ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0020

Screening Prognostic Biomarker of LUAD Based on Immune-related Lncrna

Qiuyang Chen[1, a, #], Wei Kong[1, b, *, #], Shuaiqun Wang[1, c]

[1]College of Information Engineering, Shanghai Maritime University, Shanghai, 201306, P.R. China

[a]201930310145@stu.shmtu.edu.cn, [b]weikong@shmtu.edu.cn, [c]wangsq@shmtu.edu.cn

[#]The first two authors should be regarded as Joint First Authors.

Abstract

Lung adenocarcinoma (LUAD) is a common disease, and its five-year overall survival rate is not high. In order to make a helpful role in the early screening methods and immunotherapy of LUAD patients, this study is based on the tumor microenvironment (TME) to find specific biomarkers related to immune infiltrating cells in LUAD patients. At present, there are few studies on lncRNA, and the role of a large number of lncRNA remains to be studied and explored. In this study, we searched for lncRNAs related to immune infiltrating cells based on the ImmLnc algorithm. Survival analysis of the obtained genes revealed that FENDRR, HSPC324 and LHFPL3-AS2 were significantly related to the prognosis of LUAD patients. And the high expression of FENDRR is related to poor prognosis. Low expression of HSPC324 and LHFPL3-AS2 were associated with unfavourable overall survival rate. So, we infer that FENDRR, HSPC324 and LHFPL3-AS2 may be potential prognosis biomarkers of LUAD.

Keywords

LUAD; Bioinformatics; Prognosis; Immune-related, lncRNA.

1. INTRODUCTION

As we all know, lung adenocarcinoma is a very common malignant tumor, and its incidence and mortality are very high [1]. With the development of immunology and genomics technologies, the search for effective and reliable diagnostic and prognostic biomarkers has increasingly become the focus. Biomarkers can be used for early diagnosis, prevention and treatment of lung adenocarcinoma. Therefore, it is necessary for us to find effective potential biomarkers and further understand the pathogenic mechanism of LUAD.

lncRNA is a type of RNA that does not encode protein but has biological functions. Studies have shown that lncRNA is involved in carcinogenic and tumor suppressor pathways, including proliferation, adhesion, migration and apoptosis [2], so lncRNA molecules may become effective biomarkers for cancer immunotherapy. Recently, it has been discovered that lncRNA participates in the development and differentiation of immune cells, and also participates in the expression of genes related to immune cell activation, leading to tumor immune cell infiltration [3]. Although lncRNA is involved in many biological functions of cancer, there are still few prognostic biomarkers of lung adenocarcinoma found [4]. This article will screen lncRNAs associated with tumor-infiltrating immune cells to further explore the pathogenesis of lung adenocarcinoma Biomarkers related to mechanism and prognosis provide new ideas for immunotherapy of LUAD based on the immune-related lncRNA(irlncRNA) models.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0020

In this study, in order to analyze the relationship between lncRNA and immunity in LUAD patients, we first combined edgeR with immLnc algorithm to screen lncRNAs related to immunity. Next analyze the relationship between 17 irlncRNAs and tumor-infiltrating immune cells, and screen irlncRNAs that closely related to B cells, among which B cells are tumor-infiltrating immune cells that are significantly related to the prognosis of LUAD patients. Then, the Cox regression model analysis and survival analysis were performed to screen out the differentially expressed irlncRNAs related to the prognosis. Through the above methods, three lncRNAs were found in this study, namely FENDRR, HSPC324 and LHFPL3-AS2. According to the analysis, the higher the expression level of HSPC324 and LHFPL3-AS2, the better the prognosis of LUAD patients. The expression level of FENDRR is inversely proportional to the overall survival rate of LUAD patients. We speculate that FENDRR, HSPC324 and LHFPL3-AS2 are potential prognostic biomarkers for patients with lung adenocarcinoma.

2. MATERIALS AND METHODS

2.1. Data Collection and Preprocessing

The LUAD transcriptome data and clinical data were downloaded from the TCGA database (https://cancergenome.nih.gov/). Among them, the transcriptome data includes 594 samples, of which 535 are tumor samples and 59 are normal samples. The clinical data includes demographic data (age, gender, and race), smoking history, overall survival (OS) cancer status, and pathological stage. We deleted most of the data with a value of 0 and used RESM for quantification. Then for the transcriptome data, we isolated a total of 14,370 lncRNAs expression data for subsequent research.

2.2. Differential Expression Analysis

We extracted the differentially expressed lncRNA through the "edgeR" R package. An absolute log2 fold change > 1, and adjusted P-value < 0.05 were used as the threshold for screening differentially expressed genes, and false discovery rate (FDR) will be used to control the false positive rate.

2.3. Method for Selecting Immune-related DElncRNAs

In order to identify lncRNA regulators related to immune response, we used the ImmLnc integrated algorithm for analysis. ImmLnc (http://bio-bigdata.hrbmu.edu.cn/ImmLnc/) [5] is a web-server for studying the immune-related functions and pathways of lncRNA in various cancer types. It also provides a correlation between lncRNA expression and immune cell infiltration in cancer based on TIMER. TIMER uses a deconvolution method based on expression characteristics of cell mixtures, which can quantitatively estimate the relative fraction of cell types, and it has been well verified by flow cytometry. TIMER algorithm eliminates the bias effect by screening immune signature genes and removing highly expressed genes, and can eliminate the collinearity between immune cells to ensure the accuracy of inference. Then, we will introduce the principle of ImmLnc which is used to identify immune-related lncRNAs in different cancers.

In ImmLnc, a computational method that integrates lncRNA and gene expression data was proposed. First, calculate the partial correlation coefficient (PCC) between the expression of lncRNA i and gene j, as shown in Equation 1 below.

$$PCC(i,j) = \frac{R_{LG} - R_{LP} * R_{GP}}{\sqrt{1 - {R_{Lp}}^2} * \sqrt{1 - {R_{Gp}}^2}} \quad (1)$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0020

Among them, R_{LG} is the expression of lncRNA i and coding gene j. And R_{LP} represents the expression of lncRNA i and tumor purity. R_{GP} means the correlation coefficient between the expression of gene j and tumor purity. For each lncRNA gene pair, we calculated the rank score (RS) as the following formula 2. All genes are ranked based on the RS score, and then enrichment analysis is performed.

$$RS(i,j) = -\log 10\big(P(ij)\big) * sign(PCC(ij)) \quad (2)$$

Where p(ij) is the P-Value of PCC.

2.4. Relationship between Immune Infiltrating Cells and Prognosis

Since the abnormally expressed lncRNAs in the tumor microenvironment can be used as a potential biomarker for prognosis and immunotherapy, we first identified immune cells related to the prognosis. Then analyze the relationship between lncRNAs and tumor infiltrating immune cells, and screen the lncRNAs related to tumor infiltrating immune cells to prepare for the subsequent survival analysis. A multivariate Cox regression model was used to further explore the relationship between immune infiltration cells and the prognosis of LUAD, which uses age, gender, tumor stage and other factors as variables. For all statistical analysis, P-value < 0.05 is considered significant. The above analysis was achieved through TIMER2.0 [6] (http://timer.cistrome.org/).

2.5. Survival Analysis

In order to investigate whether the lncRNA associated with tumor-infiltrating immune cells can be used as an independent prognostic factor, we performed univariate and multivariate Cox regression analysis for each variable. Multivariate analysis with the Cox regression model was performed using the clinical data (survival time and status) of patients with LUAD. According to the median expression value of identified overlapping DElncRNAs in tumor group, patients were divided into two groups: high expression group and low expression group. Kaplan-Meier curves were drawn and log-rank test was performed to analyze the relationship between genes and prognosis. When P-value is less than 0.05, it is considered statistically significant.

3. RESULTS

3.1. Identification Differentially Expressed lncRNA in LUAD

The result of edge R differential analysis showed that 366 differentially expressed DElncRNAs were found, of which 56 were up-regulated and 310 were down-regulated. Next, the up-regulated DElncRNAs were selected to obtain the immune-related lncRNA using the immLnc database. The figure 1 below showed the volcano plot of the differentially expressed genes of DElncRNA. The red part on the right side indicated the high expression gene, and the green part indicated the low expression gene.

3.2. Obtaining Immune-Related Differentially Expressed lncRNA

We used the ImmunLnc algorithm to screen out lncRNAs related to immunity. The Pearson correlation analysis was applied to the immune-related pathways of lung adenocarcinoma, and we compared the up-regulated differential lncRNAs obtained above, at last, 17 immune-related DElncRNAs were obtained.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0020

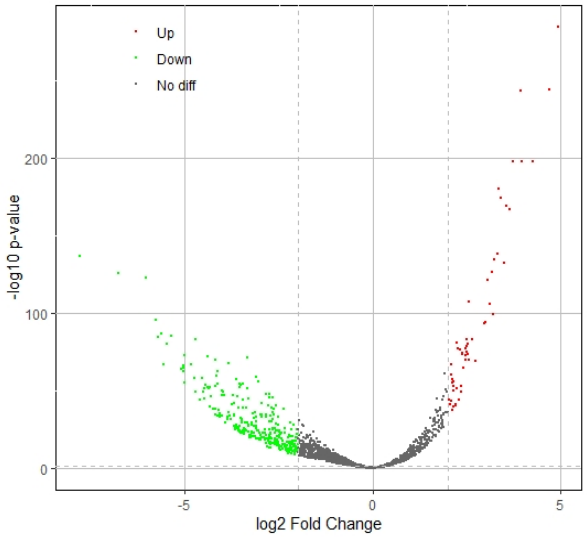

Figure 1. Volcano maps of differential expression of lncRNAs

3.3. Correlation Analysis Between Immune Cells and Prognosis of LUAD Patients

lncRNA plays a vital role in cancer immunity, it can infiltrate into cancer tissues and kill cancer cells. The abnormally expressed lncRNA in the immune microenvironment may provide new potential targets for cancer immunotherapy. Therefore, we first used the TIMER algorithm to identify immune cells related to the prognosis of lung adenocarcinoma patients. And based on this, we searched for lncRNAs related to tumor-infiltrating immune cells to find the prognostic lncRNA biomarkers. Regarding the exploration of the impact of immune infiltration on clinical outcomes, a multivariate Cox regression with infiltration abundance, age, gender, tumor purity, stage and viral infection status was constructed using the TIMER server. In table 1, the covariates included 6 kinds of cells that could be analyzed by TIMER, including B cells, $CD4^+T$ cells, $CD8^+T$ cells, Neutrophil, Macrophage and Dendritic cell. Obviously, B cells infiltration were also a key prognostic factor besides stage according to index Significant. Based on the above results, the following study will start with B cells infiltration.

Table 1. Cox risk proportional regression model

Variable	Coef	HR	95%CI_l	95%CI_u	P value	Significant
Age	0.01	1.01	0.991	1.029	0.298	
gendermale	-0.165	0.848	0.592	1.215	0.369	
raceBlack	16.271	11646434	0	Inf	0.994	
raceWhite	16.442	13827340	0	Inf	0.994	
stage2	0.833	2.299	1.488	3.553	<0.0001	***
stage3	0.932	2.541	1.62	3.984	<0.0001	***
stage4	1.112	3.041	1.546	5.982	0.001	**
Purity	0.223	1.25	0.513	3.049	0.624	
B cell	-3.203	0.041	0.002	0.68	0.026	*
CD8+Tcell	-0.593	0.553	0.065	4.72	0.588	
CD4+Tcell	1.367	3.923	0.226	68.088	0.348	
Macrophage	-0.25	0.779	0.035	17.103	0.874	
Neutrophil	-0.686	0.503	0.008	30.046	0.742	
Dendritic	0.097	1.101	0.259	4.682	0.896	

In Table 1, the coef represents the regression coefficient, and HR represents the hazard ratio. It can be seen from the significant column that B cells are also prognostic factors in addition to tumor stage.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0020

3.4. Obtaining IrlncRNAs Related to B cell

In order to further reveal the relationship between lncRNA and tumor-infiltrating immune cells, we downloaded the correlation analysis data of 6 types of immune cells and lncRNA expression data, and obtained DElncRNAs related to immune cells by controlling the threshold value of p-value<0.05 and cor>0.1. After screening, a total of 14 irlncRNAs related to B cells were obtained, and the results are shown in Table 2. Next, survival analysis will be used to determine the relationship between these lncRNAs and the prognosis of LUAD patients.

Table 2. Immune-related lncRNAs associated with B cells

lncRNA symbol	Immune cell	cor_p_value	cor_R_value
SMIM25	B cell	2.14361E-05	0.186962854
LHFPL3-AS2	B cell	1.46473E-10	0.278832101
HSPC324	B cell	5.66462E-05	0.177310798
SRGAP3-AS2	B cell	5.18778E-10	0.270643284
LINC02154	B cell	0.000347844	0.157776832
AC008268.1	B cell	3.4515E-05	0.182294004
LINC01936	B cell	2.54116E-10	0.275297988
LINC02016	B cell	0.004423496	0.125842259
LINC00968	B cell	9.29273E-09	0.250849467
MIR3945HG	B cell	0.000791623	0.148136098
LINC01863	B cell	0.001077843	0.14436565
LINC01996	B cell	1.48259E-06	0.211256492
PCAT19	B cell	0.006987908	0.119309607
FENDRR	B cell	2.51479E-05	0.185410331

This table shows the correlation results between B cells and differentially expressed irlncRNAs calculated by Pearson correlation coefficient.

4. DISCUSSION

To investigate whether 14 lncRNAs obtained above (SMIM25, LHFPL3-AS2, HSPC324, SRGAP3-AS2, LINC02154, AC008268.1, LINC01936, LINC02016, LINC00968, MIR3945HG, LINC01863, LINC01996, PCAT19 and FENDRR) were related to the prognosis of LUAD patients, we used the multivariate Cox Proportional Hazards Model for survival analysis with genes expression values of 14 lncRNAs. In Figure 2, the survival curve with high-risk and low-risk groups showed significant differences in survival time. The Receiver Operating Characteristic (ROC) curve analysis indicated that lncRNAs we found could serve as potential biomarkers for LUAD. Then, we performed a Kaplan-Meier survival curve and a log-rank test with 14 lncRNAs. The survival curves (shown in Figure 3) of LUAD patients in the effect of each lncRNA gene expression level have been plotted by “ggplot2” R package. The red line represents the survival probability of lncRNAs with high expression in lung adenocarcinoma, while the blue line represents the survival probability of lncRNAs with low expression. The x-axis means survival time and the y-axis shows survival rate. In Figure 3, it could be found that the expression level of three lncRNAs (FENDRR, HSPC324 and LHFPL3-AS2) were significantly relevant to the survival probability of LUAD patients because every P-value was less than 0.05. The remaining unqualified lncRNAs could not be used as prognostic biomarkers, so they were not considered for analysis. After analysis, LUAD patients with high expression levels of lncRNA HSPC324 and lncRNA SRGAP3-AS2 had significant better overall survival. However, the analysis results of

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0020

lncRNA FENDRR showed completely different conclusions. According to the K-M curve, it can be concluded that the higher the level of lncRNA FENDRR infiltration, the worse the overall survival rate of patients with lung adenocarcinoma.

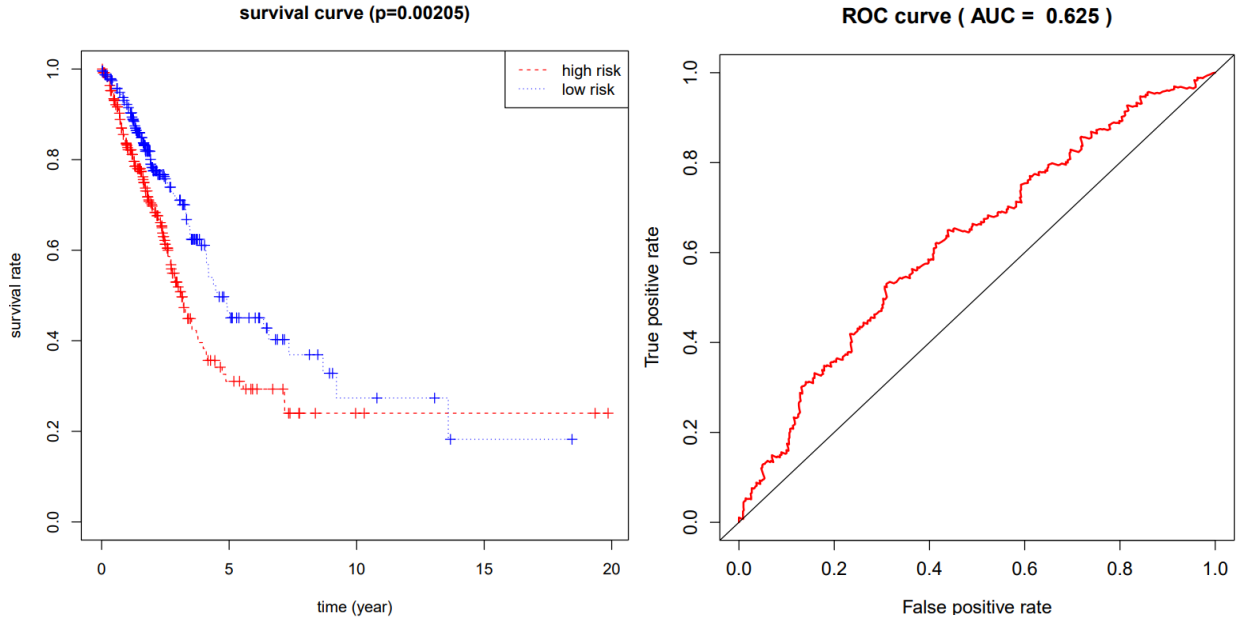

Figure 2. Multivariate Cox regression model constructed based on lncRNA expression data and clinical data

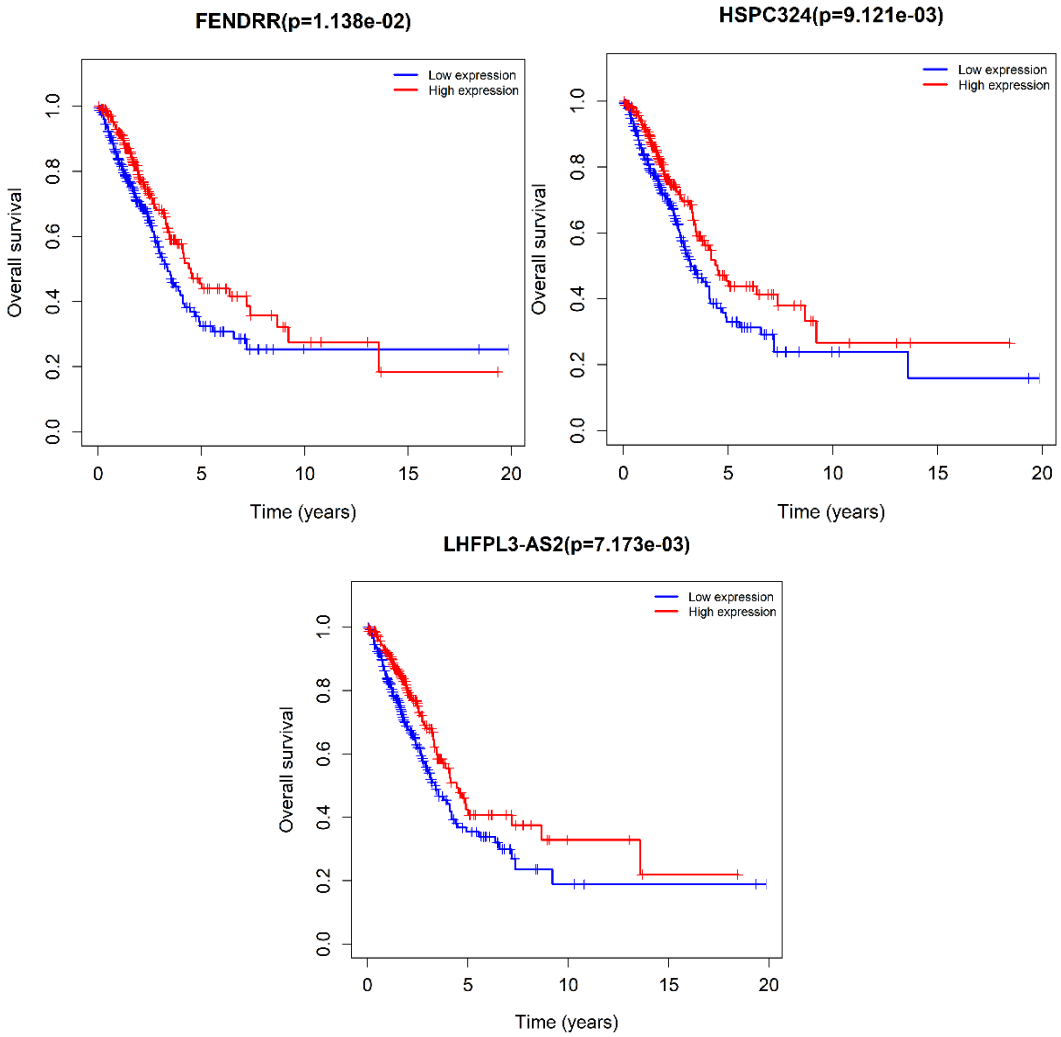

Figure 3. K-M survival curve of lncRNA screened based on P-Value<0.05

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0020

In order to verify whether the tumor-infiltrating immune cells related lncRNAs obtained in this study are really related to the prognosis of patients with lung adenocarcinoma, and whether they can be used as a potential biomarker, we consulted the literature for confirmation. The literature [7] shows that over-expressed lncRNA FENDRR reduces cell proliferation and tumorigenicity in hepatocellular carcinoma (HCC). Liu [8] pointed out that compared with normal controls, the attenuation of FENDRR is very common in colon cancer tissues. Low levels of FENDRR are related to clinical staging and poor prognosis. The role of expression of lncRNA FENDRR in different cancers is completely different. In this study, the high expression of FENDRR may play a role in promoting the development of cancer. In the future, more data and research are needed to determine the role of FENDRR in lung adenocarcinoma. According to the literature, we found that Jafarzadeh et al. [9] discovered that overexpression of lncRNA HSPC324 reduces cell proliferation, meaning that it plays a role in inhibiting the spread of lung cancer based on different databases. This conclusion was consistent with the results of survival analysis obtained in the previous text. Therefore, we can infer that the higher the expression of lncRNA HSPC324, the better the prognosis of LUAD patients. At present, there are very few studies on lncRNA LHFPL3-AS2, and we can only find Zhang's study [10], which shows that the expression of lncRNA LHFPL3-AS2 in LUSC is significantly lower than in normal lung tissue. This coincides with our research, so it can basically be determined that the high expression of lncRNA LHFPL3-AS2 plays a helpful role in the overall survival rate of cancer patients.

However, the current research on lncRNA is very rare, and the available literature is very limited. It is impossible to fully determine whether the 3 tumor-infiltrating immune cells related lncRNAs we found can be used as biomarkers for patients with lung adenocarcinoma. Therefore, the research of lncRNA still has a long way to go, and more data is needed for analysis.

5. CONCLUSION

As the prevalence and mortality of lung adenocarcinoma are still high, we need to screen suitable biomarkers to help early screening and treatment. At present, the immune microenvironment is a hot spot for tumor immunotherapy, and the function and mechanism of a large number of lncRNAs on cancer are still unclear. Therefore, this study mainly started with immune-related lncRNAs, and screened lncRNAs related to immune infiltrating cells based on the ImmLnc algorithm. Through survival analysis, 3 lncRNAs that are highly correlated with the prognosis of LUAD patients were screened out, namely FENDRR, HSPC324 and LHFPL3-AS2. We found that the high expression of FENDRR is unfavourable to the prognosis of LUAD patients. And, the higher the expression level of HSPC324 and LHFPL3-AS2, the higher the overall survival rate of LUAD patients. However, due to the current limited research and data, a large number of lncRNAs cannot be explored in depth. We look forward to verifying the conclusions of this study again in the future.

6. COMPETING INTERESTS

The authors declare no conflict of interests.

ACKNOWLEDGMENTS

This work was supported by the Natural Science Foundation of Shanghai (No. 18ZR1417200) and National Natural Science Foundation of China (No. 61803257).

REFERENCES

[1] Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015;65(2):87–108.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202106_7(6).0020

[2] Siegel RL, Miller KD, Jemal A (2017) Cancer statistics, 2017. CA Cancer J Clin 67(1):7–30.

[3] Chen Y.G., Satpathy A.T., Chang H.Y. Gene regulation in the immune system by long noncoding RNAs. Nat. Immunol. 2017; 18:962–972.

[4] Jin D, Song Y, Chen Y, Zhang P. Identification of a Seven-lncRNA Immune Risk Signature and Construction of a Predictive Nomogram for Lung Adenocarcinoma. Biomed Res Int. 2020 May 21; 2020:7929132.

[5] Li Y, Jiang T, Zhou W, et al. Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers. Nat Commun. 2020 Feb 21;11(1):1000.

[6] Li T, Fu J, Zeng Z, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. Nucleic Acids Res. 2020 Jul 2;48(W1): W509-W514.

[7] Yu Z, Zhao H, Feng X, et al. Long Non-coding RNA FENDRR Acts as a miR-423-5p Sponge to Suppress the Treg-Mediated Immune Escape of Hepatocellular Carcinoma Cells. Mol Ther Nucleic Acids. 2019 Sep 6;17: 516-529.

[8] Liu J, Du W. LncRNA FENDRR attenuates colon cancer progression by repression of SOX4 protein. Onco Targets Ther. 2019 May 30;12: 4287-4295.

[9] Jafarzadeh M, Tavallaie M, Soltani BM, et al. LncRNA HSPC324 plays role in lung development and tumorigenesis. Genomics. 2020 May;112(3):2615-2622.

[10] Zhang Y. The clinical significance and function of long non-coding RNA PGM5-AS1, HLA-F-AS1 and LHFPL3-AS2 in lung squamous cell carcinoma[D]. Guangxi Medical University, 2019.