, Sun HS. In silico identification of oncogenic potential of fyn-related kinase in hepatocellular carcinoma. Bioinformatics 2013; 29(4): 420-7; PMID:23267173; http://dx.doi.org/10.1093/bioinformatics/bts715
20. Mootha V, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, et al. PGC-1 [alpha]-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 2003; 34 (3): 267-73; PMID:12808457; http://dx.doi.org/10.1038/ng1180
21. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; 102 (43): 15545-50; PMID:16199517; http://dx.doi.org/10.1073/pnas.0506580102
22. Yang F, Lyu S, Dong S, Liu Y, Zhang X, Wang O. Expression profile analysis of long noncoding RNA in HER-2-enriched subtype breast cancer by next-generation sequencing and bioinformatics. OncoTargets and Therapy 2016; 9: 761; PMID:26929647; http://dx.doi.org/10.2147/OTT.S97664
23. Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proceedings of the National Academy of Sciences 2003; 100(14): 8418-23; PMID:12829800; http://dx.doi.org/10.1073/pnas.0932692100
24. Sørlie T, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences 2001; 98(19): 10869-74; PMID:11553815; http://dx.doi.org/10.1073/pnas.191367098
25. Zhao W, Luo J, Jiao S. Comprehensive characterization of cancer subtype associated long non-coding RNAs and their clinical implications. Scientific Reports 2014; 4: 6591; PMID:25307233; http://dx.doi.org/10.1038/srep06591
26. Wang P, Liu YH, Yao YL, Li Z, Li ZQ, Ma J, Xue YX. Long non-coding RNA CASC2 suppresses malignancy in human gliomas by miR-21. Cellular Signalling 2015; 27(2): 275-82; PMID:25446261; http://dx.doi.org/10.1016/j.cellsig.2014.11.011
27. Zhang J, Zhu N, Chen X. A novel long noncoding RNA LINC01133 is upregulated in lung squamous cell cancer and predicts survival. Tumor Biology 2015; 36(10): 7465-1; PMID:25908174; http://dx.doi.org/10.1007/s13277-015-3460-9
28. Dimitrova N, Zamudio JR, Jong RM, Soukup D, Resnick R, Sarma K, Ward AJ, Raj A, Lee JT, Sharp PA, et al. LincRNA-p21 activates p21 in cis to promote polycomb target gene expression and to enforce the G1/S checkpoint. Molecular Cell 2014; 54(5): 777-90; PMID:24857549; http://dx.doi.org/10.1016/j.molcel.2014.04.025