Mathematical Biosciences and Engineering

MBE, 17(2): 1593–1603.
DOI: 10.3934/mbe.2020083
Received: 09 May 2019
Accepted: 20 October 2019
Published: 09 December 2019

http://www.aimspress.com/journal/MBE

Research article

The signature lncRNAs associated with the lung adenocarcinoma patients prognosis

Yong Ding[1], **Jian-Hong Liu**[2,*]

[1] Department of Medical Oncology, Shaoxing Hospital of Traditional Chinese Medicine, Shaoxing 312000, China
[2] Department of Respiration, Zhejiang Jinhua Guangfu Hospital, Jinhua 321000, China

* **Correspondence:** Email: ljh2019509@163.com; Tel: +8613735662269.

Abstract: Long non-coding RNAs (lncRNAs) consist of over 200 nucleotides and are not translated into proteins. Previous studies have shown the importance of lncRNAs in the development of the lung adenocarcinoma (LUAD). Emergence of the high-throughput sequencing technology has led to the identification of a lot of lncRNAs which plays important roles in various biological events. Increasingly evidence has revealed that aberrant lncRNAs expression is related to the development of various cancers. We analyzed the RNA-seq data of 551 lung adenocarcinoma patients downloaded from The Cancer Genome Atlas (TCGA). By analyzing the pre-cancerous and cancer tissues of the relevant patient transcriptomes, we discovered the significant lncRNAs associated with the lung adenocarcinoma. Based on their median score, the prognosis of the patients was categorized as either poor or favorable. Univariate and multivariate COX analysis were used to further analyze the differential lncRNAs. Co-expression and gene enrichment analysis were performed to further investigate the function of the related lncRNA. RNA-seq data analysis led to the discovery of the lncRNA, OGFRP1 as an interesting factor involved in the lung adenocarcinoma. Therefore, lncRNAs play an important role in the clinical diagnosis and the treatment of the lung adenocarcinoma.

Keywords: lncRNAs; lung adenocarcinoma; TCGA

Abbreviations: TCGA: The Cancer Genome Atlas; lncRNA: long non-coding RNA; ROC: receiver operating characteristic

1. Introduction

According to the world health organization (WHO), in the 21st century, the lung adenocarcinoma had the highest mortality rate (31.5%) amongst the malignant tumors. In the past 20 years, the incidence of the non-small cell lung cancer has continued to increase. Although the immune checkpoint inhibitors, such as drugs for the PD-1 and the PD-L1 blockade, as the treatment for the lung adenocarcinoma made some progress in the recent years, the 5-year survival rate is still less than 15% [1,2]. Till date, the treatment for the lung adenocarcinoma is still based primarily on surgery, supplemented by radiotherapy, chemotherapy and targeted therapy. The survival rate of most patients is still less than 20%, and the prognosis is poor [3]. Therefore, the search for new biomarkers and therapeutic targets is imminent.

Long non-coding RNAs (lncRNAs) are defined as non-protein-coding RNA transcripts having nucleotides lengths which are greater than 200 bp. Previously, it was thought that lncRNAs are only transcriptional noise and does not have biological functions [4,5]. However, with new studies, it has been found that lncRNAs, take part in a multitude of functions involved in epigenetics, nuclear transport, transcriptional regulation and post-transcriptional regulation of genes at both the cis and trans levels [6,7]. lncRNA is differentially expressed in various tumors, such as the osteosarcoma, malignant glioma, lung cancer, etc. It is closely related to the process of tumor proliferation, invasion, metastasis, occurrence and development of tumors. Drug sensitivity is directly related to the type and levels of the tissue’s lncRNAs. Therefore, further exploration of the role of lncRNA in tumor resistance has important clinical value [8].

The present study aims to build an lncRNA risk score to refine LUAD patient’s prognostic classification by analyzing the LUAD patients from The Cancer Genome Atlas (TCGA) project [9]. Furthermore, the study provides significant information regarding the potential target of the prognostic lncRNAs in cis/trans/ceRNA regulation.

2. Materials and methods

The RNA-seq data and the corresponding clinical information of the LUAD patients in the TCGA project were downloaded from the Genomic Data Commons Data Portal (portal.gdc.cancer.gov/). Based on the GENCODE project long non-coding RNA annotation file (version 26, GRCh38), we obtained the Reads Per Kilobase per Million (RPKM) mapped reads expression data of the lncRNAs from level 3 RNA-seq data of the TCGA. The differentially expressed lncRNAs were calculated using edgeR by comparing the lncRNA expression of a specific class of normal samples against the tumor samples at selection cut-off fold change > 1 and false discovery rate < 0.05 [10].

The univariable Cox regression analysis calculated between the lncRNA gene expression is presented as log2 (RPKM + 1). LncRNAs were considered significantly correlated with the patient survival at a threshold of $p < 0.0001$ and false discovery rate < 0.05. The random survival forests variable hunting (RSFVH) algorithm was carried out to minimize the selection of the unrelated lncRNAs. The error rate in the RSFVH model was calculated by 1,000 permutation runs. Then, the selected lncRNAs were subjected to multivariate Cox regression analysis and the risk score was found by estimating the regression coefficients in the multivariable Cox regression and expression of the lncRNAs. The median risk score was selected in the training set as a cut-off categorizing the patients into a low-risk or a high-risk group. The Kaplan-Meier method was used to compare the survival difference between the low-risk and high-risk groups in the training and the validation sets. Multivariate Cox and Kaplan-Meier survival analysis of the cytogenetically stratified LUAD patients were used to identify the lncRNA's expression signature as an independent prognostic factor. The sensitivity and specificity of the lncRNAs expression signature was evaluated by the area under the receiver operating characteristic (AUROC) of five years. We computed the spearman correlation coefficient between the mRNA and the lncRNA expression levels. The mRNA-lncRNA paired with the absolute spearman correlation coefficient > 0.5 and P < 0.0001 in both TCGA projects, were selected for the co-expression network construction. Gene Ontology (GO) is a standardized gene function classification system that provides a dynamically updated standard vocabulary to fully describe the properties of the genes and the gene products in organisms. GO has three ontologies and describes genes as 1) molecular component, 2) cellular component, and 3) biological processes. The basic unit of the GO is term (term, node), and each word corresponds to a property [11]. GO functional analysis of the differentially expressed genes is used to find out about the significant functional enrichment analysis of differentially expressed genes in vivo, in which different genes coordinate their biological functions, and analysis is based on the signaling pathways to understand the biological functions of the genes [12]. We used KEGG, which is a commonly used database of the signaling pathways, to find out about the Pathway Significant Enrichment Analysis, which is a hypergeometric test used to find pathways that are significantly enriched in the differentially expressed genes compared to the entire genomic background [13].

3. Result

We downloaded 551 data-sets from the TCGA and analyzed the differential expression of the lncRNAs. The number of the normal samples was 54, and the tumor samples were 497. The heatmap technology was used to figure out the differences in the lncRNAs between the normal samples and the tumor samples. The heatmap uses a color code to quantify results, higher expressed genes were red in color while the lower expressed genes were green as shown in (Figure 1).

The Kaplan-Meier plot was used for the survival analysis of the lncRNAs from the clinical data downloaded from the TCGA. We analyzed the different survival curves to figure out the lncRNA's difference in survival. We identified 89 potential significant lncRNA's in our survival analysis (p < 0.05, FDR < 0.05) in the TCGA project. The survival curves of all lncRNAs were plotted for subsequent exploration of the effects of different lncRNAs on survival (Figure 2).

We used univariate and multivariable COX analysis to predict the risk score of the lung adenocarcinoma and plot the forest picture of the multivariate. Using a multivariate Cox regression model, we constructed a formula based on the expression level of the lncRNA models of the 551 TCGA lung adenocarcinoma samples. The Cox regression of the long non-coding RNAs in the TCGA is shown in Figure 3.

As shown previously, we also observed that OGFRP1 is expressed in several different lncRNAs and six lncRNAs signatures predicted the survival of the LUAD patients in the training set. We constructed a formula according to the expression level of the six lncRNAs in the 551 TCGA LUAD samples using the multivariate Cox regression model as follows: (0.097355021 x expression level of FAM83A-AS1) + (0.2879 x expression level of WWC2-AS2) + (0.367100193 x expression level of OGFRP1). Figure 4 show that patients with low-risk scores express high levels of the protective lncRNAs, while patients with the high-risk scores express high levels of the risky lncRNAs. Using the median of the risk score as the cut-off point, we predicted the lncRNAs risk for the LUAD.

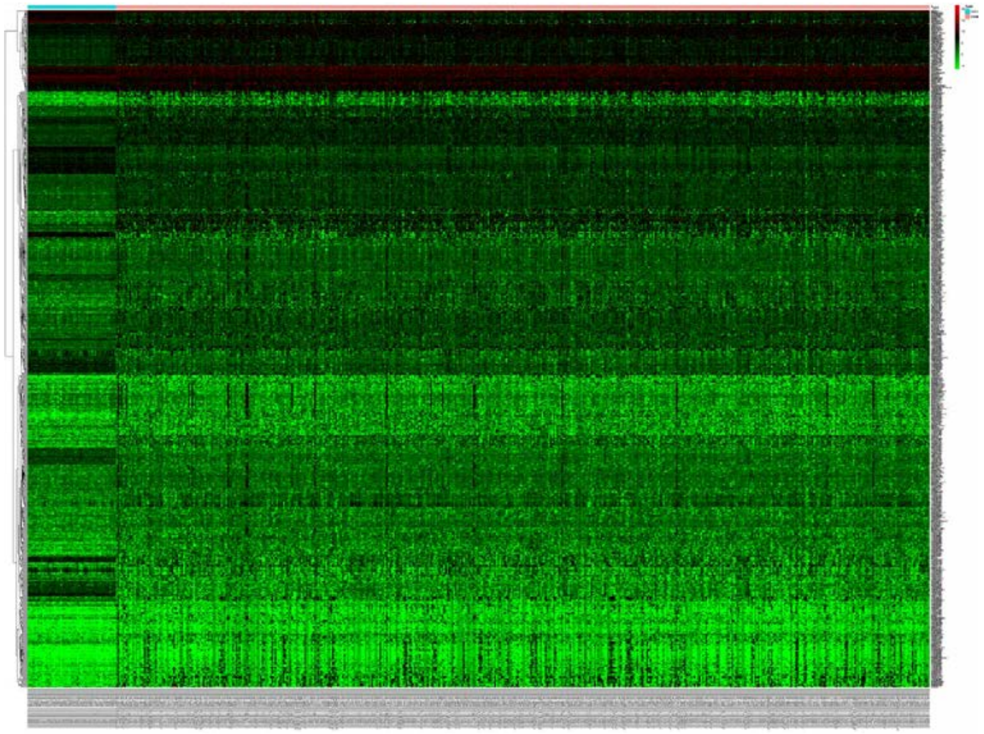

Figure 1. Expression profiles of the lncRNAs with stable expression are presented in the heatmap. Normal samples (54) are blue color coded while the tumor samples (497) are red color coded.

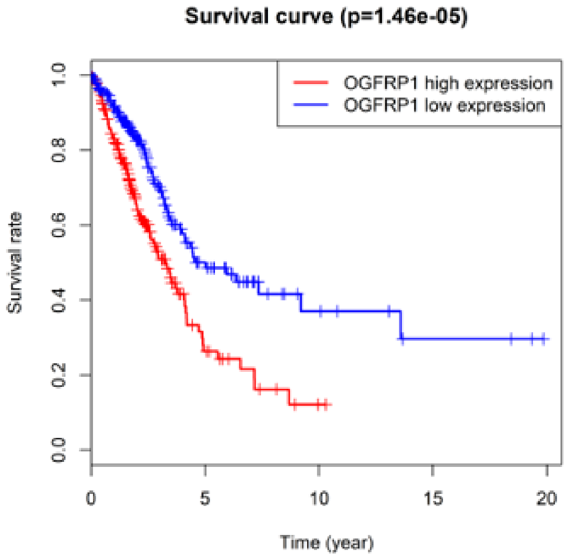

Figure 2. Kaplan-Meier estimates of the survival outcomes for patients. The clinical data download from the TCGA were analyzed for the survival of the different lncRNAs in the LUAD samples as shown in Figure 1.

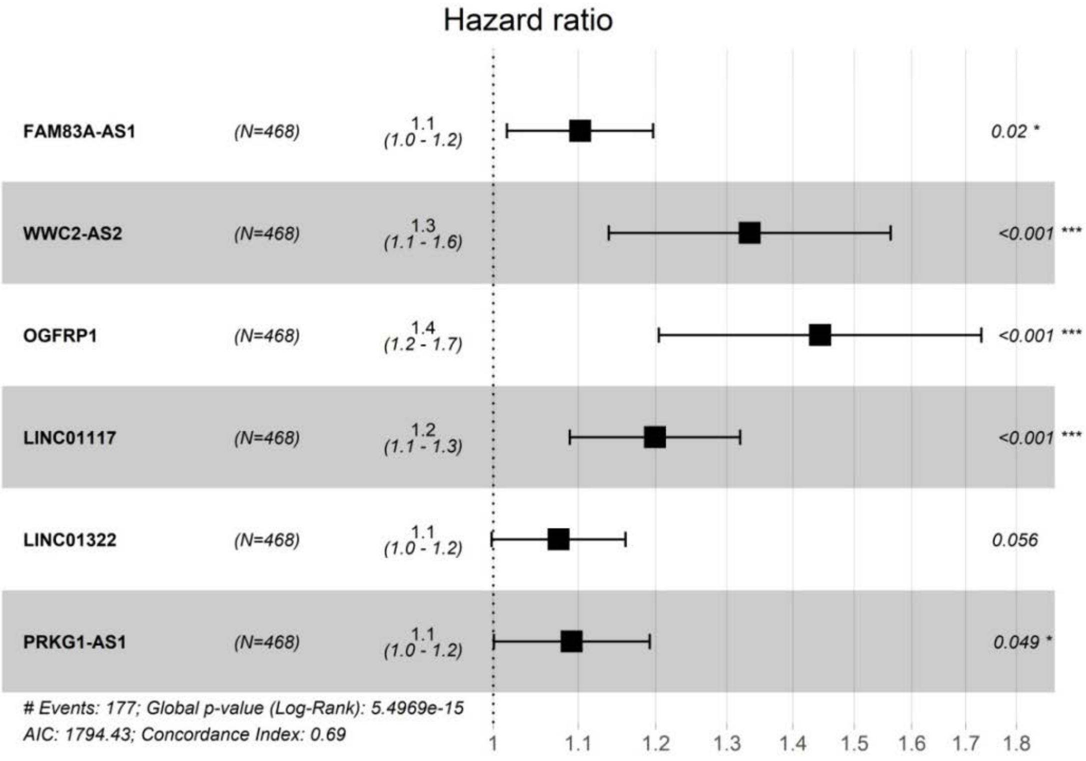

Figure 3. Forest picture of the multivariable Cox regression of the long non-coding RNAs in the TCGA.

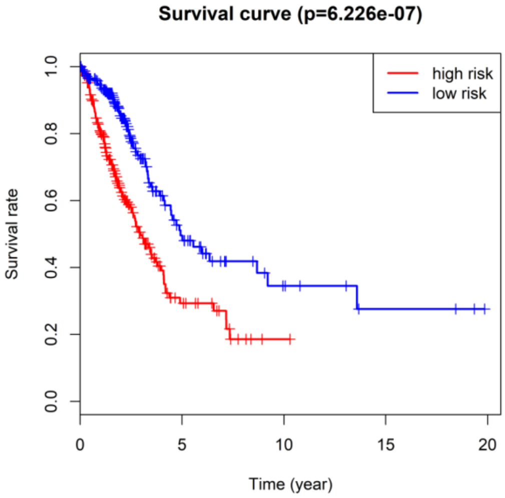

Figure 4. Kaplan-Meier estimates of the survival outcomes of patients using the six lncRNAs signatures and their risk scores.

In order to compare the sensitivity and specificity of the six-lncRNA's signature in the TCGA EFS outcome, a time dependent ROC curve analysis was performed for the six-lncRNA signature and the cytogenetics risk group. The AUROC determined was compared between these two prognostic factors for the five-year EFS in the TCGA data. Figure 5 shows that the AUROC of the six-lncRNA signature was 0.766.

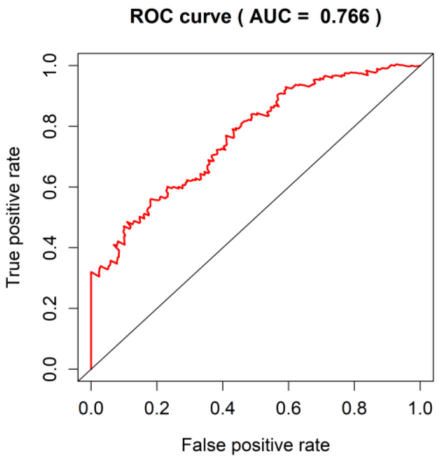

Figure 5. ROC analysis of the risk factors for survival prediction in the TCGA data. The area under the curve is the calculated ROC. Sensitivity and specificity were calculated to assess the score performance.

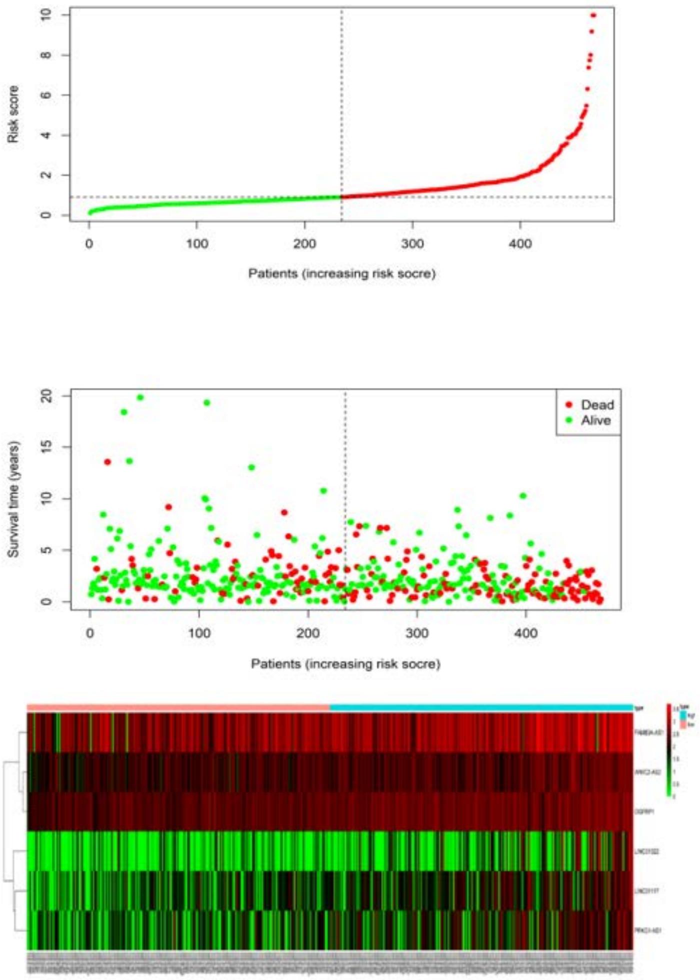

Figure 6. The six-lncRNA based risk score distribution, patient's event-free survival status and a heatmap of the three lncRNA expression profiles.

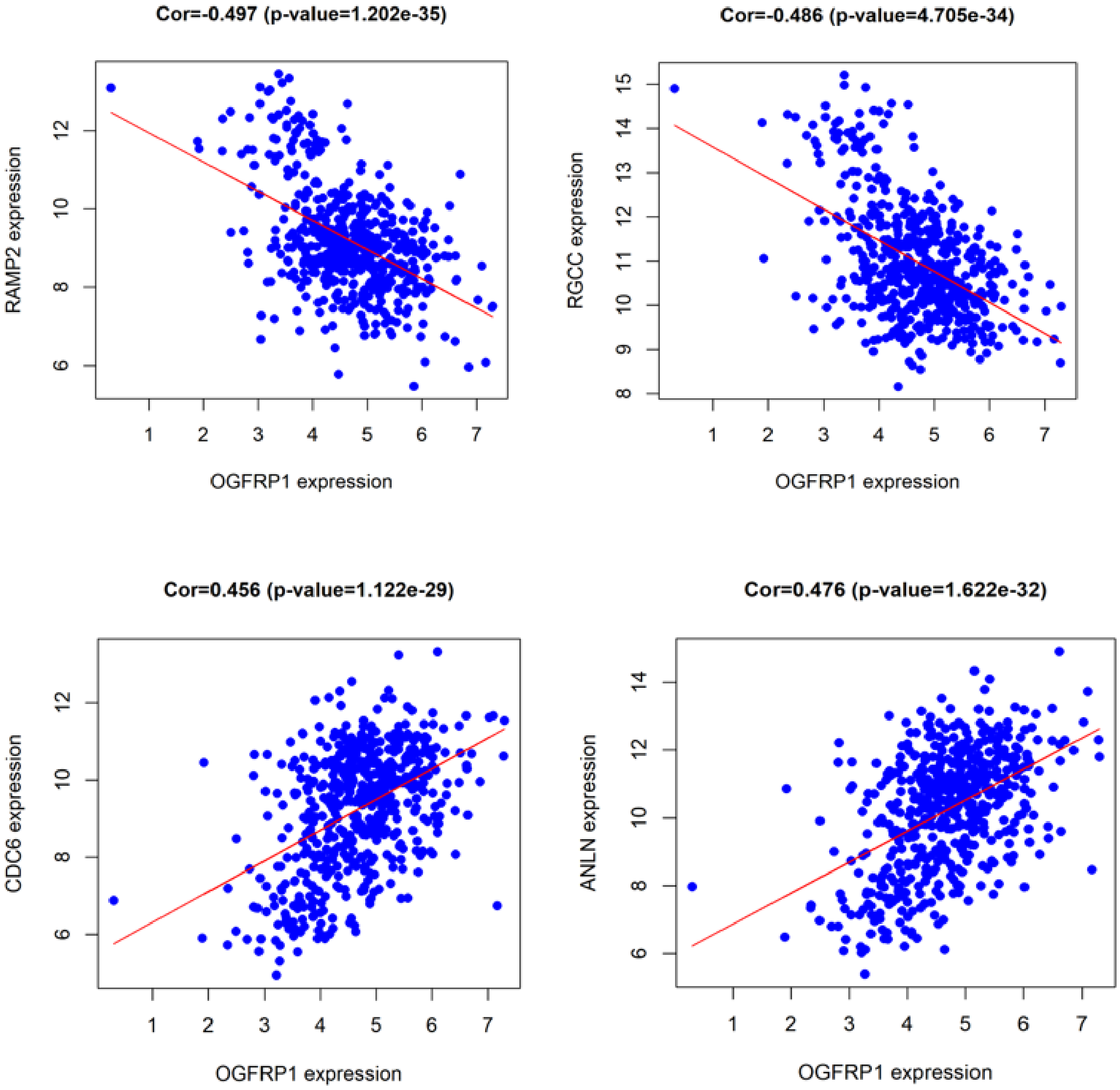

Figure 7. Co-expression analysis of the OGFRP1 screened for genes with the same expression pattern as OGFRP1. The blue line is the expression of the LUAD samples. The red line is the predicted co-expression line.

Using the median of risk score as the cut-off point, patients were categorized as a high-risk (N = 250) or a low-risk group (N = 251). Figure6shows that the low-risk scores patients expressed high levels of the protective lncRNAs (LINC01117, LINC01322, PRKG1-AS1), whereas the high-risk scores patients expressed high levels of the risky lncRNA (FAM83A-AS1, WWC2-AS2, OGFRP1). We observed that the six-lncRNA of the LUAD patients with the high-risk score had a lower survival rate. The patient's survival rate changes significantly with increasing risk score. This confirms these lncRNAs plays a key role in the lung adenocarcinoma pathogenesis and can be used to design better treatment options for these patients.

It is known that lncRNAs function through the cis-, trans- and ceRNA-regulatory mechanisms to modulate the transcription of the protein-coding genes. The genes that co-express with the OGFRP1 are shown in Figure 7.

We analyzed the genes co-expressing with the OGFRP1 and observed that the related pathways and functions are closely related to the expression of OGFRP1. We built a ceRNA regulation network to better predict these related functions. OGFRP1 has the following functions: catalytic, endo-deoxyribonuclease, DNA-dependent ATPase and tubulin binding activity. It is also related to the cell cycle pathway. The GO and the KEGG analysis shown in Figure 8, reveals that OGFRP1 plays a significant role in the LUAD.

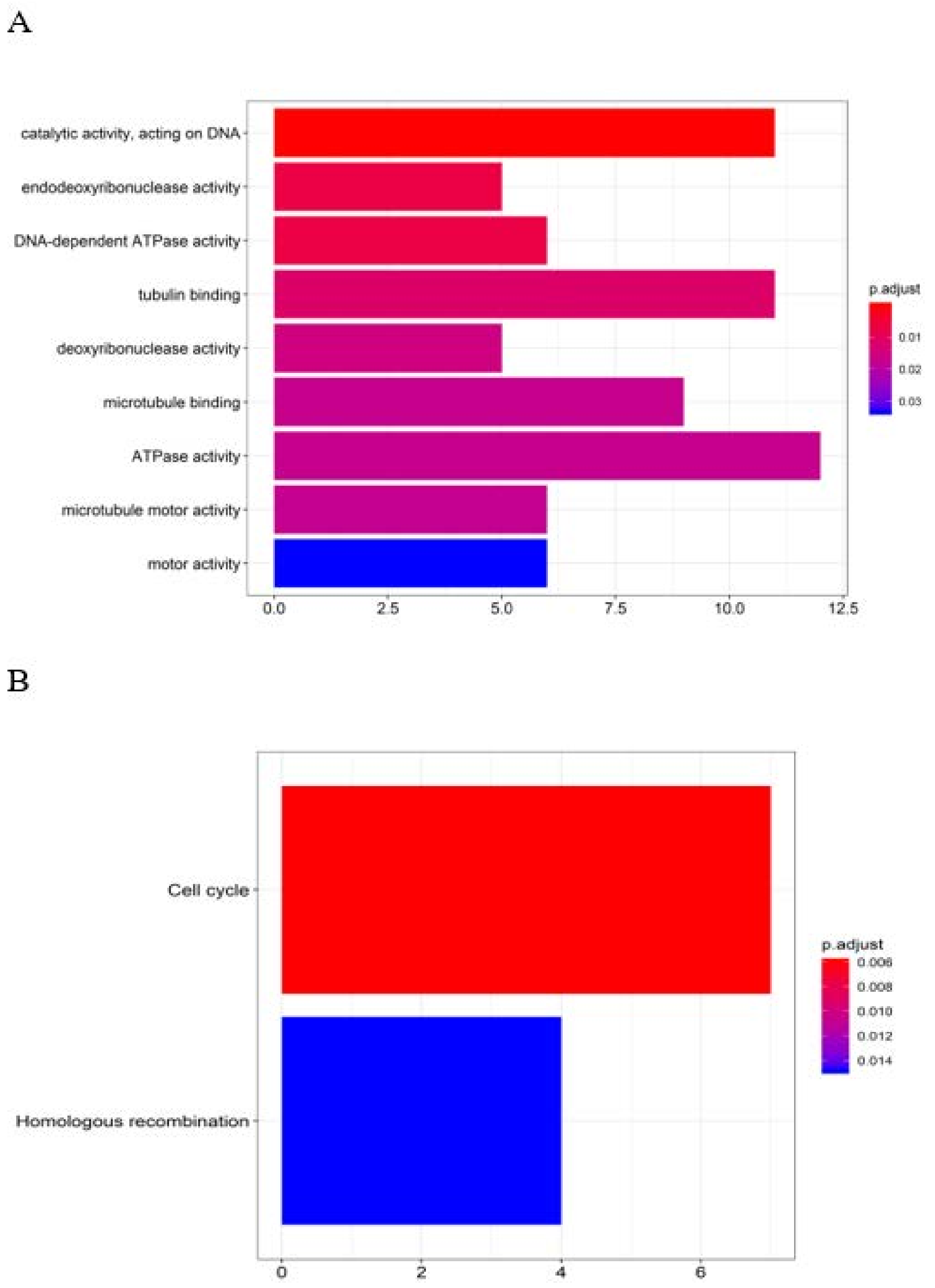

Figure 8. GO and KEGG analysis of the functions of the co-expression genes. (A) GO analysis and (B) KEGG analysis.

Table 1. Univariable Cox regression of significant long non-coding RNAs in TCGA.

gene	HR	z	p-value
LINC02178	1.211869288	5.413879258	6.17×10^{-8}
LINC00941	1.185125067	4.969783536	6.7×10^{-7}
ITGB1-DT	1.19474705	4.879717594	1.06×10^{-6}
LINC01322	1.186034627	4.818536563	1.45×10^{-6}
OGFRP1	1.550208197	4.748826079	2.05×10^{-6}
LINC01117	1.249932998	4.718363135	2.38×10^{-6}
LINC01116	1.188216116	4.295575195	1.74×10^{-5}
LINC01322	1.186034627	4.818536563	1.45×10^{-6}

Table 2. Multivariable Cox regression of long non-coding RNAs in TCGA.

ID	coef	Exp (coef)	Se (coef)	z	Pr (>|z|)
FAM83A-AS1	0.097355021	1.102251627	0.041907498	2.323093152	0.020174149
WWC2-AS2	0.287974929	1.333723866	0.080791315	3.564429269	0.000364649
OGFRP1	0.367100193	1.443542545	0.092385404	3.973573495	7.08×10^{-5}
LINC01117	0.181531227	1.199051979	0.048832249	3.717445579	0.000201247
LINC01322	0.073310735	1.076064856	0.038399289	1.90916905	0.056240288
PRKG1-AS1	0.088017663	1.09200741	0.044699624	1.969091784	0.048942552

4. Discussion

lncRNAs plays a critical role in the of development of the lung adenocarcinoma. Therefore, it is important to understand the underlying molecular mechanisms of the lncRNA to develop better treatment strategies for lung adenocarcinoma. Through transcriptome profiling of The Cancer Genome Atlas, the different lncRNAs associated with the lung adenocarcinoma were discovered, which may be further utilized to explore novel diagnostic and therapeutic strategies. Previous studies have reported some prognostic lncRNAs in the lung adenocarcinoma, such as CADM1-AS1, DLX6-AS1, lncRNASCAL1 and HOTAIR. The CADM1-AS1 and DLX6-AS1 found to be overexpressed in the lung adenocarcinoma tissues, are predicted to function as oncogenes by many studies. The lncRNASCAL1, which shows a specific expression pattern in the lung cancers can be used as a suitable classification for its diagnosis. HOTAIR, which regulates the expression of p21, is the first lncRNA found to be related to the tumors.

In the current study, we identified four significant transcription factor (TF) enrichments of the OGFRP1 co-expressed genes, The RAMP2, RGCC, ERCC6L and ANLN. The function of the RAMP2: Calcitonin and related receptors are a family of the G-protein-coupled receptors that comprises of eight subtypes, viz., CT, AMY1, AMY2, AMY3, CALCR, CGRP, AM1 and AM2. The main function of the CT receptors is to inhibit the bone reabsorption and enhance the calcium excretion by the kidneys. The function of RGCC is to modulate the activity of the cell cycle-specific kinases. It enhances the CDK1 activity, which contributes to the regulation of the cell cycle. It also inhibits the growth of the glioma cells by promoting the arrest of the mitotic progression from the G2/M transition. It is known to be associated with the pathogenesis of renal fibrosis through fibroblast activation. ANLN plays an important function during the process of cytokinesis [14], where it regulates the structural integrity of the cleavage furrow for the completion of the cleavage furrow ingression. Further, it plays a critical role in the bleb assembly process during the metaphase and the anaphase of mitosis [15]. Migration of the podocyte is another role mediated by the ANLD. Thus, it can be concluded that OGFRP1 has a critical role in the LUAD during cell cycle, cell migration and other biological processes.

The study has made significant contributions towards our understanding of the functions of the LUAD associated OGFRP1. However, it has a few limitations. We filtered the low expression lncRNAs using the formula "RPKM > 1" and only included data with count read >200 in >50% patients. As a result, various low expression lncRNAs were excluded which might be potentially correlated with the lung adenocarcinoma patients. Although the TCGA projects we chose were large and comprehensive cancer genomics datasets, the datasets were varied in terms of age, location and gender. Large scale clinical trials are necessary to uncover the functions of the lncRNAs and its underlying mechanism in lung adenocarcinoma to avoid the possibilities of false positives. .

In summary, the study provides preliminary research on the mechanism of lung adenocarcinoma associated with the lncRNAs. Risk score and survival-related relationship of the lncRNAs were screened out by computational bioinformatics to identify the signature performances of the six main lncRNAs. Furthermore, the study provids new information about several key hub genes associated with the key lncRNA which needs to be explored further to develop potential therapeutics against the lung adenocarcinoma.

Conflict of Interest

We have no conflict of interest to declare.

References

1. F. Teng, X. Meng, L. Kong, J. Yu, Progress and challenges of predictive biomarkers of anti PD-1/PD-L1 immunotherapy: A systematic review, *Cancer Lett.*, **414** (2018), 166–173.
2. L. Li, M. Peng, W. Xue, Z. Fan, T. Wang, J. Lian, et al., Integrated analysis of dysregulated long non-coding RNAs/microRNAs/mRNAs in metastasis of lung adenocarcinoma, *J. Transl. Med.*, **16** (2018), 372.
3. Z. Cong, Y. Diao, Y. Xu, X. Li, Z. Jiang, C. Shao, et al., Long non-coding RNA linc00665 promotes lung adenocarcinoma progression and functions as ceRNA to regulate AKR1B10-ERK signaling by sponging miR-98, *Cell Death Dis.*, **10** (2019), 84.
4. J. Sui, S. Yang, T. Liu, W. Wu, S. Xu, L. Yin, et al., Molecular characterization of lung adenocarcinoma: A potential four-long noncoding RNA prognostic signature, *J. Cell. Biochem.*, **120** (2019), 705–714.
5. P. Kumar, S. Khadirnaikar, S. K. Shukla, A novel LncRNA-based prognostic score reveals TP53-dependent subtype of lung adenocarcinoma with poor survival, *J. Cell. Physiol.*, **234** (2019), 16021–16031.
6. D. Li, W. Yang, C. Arthur, J. S. Liu, C. Cruz-Niera, M. Q. Yang, Systems biology analysis reveals new insights into invasive lung cancer, *BMC Syst. Biol.*, **12** (2018), 117.

7. X. Zhao, X. Li, L. Zhou, J. Ni, W. Yan, R. Ma, et al., LncRNA HOXA11-AS drives cisplatin resistance of human LUAD cells via modulating miR-454-3p/Stat3, *Cancer Sci.*, **109** (2018), 3068–3079.
8. M. Qiu, Y. Xu, J. Wang, E. Zhang, M. Sun, Y. Zheng, et al., A novel lncRNA, LUADT1, promotes lung adenocarcinoma proliferation via the epigenetic suppression of p27, *Cell Death Dis.*, **6** (2015), e1858.
9. X. Shi, H. Tan, X. Le, H. Xian, X. Li, K. Huang, et al., An expression signature model to predict lung adenocarcinoma-specific survival, *Cancer Manag. Res.*, **10** (2018), 3717–3732.
10. D. Li, W. Yang, J. Zhang, J. Y. Yang, R. Guan, M. Q. Yang, Transcription factor and lncRNA regulatory networks identify key elements in lung adenocarcinoma, *Genes*, **9** (2018), 12.
11. S. Shukla, Unravelling the Long Non-Coding RNA Profile of Undifferentiated Large Cell Lung Carcinoma, *Non-Coding RNA*, **4** (2018), 4.
12. X. Li, B. Li, P. Ran, L. Wang, Identification of ceRNA network based on a RNA-seq shows prognostic lncRNA biomarkers in human lung adenocarcinoma, *Oncol. Lett.*, **16** (2018), 5697–5708.
13. M. Qiu, D. Feng, H. Zhang, W. Xia, Y. Xu, J. Wang, et al., Comprehensive analysis of lncRNA expression profiles and identification of functional lncRNAs in lung adenocarcinoma, *Oncotarget*, **7** (2016), 16012–10622.
14. W. M. Zhao, G. Fang, Anillin Is a Substrate of Anaphase-promoting Complex/Cyclosome (APC/C) That Controls Spatial Contractility of Myosin during Late Cytokinesis, *J. Biol. Chem.*, **280** (2005), 33516–33524.
15. T. Kiyomitsu, I. M. Cheeseman, Cortical Dynein and Asymmetric Membrane Elongation Coordinately Position the Spindle in Anaphase, *Cell*, **154** (2013), 391–402.