

Identification of alternative splicing and lncRNA genes in pathogenesis of small cell lung cancer based on their RNA sequencing

Yuming Lei^{A,D}, Yunfei Shi^{A,B}, Jin Duan^{B,C}, Yinqiang Liu^{C,E}, Guoli Lv^{C,E}, Rou Shi^{B,E}, Fajun Zhang^{B,E}, Qingmei Yang^{B,C}, Wei Zhao^{A,F}

The First Affiliated Hospital of Kunming Medical University, China

A – research concept and design; B – collection and/or assembly of data; C – data analysis and interpretation;

D – writing the article; E – critical revision of the article; F – final approval of the article

Advances in Clinical and Experimental Medicine, ISSN 1899–5276 (print), ISSN 2451–2680 (online)

Adv Clin Exp Med. 2019;28(8):1043–1050

Address for correspondence

Wei Zhao

E-mail: weizhaowz6@sohu.com

Funding sources

The study was supported by the Scientific Research Projects of Institutions of Medical and Health Institutions in Yunnan Province

The Role of NOD-like Receptors and Inflammatory Bodies in the Development of Xuanwei Lung Cancer (project No. 2016NS017); Number of Training Objects for Leaders of Medical Disciplines in Yunnan Province (project No. D2017013); Doctoral Research Fund of the First Affiliated Hospital of Kunming Medical University

The research about PDGFRB functions on lung squamous cell carcinoma progression and its potential usage as a clinical lung squamous cell carcinoma maker.

Conflict of interest

None declared

Received on July 28, 2017

Reviewed on October 15, 2017

Accepted on August 17, 2018

Published online on August 8, 2019

Cite as

Lei Y, Shi Y, Duan J, et al. Identification of alternative splicing and lncRNA genes in pathogenesis of small cell lung cancer based on their RNA sequencing. *Adv Clin Exp Med.* 2019;28(8):1043–1050. doi:10.17219/acem/94392

DOI

10.17219/acem/94392

Copyright

Copyright by Author(s)

This is an article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Abstract

Background. The molecular mechanisms involved in small-cell lung cancer (SCLC) are largely unknown. Recent studies have suggested that long non-coding RNAs (lncRNAs) are likely to play a critical role.

Objectives. There is an urgent need for suitable molecular biomarkers for SCLC diagnosis and for assessing patient prognosis.

Material and methods. In this study, we used public databases to identify mRNA-like candidate lncRNAs. A multi-step computational approach was used to construct a functional SCLC lncRNAs-mediated competing with endogenous RNA (ceRNA) network (LMCN) by integrating genome-wide lncRNAs and mRNA expression profiles, miRNA-target interactions, functional analyses, and clinical survival analyses.

Results. The results revealed the significance of lncRNAs interactions with ceRNAs in SCLC, indicating that integration of expression profiles and alternative splicing could be used to identify biomarkers and the underlying pathological changes. The following genes: *EPB41L4A-AS1*, *HOXA-AS2*, *XIST*, *DLEU2*, *FGD5-AS1*, *ALMS1-IT1*, *SNHG12*, *MIR17HG*, *MIR4720*, and *SCARNA10* in cluster, as well as shared alternative splicing events, were considered to be critical genes.

Conclusions. Olfactory transduction and endocytosis were the top-enriched pathways in SCLC. The selected cluster, including critical genes, might also be a potential pathway of SCLC pathogenesis. As a result, this research provides the perspective information to explore the potential critical genes and its pathways in SCLC therapy.

Key words: alternative splicing, small-cell lung cancer (SCLC), critical genes

Introduction

Small-cell lung cancer (SCLC) is an aggressive neuroendocrine carcinoma. Lung cancer is the leading cause of cancer-related deaths worldwide. About 15% of lung cancer cases are SCLC cases.¹ Its prognosis is poor, and has a trend toward metastasizing to the brain at an early stage, despite advances in treatment modalities that have improved survival.² Histological diagnosis can provide information valuable in planning the treatment program, but it is insufficient for predicting clinical outcomes.³ At present, suitable molecular biomarkers for SCLC diagnosis are needed to assess the patient's prognosis.

Recently, noncoding RNAs (ncRNAs) have received considerable attention due to their key role in affecting shared gene expression programs.⁴ These RNAs are usually microRNAs or long-non-coding RNAs (lncRNAs). With the development of high-throughput sequencing technology, ncRNAs have been discovered in a wide range of biological processes. However, the functions of lncRNAs in SCLC are not well characterized and the identification of lncRNA biomarkers is challenging. Typically, lncRNA functions are characterized using a "guilt by association" strategy.⁵ Recent studies have demonstrated that lncRNAs compete with endogenous RNAs (ceRNAs) in regulating expression level of other transcripts.^{6,7} For example, the lncRNA HULC acts as an endogenous ceRNA and plays an important regulatory role in lung cancer.⁸ In addition, a recent study using network analysis identified a lncRNA-associated ceRNA network (across 12 different cancers) and prognostic markers of lncRNAs.⁷ Another study proposed that the ceRNA interaction network of SCLC could reveal canonical oncogenic pathways.⁹ Collectively, this data underscores the importance of lncRNAs interaction with ceRNAs and indicates that the integration of expression profiles with network analysis could identify the risk of lncRNAs and the underlying tumor pathology.

Current studies have suggested that the alternative splicing of the RNAs, as an essential layer of gene expression regulation that expands the cell proteome, plays complex and critical role in breast cancer.¹⁰ Additionally, the alternative splicing of *EGFR*, *CD44*, *PIK3C3*, *RRAS2*, *MAPKAP1*, and *FGFR2* was shown to raise the risk of stratification in non-small cell lung cancer (NSCLC) patients.¹¹ Alternative splicing of *S6K1* gene can regulate the mammalian target of rapamycin (mTOR) complex 1 signaling pathway. This has been strongly linked with the onset and progression of NSCLC. Alternative placing of this gene may, therefore, constitute a relevant and innovative biomarker. The short isoforms of *S6K1* are activated mTORC1, leading to increased 4E-BP1 phosphorylation.^{12,13}

In this research, a multi-step computational approach was used for constructing a functional lncRNAs-mediated ceRNA network (LMCN) in SCLC. We utilized the high-throughput profiles of SCLC samples, which were collected from the ArrayExpress database. Intergradations

of alternative splicing, genome-wide lncRNA/mRNA expression profiles, comprehensive miRNA-target interactions, and functional analyses were entailed in the systematic analysis. Based on LMCN, we found that several lncRNAs exhibited specific topological features, which are consistent with coding mRNAs regulatory associated with SCLC pathology. The alternative RNA splicing involved in RNA processing significantly related with cell cycle functions was also identified. Additionally, by exploring a competitive sub-network with in the LMCN we determined that MCM3AP-AS and MIR17HG were involved in a ceRNA regulatory module and synergistically competed with *MATR3*, *XPO1* and *ZCCHC14*. These analyses demonstrated that the integration of alternative splicing profile with cancer-associated lncRNA-mediated ceRNA networks could be used to promote the discovery of molecular biomarkers and SCLC therapeutics.

Methods

Data recruitment and preprocess

Sequence data was deposited in Array Express, E-GEOD-60052 (<https://www.omicsdi.org/dataset/arrayexpress-repository/E-GEOD-60052>), including 79 SCLC patients and 7 normal lung tissues. This dataset was processed on the Illumina HiSeq 2000 platform (Illumina, Inc., San Diego, USA). For sequencing, samples for non-coding gene datasets of SCLS patients and controls were prepared using Illumina mRNA-Seq kit (Part #RS-100–0801) (Illumina, Inc.).

For data preprocess, Illumina reads were trimmed for adaptor sequence, and low-complexity or low-quality sequence was removed, then mapped to Human genome reference hg19 via <http://bowtie-bio.sourceforge.net/bowtie2/index.shtml> (data was normalized using DESeq in R package to analyse count data for differential expression(<https://bioconductor.org/packages/release/bioc/html/DESeq.html>). For data analysis, RNA-seq data was analyzed using the set of open source software programs of the Tuxedo suite: TopHat v. 1.3.3 and Cufflinks (<http://ccb.jhu.edu/software/tophat/tutorial.shtml> and <http://cole-trapnell-lab.github.io/cufflinks/>).^{14,15} The R platform available as free software (<http://cran.r-project.org/>) was applied. In brief, PolyAplus RNA (RNA transcripts) was purified from 100 ng of total RNA with oligo-dT beads, and then the samples were fragmented with divalent cations under high temperature. First strand synthesis was performed with random hexamer and reverse transcriptase, and 2nd strand synthesis with RNase H and DNA polymerase I. Following cDNA synthesis, the double-stranded products were end-repaired, a single "A" was added and then the Illumina Paired End Adapters (https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/experiment-design/

illumina-adapter-sequences-1000000002694-09.pdf) were ligated on to the cDNA products. The ligation products were purified using gel electrophoresis. The target size range for these libraries was ~300 bp, such that the final library for sequencing would have cDNA inserts with sizes of ~200 bp long. One run of 2×50 bp paired-end sequencing was performed on the HiSeq2000 instrument, using 1 lane per tissue, to produce approx. 80 million read pairs per tissue.

Reconstructing transcriptomes

To determine splice variants within each tissue, we aligned reads to the hg19 genome using TopHat v. 1.3.3 with an intron database, which extracted from the UCSC Known Genes data set (<http://genome.ucsc.edu>). Aligned reads were then assembled into transcript fragments using Cufflinks software (with parameters '-F 0.05'). Then, the Cuffcompare (<http://cufflinks.cbcb.umd.edu/>) was used to compare these transfrags to the Ensembl annotation. After eliminating likely artifacts and assemblies not associated with Ensembl genes, the Cuffdiff was used to redistribute reads with a high-confidence transcript sets obtained.^{15,16} Cuffcompare classifies assembled transcripts into multiple categories: the known coding genes, the known noncoding genes, the potentially novel transcripts, and the undefinable transcripts. The potentially novel transcripts and the undefinable transcripts were used for downstream analysis. FPKM expression level for this set was then re-estimated from the original alignments using Cuffdiff.

Discovery of alternative splicing events

To determine alternative splicing events in the LMCN, we developed a software package that can identify the average shortest (AS) profile and analyse the exons-involved in 1 transcript in the transcripts profiles. We restricted the analysis to Ensembl genes with FPKM ≥ 0.1 re-estimated using Cuffdiff as described above. We define the exon skipping event as the pair between exon-containing splice form and exon-skipping splice form, where the boundaries of flanking introns are required to match precisely. To determine the novel events, the exons and spanning introns were compared against the annotation data sets (UCSC Genes and Ensembl).

Comparison of alternative splicing events among groups

For each event, we calculated the exon inclusion ratio $R = \text{FPKMon} / (\text{FPKMon} + \text{FPKMoff})$ for each tissue. To account for minor differences in the annotation of splice junctions, the differentially expressed level of average shortest (AS) mRNA was calculated with limma package in R. All differentially expressed mRNA across groups data were

further screened by limma package based on Bayes method (<http://www.bioconductor.org/>). After processing using the top table method of limma package, the differentially expressed genes were analyzed. Then, they were compared with AS genes. The AS genes and DEG were selected for further network assay. The procedure was as follows: lncRNA, miRNA and mRNA expression profiles were obtained from starBase v. 2.0 (<http://starbase.sysu.edu.cn/starbase2/index.php>). Interactions between miRNA-lncRNAs were identified using existing miRNA target prediction methods. Putative miRNA-lncRNA interactions were identified using the miRanda algorithm with the default parameters. The miRNA target binding sites were predicted on entire lncRNAs. The data of the miRNA-mRNA interactions was downloaded from miRNA reference databases, the TarBase (v. 6, <http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=tarbasev6/index>) and the mirTarBase (v. 4.5, <http://mirtarbase.mbc.nctu.edu.tw/php/download.php>). Both databases store manually curated collections of experimentally supported miRNA targets.

Functional enrichment analysis

The gene ontology based on gene ontology database and the pathway enrichment analysis based on the database of KEGG pathway were carried out using Fisher's test. P-value < 0.05 and false discovery rate (FDR) < 0.05 were considered to be statistically significant.

Results

Integrated analysis of gene signatures

After preprocessing the expression profile of mRNA datasets, we identified 3,105 differentially expressed genes of SCLA using limma package in R language. Limma is an R/Bioconductor software package that provides an integrated solution for analyzing data from gene expression experiments. Over the past decade, limma has been a popular choice for gene discovery through differential expression analyses of microarray and high-throughput PCR data.¹⁶ Figure 1A shows the log fold changes (log FC) in top 10 gene signatures. Figure 1B shows the volcano plot of these genes (log odds). Log FC value (Fig. 1A) indicated that the upregulated genes were *SNAP25* and *MANEAL*; the downregulated genes were *AGER*, *EPAS1*, *EMP2*, *RP 11-114H24.5*, *HBA2*, *NFKBIA*, *ABCA3*, and *CA4*, and they were considered as suppressor genes in SCLC (Table 1).

Construction of lncRNAs-mediated ceRNA network

For the network, the Pearson's correlation coefficient of proteins encoded by lncRNA gene signatures and their

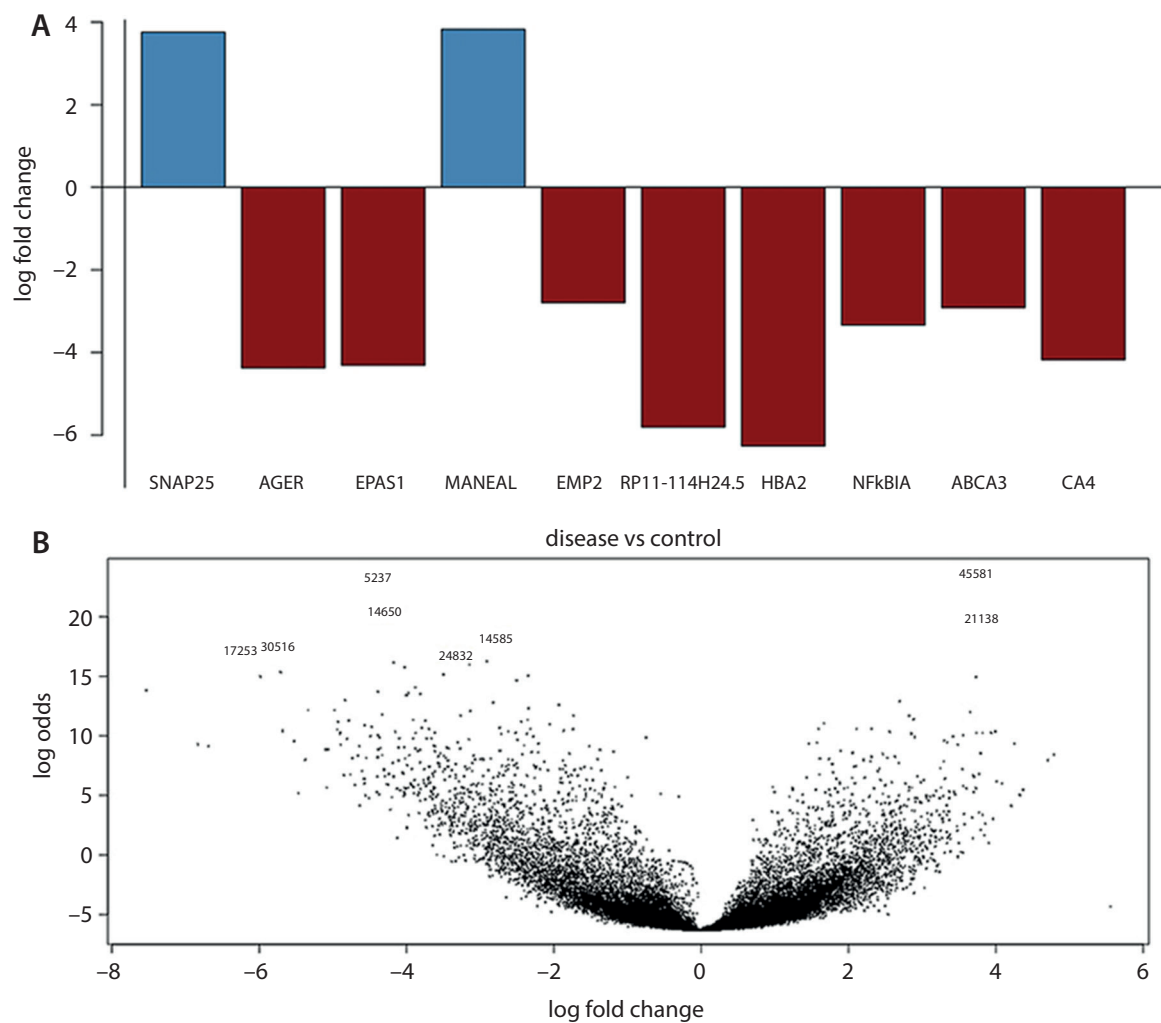


Fig. 1. A – stacked bar plots of fold change for top 10 ranked genes in microarray datasets of small cell lung cancer; B – volcano plot of top 10 ranked genes

mRNA mediated by ceRNA was calculated. The constructed LMCN network (Fig. 2) included 4,987 nodes and 15,579 edges, where nodes represented genes and edges between nodes indicated gene interactions in the network.

Identification of the cluster in LMCN

While the LMCN could provide a global view of all possible competing ceRNA interactions that could be used to investigate the regulatory properties of the lncRNAs, the partial sub-networks revealed a more detailed pattern of how

Table 1. The top 10 gene signatures identified by gene expression profile in small cell lung cancer

Gene	Log FC	AveExpr	t	p-value	Adjusted p-value	B
<i>SNAP25</i>	3.756787	5.864792	9.581308	3.68E-15	1.38E-10	23.67952
<i>AGER</i>	-4.37513	5.578326	-9.49856	5.40E-15	1.38E-10	23.31936
<i>EPAS1</i>	-4.31265	7.866335	-8.82805	1.23E-13	2.09E-09	20.39216
<i>MANEAL</i>	3.825731	6.030139	8.700969	2.21E-13	2.83E-09	19.83646
<i>EMP2</i>	-2.79616	7.953902	-8.32082	1.30E-12	1.33E-08	18.17466
<i>RP11-114H24.5</i>	-5.80775	2.128777	-8.17396	2.56E-12	2.19E-08	17.53344
<i>HBA2</i>	-6.26852	2.269522	-8.13152	3.12E-12	2.28E-08	17.34829
<i>NFKBIA</i>	-3.33741	6.38071	-7.98145	6.26E-12	4.00E-08	16.69415
<i>ABCA3</i>	-2.91447	7.701587	-7.87974	1.00E-11	5.69E-08	16.25149
<i>CA4</i>	-4.17853	1.285196	-7.84961	1.15E-11	5.88E-08	16.12048

Log FC – logarithmic fold change; AveExpr – the average expression; t – two-sample t-test; B – potential scenario for the statistical test for differential expression.



Fig. 2. LMCN of gene signatures in small cell lung cancer. Purple nodes refer to transcription factors (TFs), red ones refer to the common genes in the cluster

the lncRNAs synergized with competing mRNAs. We derived a cluster from the LMCN by applying a Pearson’s correlation coefficient threshold >0.9 (Fig. 3). The genes with a top degree were *EPB41L4A-AS1*, *HOXA-AS2*, *XIST*, *DLEU2*, *FGD5-AS1*, *ALMS1-IT1*, *SNHG12*, *MIR17HG*, *MIR4720*, and *SCARNA10* (Table 2).

Functional enrichment analysis

Functional enrichment analysis of the obtained gene signatures was carried out, including Gene Ontology (GO) term enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. The GO analysis was carried out in 3 categories, including biological processes (BP), molecular functions (MF) and cellular components (CC). The results of GO analysis showed that cluster genes were significantly enriched in vacuolar transport, Golgi vesicle

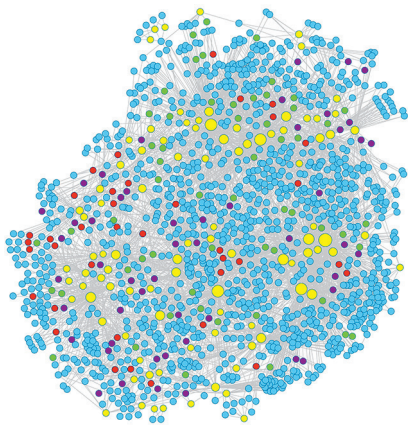


Fig. 3. Gene signatures in the cluster. Purple nodes refer to TFs, red ones refer to the common genes in the cluster 1 and KEGG pathway of oxidative phosphorylation, Parkinson’s disease, Huntington’s disease, and Alzheimer’s disease

transport, mitotic cell cycle phase transition, endosomal transport, and cell cycle phase transition. After the false discovery rate was calculated, the most significant 10 GO terms including BP, MF and CC were presented in Table 3. For KEGG pathway enrichment, 4 terms were obtained and the most significant ones were Olfactory transduction [PATH:hsa04740], Neuroactive ligand-receptor interaction [PATH:hsa04080], Cell cycle [PATH:hsa04110], Endocytosis [PATH:hsa04144], and Ubiquitin mediated proteolysis [PATH:hsa04120] (Table 4).

Identification of the hub genes with average shortest events in LMCN

The genes in the LMCN-shared high degree were *HOXA-AS2*, *EPB41L4A-AS1*, *DLEU2*, *XIST*, *MIR17HG*,

Table 2. The average centrality value of cluster

Gene	Degree	Average shortest path length	Closeness centrality	Neighborhood connectivity	Stress	Topological coefficient
NEAT1	38	3.47019424	0.28816831	3.39473684	1481190	0.07483553
FOXP1	6	3.64567984	0.27429726	50.83333333	518298	0.20507545
STX12	5	3.88948426	0.2571035	34.4	310398	0.21410256
KLF7	5	3.68318821	0.27150391	52.8	344072	0.23981481
ITPR1	5	3.55860683	0.28100885	69.2	422280	0.22434211
ATF7IP2	4	3.78767582	0.26401415	52.25	181166	0.29454023
CPEB4	4	3.61888814	0.27632797	71.75	222772	0.26498127
LDLR	4	3.561286	0.28079744	83.25	244082	0.26791531
NCKAP5	4	3.75954454	0.26598967	69	311492	0.29310345
VPS37B	4	3.7635633	0.26570564	59.5	157274	0.28398058
TNKS1BP1	4	3.66175486	0.2730931	73.25	381286	0.27471483
TANC2	4	4.02478232	0.24846064	31.75	46432	0.27702703
CSF1	4	3.73945077	0.26741895	53.75	223642	0.26641414
ADARB1	4	3.81982585	0.26179204	46.5	139536	0.27409639
RBMS1	4	3.71131949	0.26944595	65.75	395634	0.28650442

Table 3. Significant terms of GO enrichment

GO ID	Description	Gene ratio	Bg ratio	p-value	Adjusted p-value	Count
GO:0007034	vacuolar transport	129/4397	272/16672	5.20E-14	1.55E-10	129
GO:0048193	Golgi vesicle transport	141/4397	307/16672	8.54E-14	1.55E-10	141
GO:0044772	mitotic cell cycle phase transition	190/4397	450/16672	1.13E-13	1.55E-10	190
GO:0016197	endosomal transport	118/4397	244/16672	1.14E-13	1.55E-10	118
GO:0044770	cell cycle phase transition	198/4397	477/16672	2.45E-13	2.68E-10	198
GO:0006914	autophagy	176/4397	428/16672	1.37E-11	1.24E-08	176
GO:0016482	cytosolic transport	69/4397	131/16672	1.25E-10	9.76E-08	69
GO:0072657	protein localization to membrane	183/4397	478/16672	5.03E-09	3.44E-06	183
GO:0006892	post-Golgi vesicle-mediated transport	37/4397	60/16672	9.30E-09	5.65E-06	37
GO:0000082	G1/S transition of mitotic cell cycle	98/4397	227/16672	2.69E-08	1.47E-05	98

GO ID – Gene Ontology identification; Bg ratio – background ratio indicates the number of target genes regulated by the transcription factor; Count – indicates how many reads map to each feature.

Table 4. Significant KEGG pathways

Pathways	p-value	FDR
4740 Olfactory transduction [PATH:hsa04740]	7.88E-48	2.28E-45
4080 Neuroactive ligand-receptor interaction [PATH:hsa04080]	7.15E-16	1.04E-13
4110 Cell cycle [PATH:hsa04110]	7.34E-10	7.10E-08
4144 Endocytosis [PATH:hsa04144]	3.49E-08	2.53E-06
4120 Ubiquitin mediated proteolysis [PATH:hsa04120]	3.61E-07	2.09E-05
5204 Chemical carcinogenesis [PATH:hsa05204]	7.28E-07	3.52E-05
982 Drug metabolism – cytochrome P450 [PATH:hsa00982]	2.92E-06	8.61E-05
4668 TNF signaling pathway [PATH:hsa04668]	2.97E-06	8.61E-05
4114 Oocyte meiosis [PATH:hsa04114]	2.54E-06	8.61E-05
4914 Progesterone-mediated oocyte maturation [PATH:hsa04914]	2.18E-06	8.61E-05

FDR – false discovery rate.

FGD5-AS1, *SNHG12*, *ALMS1-IT1*, *ZRANB2-AS2*, and *MIR4720*. In the obtained cluster, the average rank value based on p-value was calculated and genes with the highest value were considered as hub genes. The result was shown in Table 2: genes in a cluster with an average degree value of 6.6 were connected significant. This is consistent with our earlier AS analysis. The average shortest events (AS events) present in several genes (*NEAT1*, *FOXP1*, *STX12*, *KLF7*, *ITPR1*) in the cluster of cell cycle pathways further confirmed the significance of hub genes.

Discussion

Pathway analysis of small-cell lung cancer pathogenesis

In the current study, based on LMCN and pathway enrichment of gene signatures, several KEGG pathways were found. Some of these pathways have already been reported by other researchers, such as cell cycle and endocytosis.^{17–19} One of the significant pathways was endocytosis

with the p-value of 3.49E-08. Endocytosis is a form of active transport and an important process for the regulation of cell signal transduction, cellular dynamics and cell-to-cell communication. Endocytosis and its counterpart, exocytosis, take place in all cells, which include pinocytosis and phagocytosis. The current study showed that in this pathway ligand-induced endocytosis requires NM IIA to induce the degradation of EGFR, causing the endocytosis process.²⁰ Our present study verified the endocytosis process in cancer and predicted that the interaction of *ALMS1-IT1* and *VPS37B* as a novel mechanism in trans-Golgi network (Fig. 4). *VPS37B* is located in chromosome 12. This gene binds to *TSG101* at several different sites, which function in vacuolar protein sorting. Several other genes are also in the pathway list, e.g., *GIT2*, *ZFYVE9*, *HSPA6*, *RAB11FIP4*, *CHMP2B*, *TGFB2*, *RAB4A*, *RAB5A*, *RK5*, and others. Some of them are lncRNAs, such as *XIST*, *TRAF3IP2-AS1*, *DLEU2*, *DLEU2*, *MIR4720*, and *SNHG12*. These genes were analyzed via the LMCN, indicating that they were involved in all the 4 main functions (receptor-mediated endocytosis, caveolae, macropinocytosis, and phagocytosis). They were considered as the critical genes. Therefore, endocytosis pathway as well as these genes was thought to be very important in the pathogenesis of SCLC. The map of OXPHOS pathway was shown in Fig. 4.

Critical genes in pathogenesis of SCLC

The genes in the cluster were considered as critical genes for the pathogenesis of SCLC. These results strongly support SRSF1 as a prognostic biomarker in SCLC and provide a rationale for personalized therapy in SCLC. The critical genes with AS events in this study included *NEAT1*,

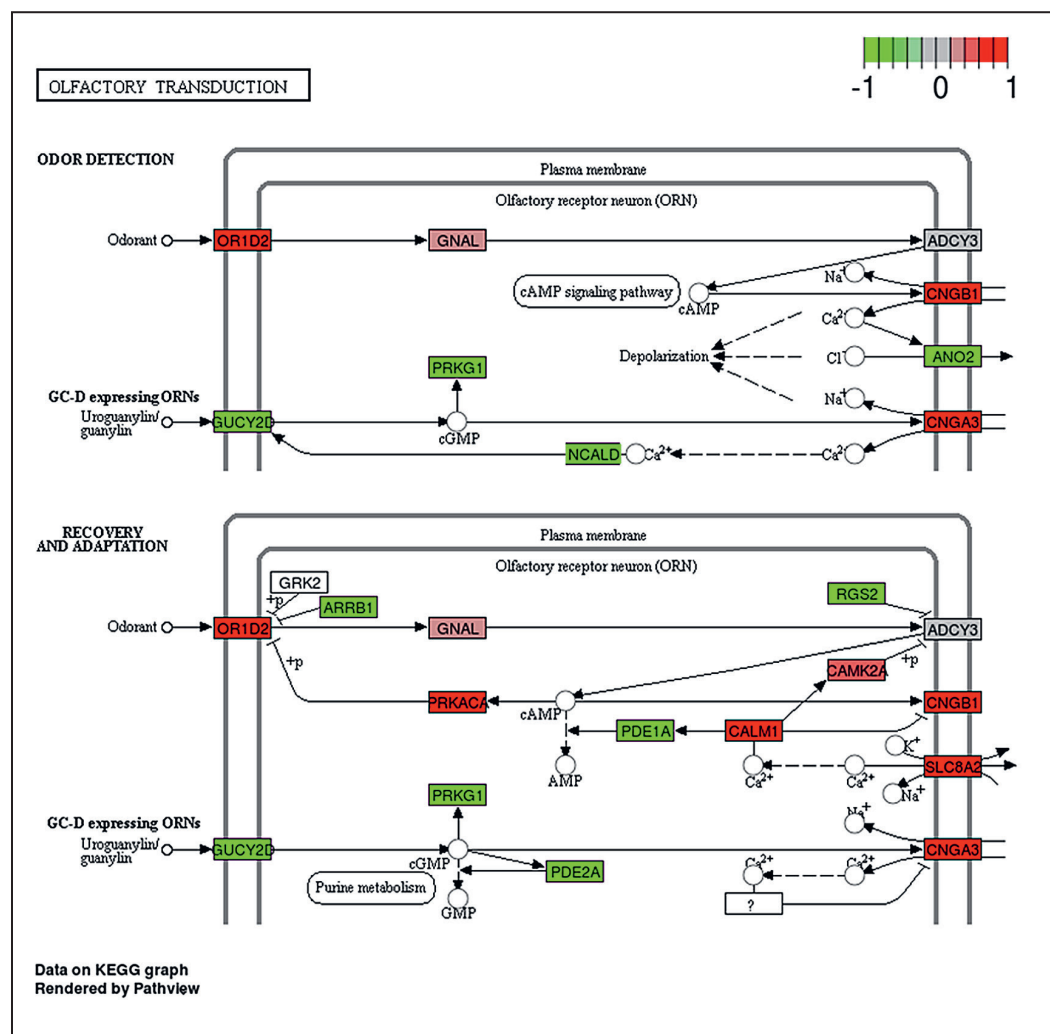


Fig. 4. The map of oxidative phosphorylation pathway

FOXPI, *STX12*, *KLF7*, and *ITPR1*. Among these genes, *NEAT1* and *FOXPI* were reported to be associated with cancer. Nuclear-enriched abundant transcript 1 (*NEAT1*, non-protein coding) produces a lncRNA, which is transcribed from the multiple endocrine neoplasia locus. This lncRNA is retained in the nucleus where it forms the core structural component of the paraspeckle sub-organelles. Its high expressions are associated with poor prognosis in different types of solid tumors, such as lung cancer, esophageal cancer, colorectal cancer, and hepatocellular carcinoma.²¹ It was aberrantly upregulated and may act as a transcriptional regulator for numerous genes, including some genes involved in unfavorable prognosis in cancer patients.²² Increased serine/arginine splicing factor 1 (*SRSF1*) DNA copy number and mRNA overexpression has been linked to poor prognosis for SCLC. Functional studies in vitro and in vivo demonstrate that *SRSF1* may play a central role in DNA repair and chemosensitivity.²³ On the other hand, downregulated *NEAT1* promotes leucocyte differentiation in acute promyelocytic leukemia. Our study examined the relationship between *NEAT1* and SCLC and confirmed that the defection of this gene was closely related to the pathogenesis of SCLC.

FOXPI gene belongs to the subfamily P of the Forkhead box (FOX) transcription factor family. Forkhead box transcription factors play an important role in the regulation of tissue- and cell type-specific gene transcription during both development and adulthood. This gene may express in several types of human malignant tumors. It also has been reportedly associated with the metastasis and patient prognosis in NSCLC.²⁴ The expression of *FoxP1* in NSCLC was significantly higher. Since the *FOXPI* protein contains both DNA-binding- and protein-protein binding-domains, it was reported to be a regulator in cancer cells. The low expression of *FOXPI* was independent factors for predicting the poor prognosis for NSCLC.²⁴ Alternative splicing of *FOXPI* results in multiple transcript variants encoding different isoforms, which is also related to many other cancers.²⁵ *NEAT1* and *FOXPI* genes connected closely in the cluster and might be another potential way for the pathogenesis of SCLC. The other genes, *STX12*, *KLF7* and *ITPR1*, had not been reported to be in any relationship with SCLC and they need to be investigated further.

In this study, a comprehensive process of data in datasets of SCLC for protein–protein interactions (PPI) network was conducted. Some clusters where the genes connected

closely to SCLC were selected. Then, the function and signaling pathways for SCLC were presented systematically. Finally, the potential pathways and 2 critical genes in SCLC pathogenesis were identified. These may aid in SCLC therapy. However, this study was based on the bioinformatics reports. These results need to be further proven with laboratory tests.

References

1. Taromi S, Catusse J, Elverfeldt DV, et al. PO-33-Heparin suppresses progression of small cell lung cancer (SCLC) in an orthotopic mouse model. *Thromb Res*. 2016;140(Suppl 1):S188.
2. Liu Y, Liu YS, Wu PF, et al. Brain microvascular endothelium induced-annexin A1 secretion contributes to small cell lung cancer brain metastasis. *Int J Biochem Cell Biol*. 2015;66:11–19.
3. Verhaak RG, Hoadley KA, Purdom E, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010;17(1):98–110.
4. Guttman M, Donaghey J, Carey BW, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*. 2011;477(7364):295–300.
5. Guo Q, Cheng Y, Liang T, et al. Comprehensive analysis of lncRNA-mRNA co-expression patterns identifies immune-associated lncRNA biomarkers in ovarian cancer malignant progression. *Sci Rep*. 2015;5:17683.
6. Ebert MS, Sharp PA. Emerging roles for natural microRNA sponges. *Curr Biol*. 2010;20(19):R858–R861.
7. Wang P, Ning S, Zhang Y, et al. Identification of lncRNA-associated competing triplets reveals global patterns and prognostic markers for cancer. *Nucleic Acids Res*. 2015;43(7):3478–3489.
8. 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(16):5366–5383.
9. Sumazin P, Yang X, Chiu HS, et al. An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell*. 2011;147(2):370–381.
10. Pradella D, Naro C, Sette C, Ghigna C. EMT and stemness: Flexible processes tuned by alternative splicing in development and cancer progression. *Mol Cancer*. 2017;16(1):8.
11. Li Y, Sun N, Lu Z, et al. Prognostic alternative mRNA splicing signature in non-small cell lung cancer. *Cancer Lett*. 2017;393:40–51.
12. Huertas CS, Dominguez-Zotes S, Lechuga LM. Analysis of alternative splicing events for cancer diagnosis using a multiplexing nanophotonic biosensor. *Sci Rep*. 2017;7:41368.
13. Mei H, Wang Y, Fan J, Lin Z. Alternative splicing of S6K1 promotes non-small cell lung cancer survival. *Tumour Biol*. 2016;37(10):13369–13376.
14. Ghosh S, Chan CK. Analysis of RNA-seq data using TopHat and Cufflinks. *Methods Mol Biol*. 2016;1374:339–361.
15. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7(3):562–578.
16. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
17. Rainero E. Extracellular matrix endocytosis in controlling matrix turnover and beyond: Emerging roles in cancer. *Biochem Soc Trans*. 2016;44(5):1347–1354.
18. Vassie JA, Whitelock JM, Lord MS. Endocytosis of cerium oxide nanoparticles and modulation of reactive oxygen species in human ovarian and colon cancer cells. *Acta Biomater*. 2017;50:127–141.
19. Yu XW, Lin S, Du HZ, et al. Synergistic combination of DT-13 and topotecan inhibits human gastric cancer via myosin IIA-induced endocytosis of EGF receptor in vitro and in vivo. *Oncotarget*. 2016;7(22):32990–33003.
20. Kim JH, Wang A, Conti MA, Adelstein RS. Nonmuscle myosin II is required for internalization of the epidermal growth factor receptor and modulation of downstream signaling. *J Biol Chem*. 2012;287(33):27345–27358.
21. Yu X, Li Z, Zheng H, Chan MT, Wu WK. NEAT1: A novel cancer-related long non-coding RNA. *Cell Prolif*. 2017;50(2). doi:10.1111/cpr.12329
22. Barry G, Briggs JA, Hwang DW, et al. The long non-coding RNA NEAT1 is responsive to neuronal activity and is associated with hyperexcitability states. *Sci Rep*. 2017;7:40127.
23. Jiang L, Huang J, Higgs BW, et al. Genomic landscape survey identifies SRSF1 as a key oncogene in small cell lung cancer. *PLoS Genet*. 2016;12(4):e1005895.
24. Feng J, Zhang X, Zhu H, Wang X, Ni S, Huang J. High expression of FoxP1 is associated with improved survival in patients with non-small cell lung cancer. *Am J Clin Pathol*. 2012;138(2):230–235.
25. Santos ME, Athanasiadis A, Leitao AB, DuPasquier L, Sucena E. Alternative splicing and gene duplication in the evolution of the FoxP gene subfamily. *Mol Biol Evol*. 2011;28(1):237–247.