Six-long Non-coding RNA Signature to Predict the Prognosis of Lung Adenocarcinoma

Yang Wang
Xiangya Hospital Central South University
Chengping Hu (huchengp28@csu.edu.cn)
Department of Respiratory Medicine, Xiangya Hospital (Key Cite of National Clinical Research Center for Respiratory Disease), Central South University, Changsha, Hunan, 410008, PR China.
https://orcid.org/0000-0002-1285-8579

Research article

Keywords: lncRNA, LUAD, prognosis, biomarkers

Posted Date: August 24th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-56767/v1

Abstract

Background: Long non-coding RNAs (lncRNAs) have been reported to play essential roles in tumorigenesis and cancers prognosis, and they can be a potential cancer prognostic markers. However, in lung adenocarcinoma(LUAD), how lncRNA signatures predict the survival of patients is poorly understood. Our study aims to explore lncRNA signatures and prognostic function in LUAD.

Methods: The expression and prognosis data of lncRNAs in LUAD patients was collected from the Cancer Genome Atlas (TCGA) data. All analyses were performed using the R package (version 3.6.2). Metascape, STRING and Cytoscape were used for enrichment analysis and function prediction of the lncRNA co-expressed protein-coding genes.

Results: We have collected lncRNA expression data in 466 LUAD tumors, and a six-lncRNA signature(RP11-79H23.3, RP11-309M7.1, CTD-2357A8.3, RP11-108P20.4, U47924.29, LHFPL3-AS2) has been shown to be significantly related to LUAD patients' overall survival. According to the lncRNA signatures, the high-risk and low-risk groups were divided in LUAD patients with different survival rates. Further multivariable cox regression analysis showed that the prognostic value of this signature was independent of clinical factors. The potential functional roles and hub co-expressed protein-coding genes in the six prognostic lncRNAs are shown in the functional enrichment analysis.

Conclusions: These results showed that these six lncRNAs could be independent predicted prognostic biomarkers in LUAD patients.

Introduction

The incidence and mortality rate of lung cancer worldwide is still in the forefront, and the 5-year survival rate is less than 20%[1]. Lung cancer is classified as lung squamous cell carcinoma, small cell lung cancer and lung adenocarcinoma (LUAD) which is the most common type of lung cancer. The standard treatments for LUAD are chemotherapy, target therapy and surgery, but many patients still got into deterioration. Nowadays, no significant molecular biomarkers for LUAD have been accepted. It is necessary to prevent the development of early LUAD actively. Long non-coding RNAs (lncRNAs) are RNA transcripts longer than 200 bp with little or no protein-coding capacity[2–3]. More and more studies showed that lncRNAs were involved in chromosome regulation, transcription and post-transcriptional regulation[4–5]. It has also been observed to be involved in the development and progression of other cancers, such as gastric cancer and liver cancer[6–7]. Several prognostic biomarkers for LUAD have been reported such as lncRNA LOC100132354[8]and lncRNA CTB-193M12.5[9], etc. In this study, we used a cohort of 466 LUAD patients from The Cancer Genome Atlas (TCGA) to predict whether potential lncRNA was related to the survival of LUAD patients. We demonstrated that six-lncRNA signature was independent of clinical factors and suggested the potential roles in LUAD pathogenesis.

Materials And Methods

Ethics Statement

The current study received approval from the Xiangya Hospital of Central South University according to the Declaration of Helsinki. Every data was collected from the open web, confirming that all the written informed consents were attained.

The LUAD patient information

The lncRNAs expression and clinical data of LUAD patients were collected from the TCGA database. A total of 466 LUAD patients were enrolled in this study. We also downloaded the related clinical data of LUAD patients, including age, gender, tumor stage, smoked pack-year, kras state. We divided samples from TCGA data set into training (n = 233) and testing sets (n = 233).

lncRNA expression profile

LUAD RNA-seq data were collected from TCGA data portal (https://tcga-data.nci.nih.gov/tcga/). According to the human genome (Ensembl database v72 assembly), the expression level of ln-cRNAs and mRNAs were counted by the value of Reads Per Kilobase of exon model per Million mapped reads (RPKM). We evaluated lncRNAs from TCGA dataset according to the foll-owing criteria: 1) transcripts do not locate in any protein-coding region; 2) transcript sequences connect with GENCODE project[10]; 3) transcripts express in more than half of the LUAD samples. The lncRNA expression profiles were defined as those with an average RPKM $\geq$ 0.1 across 466 LUAD samples. At last, a total of 6102 lncRNAs in the dataset were enrolled. We used edgeR[11] software to identify the expression difference.

Statistical analysis

Based on the training set, we used univariate cox regression to calculate the association between the expression level of every lncRNA and overall survival in patients. Those lncRNAs were significant if P-values were less than 0.05. Then, we used multivariate analysis by the Random Survival Forests method to calculate the selected lncRNAs. Risk scores were considered to be involving in these selected lncRNAs, which were weighted by their estimated regression coefficients in the multivariable cox regression model. The risk score can be estimated for each LUAD patient according to prognostic six lncRNAs. We divided LUAD patients into high-risk and low-risk groups according to the risk score. The Kaplan-Meier survival analyses can estimate the survival differences in those two groups. We used multivariate Cox regression and stratified analyses to demonstrate whether the lncRNA was independent of other clinical factors. The receiver operating characteristic (ROC) curve of 5 years were used to estimate the sensitivity and specificity of the survival prediction based on the risk score. All analyses were performed using the R package (version 3.6.2).

Functional Enrichment Analysis

Metascape(http://metascape.org)[12–19] is a website for gene annotation and analysis. In this study, metascape was used to identify pathway and process enrichment of lncRNA correlated genes. On the Metascape online tool, enrichment is shown by the Gene Ontology (GO) terms for biological process,

cellular component, and molecular function categories, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The terms with the characteristics of P-value < 0.01,minimum count of 3 and enrichment factor of > 1.5 were considered significant. The network was plotted with the similarity of > 0.3 to connect edges. Search Tool for the Retrieval of Interacting Genes (STRING; http://string-db.org) (version 10.0)[20] constructed the PPI network of six lncRNA co-expressed protein-coding genes. Cytoscape (version 3.4.0) is an open web for visualizing networks of molecular interaction[21]. We used The plug-in Molecular Complex Detection (MCODE) (version 1.4.2) of Cytoscape to cluster PPI network so that we can get the most significant interaction network.

Results

Identifying the prognostic lncRNAs from the training set

The 466 TCGA LUAD patients were randomly divided into a training (n = 233) set or a testing set (n = 233), respectively. Based on the training set, the lncRNAs were subjected to the Cox regression model, and a total of six lncRNAs were significantly correlated with the patients' overall survival (P-value < 0.05; Table 1). Three of them (RP11-79H23.3, RP11-309M7.1, CTD-2357A8.3) had positive coefficients, suggesting that the higher expression level was related to shorter survival. The negative coefficients for the other three lncRNAs (RP11-108P20.4, U47924.29, LHFPL3-AS2) with higher levels of expression were related to longer survival.

Table 1
six lncRNAs significantly associated with the overall survival

Gene ID	Gene symbol	Chromosome	Coeffcient	HR	95%CI	P-value
ENSG00000261618.1	RP11-79H23.3	chr8:78837529–78840522	0.37	1.45	1.18–1.78	0.00037
ENSG00000261298.1	RP11-309M7.1	chr2:234223314–234224514	0.21	1.23	1.10–1.39	0.0005
ENSG00000267123.4	CTD-2357A8.3	chr17:78617469–78632057	0.15	1.16	1-1.34	0.0388
ENSG00000267593.1	RP11-108P20.4	chr18:58813880–58834364	-0.22	0.81	0.7–0.93	0.0031
ENSG00000271969.1	U47924.29	chr12:6964949–6965382	-0.16	0.85	0.75–0.97	0.018
ENSG00000225329.2	LHFPL3-AS2	chr7:104894628–104926645	-0.12	0.89	0.81–0.98	0.0192

The six-lncRNA signature and survival analysis in the training set

Acording to the expression level of six lncRNAs, there is a risk-score formula for LUAD patients' survival prediction. The risk score formula is as following: Risk score= (0.37 × expression level of RP11-79H23.3) + (0.21 × expression level of RP11-309M7.1) + (0.15 × expression le-vel of CTD-2357A8.3)+ (-0.22 × expression level of RP11-108P20.4)+ (-0.16 × expression level of U47924.29)+ (-0.12 × expression level of LHFPL3-AS2). Then, we calculated the lncRNA-based risk score for each LUAD patient in the training set and divided LUAD patients into high-risk (n = 117) and low-risk groups (n = 116) by a threshold as the median risk score value. The Kaplan-Meier curves showed that patients in the high-risk group had a worse prognosis th-an those in the low-risk group (28.5 months vs 59.3months, P-value < 0.0001; Fig. 1a). Time dependent ROC curve analysis was measured to estimate the competitive performance of the six lncRNA features, and the AUC score of the six-lncRNA features was 0.746 (Fig. 1b), suggesting the better survival prediction in the training dataset. Univariate cox regression analysi-s showed that the six-lncRNA risk score were significantly associated with patients' survival (P-value < 0.05, HR = 1.41, 95% CI = 1.26–1.58; Table 2). Then, the aggregation effect of samples was shown (red represents tumor, black represents normal sample) in Fig. 1c. Figure 1d showed the gene distribution between expression logFC and CPM (counts of per Million mapped reads). Figure 1e signified the lncRNA gene distribution between expression logFC and FDR (FDR is adjusted P value). Figure 1f showed CPM heat map of differentially expressed lncRNAgenes. Figure 1g showed heatmap of the six lncRNA expression profiles, which were ranked according to the risk score value. Patients with high-risk scores had higher mortality than patients with low-risk scores. For patients with high risk scores, the expression profiles of lncRNAs (CTD-2357A8.3, LHFPL3-AS2) were significantly unregulated, whereas the expressionsof remaining lncRNAs (RP11-79H23.3, RP11-309M7.1, RP11-108P20.4, U47924.29) were down-regulated. Figure 1h signified the risk score scatter plot. Figure 1I signified the follow-up scatter plot.

Table 2: Univariable and multivariable Cox regression analyses in training set.

Variables	Univariable model[a]			Multivariable model		
	HR	95% CI	P	HR	95% CI	P
six-lncRNA risk score	1.41	1.26-1.56	2.47E-09	1.36	1.20-1.55	2.20E-06
Gender	0.92	0.56-1.50	7.36E-01	1.27	0.76-2.13	3.61E-01
Age	1	0.98-1.03	5.60E-01	1.03	1-1.05	5.49E-02
Tumor stage	1.75	1.39-2.19	1.58E-06	1.69	1.30-2.20	7.85E-05
Smoked pack year	1	0.99-1.01	8.37E-01	1	0.99-1.01	9.78E-01
kras status	1.31	0.31-5.49	7.11E-01	1.33	0.31-5.67	6.97E-01

a In both univariable and multivariable Cox regression analyses, risk score, age, gender, stage□smoked pack-year and kras status were evaluated as continuous variables. P<0.05 was considered statistically significant in all analyses.

The survival prediction of the six-lncRNA signature in the testing set and the entire TCGA data set

In the testing set, patients in the high-risk group had significantly shorter survival than those in the low-risk group (33.3 months vs. 57.5 months, P-value = 0.0207; Fig. 2a). ROC curves for the six-lncRNAs-based model got an AUC score of 0.737 in the testing set (Fig. 2c). In the entire TCGA data set, patients in the high-risk group had significantly shorter survival than those in the low-risk group (31.0 months vs. 57.5 months, P-value < 0.0001; Fig. 2b). ROC curves analysis for the six-lncRNAs based model got an AUC score of 0.746 in the entire set(Fig. 2d).

Independence of the six-lncRNA signature and the other clinical factors

We estimated if the survival prediction based on six-lncRNA signature was independent of clinical variables. We used multivariate cox regression analysis to estimate lncRNA-based risk score and other clinical information, such as age, gender, tumor stage, smoked pack-year and kras state (Table 2). The result showed that six-lncRNA risk score is tightly related to survival after regulating the clinical factors. Then, we found that the tumor stages were also significantly related to overall survival. We carried out the stratified analysis. The entire TCGA data set were divided into younger stratum (age ≤ 66, n = 236) and older stratum (age > 66, n = 230), male stratum(MALE, n = 213) and female stratum(FEMALE, n = 253), low smoked pack-year stratum (low smoked-pack year < 25, n = 247) and high smoked pack-year stratum (high smoked-pack year > 25, n = 219), low tumor stages (stage I/II, n = 364) and high tumor stages (stage III/IV, n = 102), no kras state stratum (kras_no, n = 443) and kras state stratum (kras_yes, n = 23). The result showed that the six-lncRNA risk score could divide LUAD patients into high-risk and low-risk subgroup within each stratum. These result indicated that the prognostic value of six-lncRNA signature is independent of age (Fig. 3). The different stratums of age showed statistical significance in risk score, and we can see the stratification analysis of gender (Fig. 4), tumor stage I/II (Fig. 5a), entire tumor stage (Fig. 5c), smoked pack-year (Fig. 6), kras_no state (Fig. 7a) and entire kras state (Fig. 7c) show similar results in the entire set. However, the stratum of tumor stage III/IV and kras_yes state did not show significance in risk score, and the reason may be the number of patients in the two stratums were small. These findings demonstrated that six-lncRNA risk score displayed a competitive survival prediction of LUAD patients.

Functional characteristics of six prognostic lncRNAs

To explore the functional implication of six prognostic lncRNAs in LUAD tumorigenesis, we performed functional category enrichment analysis by analyzing GO and KEGG in Metascape. The biological functions of lncRNAs are still largely unknown. Here, we predicted their potential functions according to the co-expression network. We calculated Spearman correlation coefficients between lncRNAs and protein-coding genes according to their expression values. We selected protein-coding genes as co-expressed partners of six prognostic lncRNAs, whose spearman coefficient > 0.42. At last, a total of 1003 protein-coding genes were significantly correlated with at least one prognostic lncRNAs. Functional enrichment analysis showed that lncRNA correlated protein-coding genes mainly enriched in mitotic

nuclear division, T cell activation, addative immune system, PID IL12 2 pathway and several pathways (Fig. 8a, Table 3). Figure 8b showed dot plot of the functional enriched GO terms where terms containing more genes tend to have big circles; Fig. 8c showed a network of KEGG enriched terms colored by p-value, where terms containing more genes tend to have a more significant p-value. Figure 8d showed the network of enriched terms: colored by cluster-ID, where nodes that share the same cluster-ID are typically close to each other; Fig. 8e showed the network of enriched terms: colored by p-value, where terms containing more genes tend to have a more significant p-value. These results suggested that the six prognostic lncRNAs may be related to immune reaction through regulating protein-coding genes to influence lung tumorigenesis. The PPI network of six lncRNA co-expressed protein-coding genes was constructed by STRING, and the most significant module was obtained using Cytoscape (Fig. 8f).

Table 3
Top 20 clusters with their representative enriched terms (one per cluster).

GO	Category	Description	Count	%	Log10(P)	Log10(q)
GO:0140014	GO Biological Processes	mitotic nuclear division	57	5.75	-24.30	-19.99
GO:0042110	GO Biological Processes	T cell activation	72	7.27	-21.75	-18.03
R-HSA-1280218	Reactome Gene Sets	Adaptive Immune System	87	8.78	-17.47	-14.30
M54	Canonical Pathways	PID IL12 2PATHWAY	24	2.42	-16.87	-13.79
GO:0001816	GO Biological Processes	cytokine production	85	8.58	-15.31	-12.34
GO:0002250	GO Biological Processes	adaptive immune response	76	7.67	-14.70	-11.76
GO:0044770	GO Biological Processes	cell cycle phase transition	71	7.16	-13.87	-11.02
GO:0019221	GO Biological Processes	cytokine-mediated signaling pathway	81	8.17	-13.27	-10.48
GO:0071346	GO Biological Processes	cellular response to interferon-gamma	33	3.33	-12.38	-9.70
R-HSA-5683826	Reactome Gene Sets	Surfactant metabolism	14	1.41	-11.56	-8.97
GO:0043368	GO Biological Processes	positive T cell selection	14	1.41	-10.65	-8.13
GO:0032609	GO Biological Processes	interferon-gamma production	24	2.42	-10.58	-8.07
GO:0008608	GO Biological Processes	attachment of spindle microtubules to kinetochore	13	1.31	-9.84	-7.40
M14	Canonical Pathways	PID AURORA B PATHWAY	14	1.41	-9.69	-7.27

"Log10(P)" is the p-value in log base 10. "Log10(q)" is the multi-test adjusted p-value in log base 10.

GO	Category	Description	Count	%	Log10(P)	Log10(q)
hsa04110	KEGG Pathway	Cell cycle	23	2.32	-8.93	-6.62
hsa04514	KEGG Pathway	Cell adhesion molecules (CAMs)	24	2.42	-8.32	-6.06
GO:0019883	GO Biological Processes	antigen processing and presentation of endogenous antigen	10	1.01	-8.27	-6.02
GO:0051310	GO Biological Processes	metaphase plate congression	15	1.51	-7.88	-5.65
R-HSA-451927	Reactome Gene Sets	Interleukin-2 family signaling	13	1.31	-7.86	-5.64
GO:0045088	GO Biological Processes	regulation of innate immune response	46	4.64	-7.74	-5.53

"Log10(P)" is the p-value in log base 10. "Log10(q)" is the multi-test adjusted p-value in log base 10.

Discussion

lncRNAs play vital roles in tumorigenesis and progression of tumor, and many researchers have studied whether some specific lncRNAs in cancers can be involved in diagnosis and prognosis. Although lncRNA in LUAD tumorigenesis has been reported many times, the competitive prognostic values of lncRNA in LUAD are not fully understood. So, it is necessary to find reliable prognostic biomarkers of LUAD. In our work, a six-lncRNA prognostic feature was identified according to the lncRNA expression of LUAD patients. This study determined the potential six-lncRNA feature to predict the prognosis of LUAD. The performance of six-lncRNA feature was estimated using ROC analysis, displaying that the prognostic of performance the six-lncRNA feature is suitable for survival prediction. ROC analysis has got an AUC of 0.746 in the training set, which indicated high sensitivity and specificity of the six lncRNA model. Further functional annotation of these prognostic lncRNAs correlated protein-coding genes indicated that they might be involved in immune reaction to regulate LUAD. Also, the result showed that the prognostic value of six-lncRNA signature was independent of other clinical factors in LUAD. Although these six prognostic lncRNAs have not been investigated previously in lung cancers, we assume that these lncRNAs may be related to LUAD, and many studies are needed to validate the deep mechanism in the future.

Previous works have reported some prognostic lncRNAs in LUAD, such as LOC100132354, CTB-193M12.5, NEAT1, HOTAIR, H19, NNT-AS1, linc01088, TP73-AS1, TINCR, LINC00336, [8–9, 22–29], which are not included in TCGA-based prognostic lncRNAs. According to LNCipedia[30], lncRNA RP11-79H23.3 is an intergenic non-coding RNA. One study showed that lncRNA RP11-79H23.3 was significantly overexpressed in old tendon[31]. Another study reported that lncRNA RP11-79H23.3 may be a sponge for

miR-107 to suppress tumorigenesis of bladder cancer[32]. LncRNA CTD-2357A8.3 is also an intergenic non-coding RNA[30], and it was reported to be one of eight prognostic lncRNAs of esophageal cancer[33]. LncRNA RP11-309M7.1 and lncRNA RP11-108P20.4 are also intergenic non-coding RNAs, and they have not been previously reported. LncRNA U47924.29 and lncRNA LHFPL3-AS2 are antisense lncRNAs, and they have not been previously reported[30].

We should explain some limitations of our study. First, we only analyzed the prognostic power of the six-lncRNA feature in the TCGA data, and the deep mechanisms of lncRNAs need to be further explored.

Conclusion

We have explored the prognostic values of lncRNAs in LUAD patients in our study. Our result has suggested that the six-lncRNA signature is useful in predicting the clinical outcome, and may be effective prognostic biomarkers in LUAD patients survival prediction.

Abbreviations

A. NSCLC
non-small cell lung cancer; LUAD:lung adenocarcinoma;AUC, area under the curve; MCODE:molecular complex detection; GO:gene ontology;KEGG:Kyoto Encyclopedia of Gene and Genome; TCGA:The Cancer Genome Atlas; RPKM:Reads Per Kilobase of exon model per Million mapped reads;MDS:multidimensional scaling;CPM:counts of per Million mapped reads;

Declarations

Acknowledgment

We sincerely thank the researchers for providing their databases information online, it is our pleasure to acknowledge their contributions.

Disclosure

The authors report no conflicts of interest in this work.

Author Contributions

Yang Wang conceived, designed, analyzed the data, and write the manuscript. Chengping Hu conceptualized and developed an outline for the manuscript and revised the manuscript. All authors read and approved the final manuscript.

Funding

This investigation was supported by National Key R&D Program of China (NO.2018YFC1311900).

Data Availability

All data included in this study are available upon request by contact with the corresponding author.

Code availability

All code included in this study are available upon request by contact with the corresponding author.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Bouzbid S, Hamdi-Chérif M, Zaidi Z, et al (2018) Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet 391(10125):1023-1075.
2. Gloss BS, Dinger ME (2016) The specificity of long noncoding RNA expression. Biochim Biophys Acta 1859:16–22.
3. Malih S, Saidijam M, Malih N (2016) A brief review on long noncoding RNAs: a new paradigm in breast cancer pathogenesis, diagnosis and therapy. Tumour Biol 37:1479–85.
4. Rinn JL, Chang HY(2012) Genome regulation by long noncoding RNAs. Annu Rev Biochem 81:145–66.
5. Li CH, Chen Y (2013) Targeting long non-coding RNAs in cancers: progress and prospects. Int J Biochem Cell Biol 45:1895–910.
6. Xiaohui Su, Jianjun Zhang, Xianfeng Luo,et al (2019) LncRNA LINC01116 Promotes Cancer Cell Proliferation, Migration And Invasion In Gastric Cancer By Positively Interacting With lncRNA CASC11.Onco Targets Ther 12: 8117–8123.
7. Ze Zhang, Shouqian Wang, Fan Yang, et al (2020) LncRNA ROR1-AS1 high expression and its prognostic significance in liver cancer.Oncol Rep 43(1): 55–74.
8. Yumin, Wang, Fan, et al (2018) lncRNA LOC100132354 promotes angiogenesis through VEGFA/VEGFR2 signaling pathway in lung adenocarcinoma.Cancer Manag Res 10: 4257–4266.
9. Wang X, Li G , Luo Q, et al (2018) Integrated TCGA analysis implicates lncRNA CTB-193M12.5 as a prognostic factor in lung adenocarcinoma. Cancer Cell International 18(1):27.

10. Derrien T, Johnson R, Bussotti G, et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22:1775–89.
11. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–40.
12. Yingyao Zhou, Bin Zhou, Lars Pache (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nature Communications 10(1):1523.
13. Zar, J.H□Hochberg Y, Benjamini Y, et al (1990) More powerful procedures for multiple significance testing. Statistics in Medicine 9:811-818.
14. Cohen, J (1960) A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20:27-46.
15. Paul Shannon, Andrew Markiel, Owen Ozier, et al(2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11):2498-2504.
16. Chris Stark, Bobby-Joe Breitkreutz, Teresa Reguly, et al (2006) BioGRID: a general repository for interaction datasets. Nucleic Acids Res 34:D535-539.
17. Taibo Li, Rasmus Wernersson, Rasmus B Hansen,et al (2017) A scored human protein-protein interaction network to catalyze genomic interpretation. Nat. Methods 14(1):61-64.
18. Taibo Li, Rasmus Wernersson, Rasmus B Hansen, et al (2016) A scored human protein-protein interaction network to catalyze genomic interpretation. Nat. Methods 14(1):966-967.
19. Gary D Bader, Christopher W V Hogue (2003) An automated method for finding molecular complexes in large protein interaction networks. BMC bioinformatics 4:2.
20. Franceschini A, Szklarczyk D, Frankild S, et al (2013) STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41 D808-D815.
21. Smoot ME, Ono K, Ruscheinski J, et al (2011) Cytoscape 2.8: New features for data integration and network visualization. Bioinformatics 27: 431-432.
22. Zhou W, Chen X, Hu Q, et al (2018) Galectin-3 activates TLR4/NF-κB signaling to promote lung adenocarcinoma cell proliferation through activating lncRNA-NEAT1 expression. BMC Cancer 18(1):580.
23. Wang R, Shi Y, Chen L, et al (2015) The ratio of FoxA1 to FoxA2 in lung adenocarcinoma is regulated by LncRNA HOTAIR and chromatin remodeling factor LSH. Scientific Reports 5:17826.
24. Liu L, Liu L, Lu S (2019) lncRNA H19 promotes viability and epithelial-mesenchymal transition of lung adenocarcinoma cells by targeting miR-29b-3p and modifying STAT3. International Journal of Oncology 54 (3), 929-941.
25. He W, Zhang Y, Xia S, et al (2020) LncRNA NNT-AS1 promotes non-small cell lung cancer progression through regulating miR-22-3p/YAP1 axis. Thorac Cancer 11 (3), 549-560.
26. Liu JQ, Feng YH, Zeng S,et al (2020) Linc01088 promotes cell proliferation by scaffolding EZH2 and repressing p21 in human non-small cell lung cancer. Life Sci 241:117134.
27. Chen C, Shu L, Zou W,et al(2019) Role of long non-coding RNA TP73-AS1 in cancer. Biosco Rep 39(10). pii: BSR20192274.

28. Ya-Wen Gao, Fang Ma, Yang-Chun Xie, et al (2019) Sp1-induced upregulation of the long noncoding RNA TINCR inhibits cell migration and invasion by regulating miR-107/miR-1286 in lung adenocarcinoma. Am J Transl Res 11(8): 4761–4775.
29. Wang M, Mao C, Ouyang L, et al (2019) Correction to: Long noncoding RNA LINC00336 inhibits ferroptosis in lung cancer by functioning as a competing endogenous RNA. Cell Death Differ 27 (4), 1447.
30. Pieter-Jan V , Jasper A , Kenneth V , et al (2019) LNCipedia 5: towards a reference set of human long non-coding RNAs. Nucleic Acids Res 47(D1):D135-D139.
31. Peffers M, Fang Y, Cheung K , et al (2015) Transcriptome analysis of ageing in uninjured human Achilles tendon. Arthritis Research & Therapy 17(1):33.
32. Hong C , Rui Y , Xiaying Z , et al (2018) LncRNA RP11-79H23.3 Functions as a Competing Endogenous RNA to Regulate PTEN Expression through Sponging hsa-miR-107 in the Development of Bladder Cancer. International Journal of Molecular Sciences 19(9):2531.
33. Fan Q, Liu B (2016) Identification of a RNA-Seq Based 8-Long Non-Coding RNA Signature Predicting Survival in Esophageal Cancer. Medical science monitor 22:5163-5172.

Figures

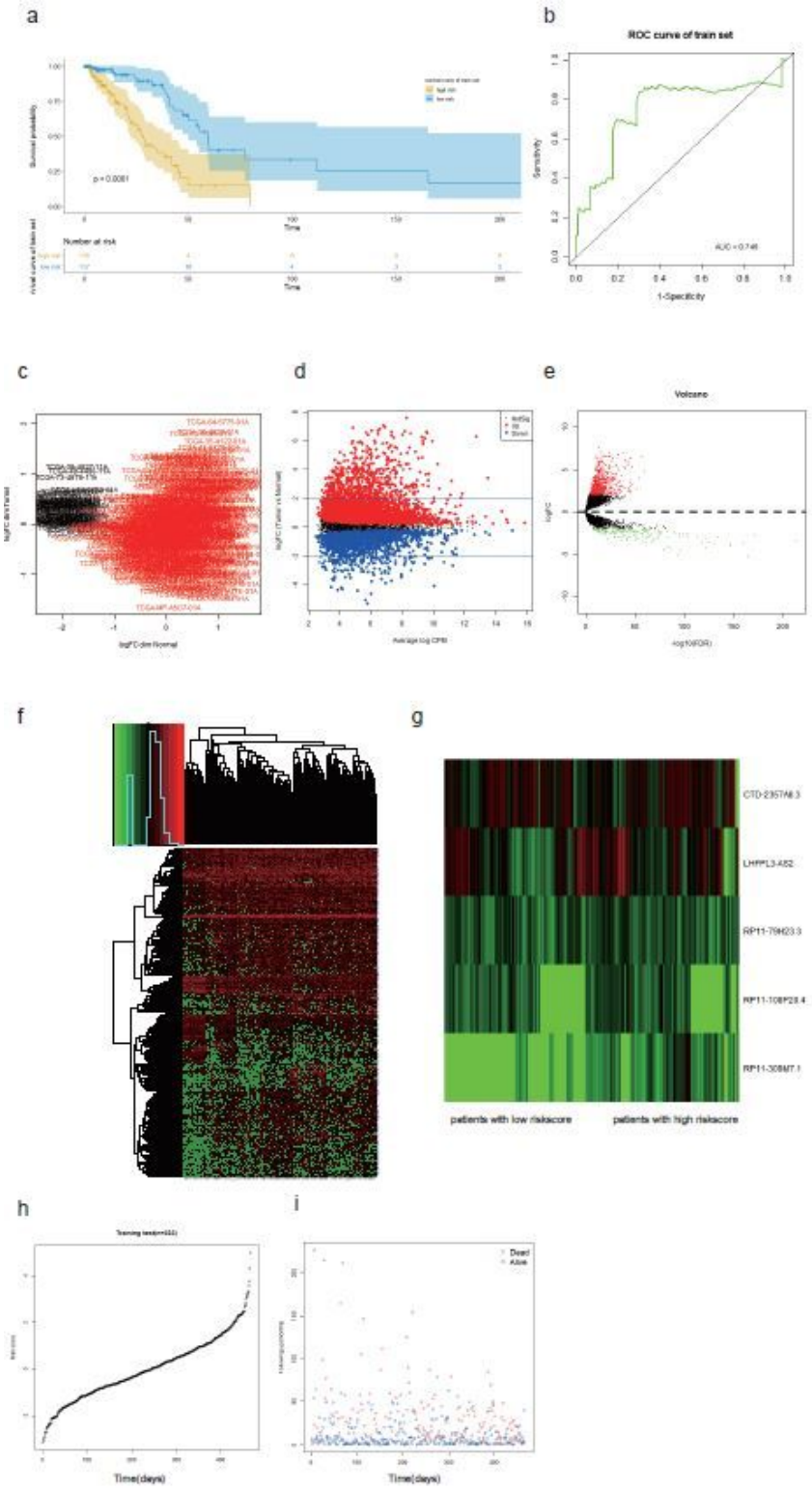

Figure 1

The risk score model of six-lncRNA predicts overall survival of patients with LUAD in the training set. (a)Kaplan-Meier estimates plots of the survival of the LUAD patients in high- and low-risk groups. The P-value indicates the differences in the two curves from the results of two sided log-rank tests. The number below the curve is the number of the patients in the high- and low-risk groups; (b).The Receiver operating characteristic (ROC)analysis of risk score for survival prediction in the training set. The area below the

curve (AUc) was calculated for ROC curves. (c)MDS(multidimensional scaling).(d).The volcano map of CPM(counts of per Million mapped reads) and logFC. (e).The volcano map between logFC and FDR. (f).CPM heat map of differentially expressed genes.(g). heatmap of the six lncRNA expression profiles. (h)risk score scatter plot.I. follow-up scatter plot.

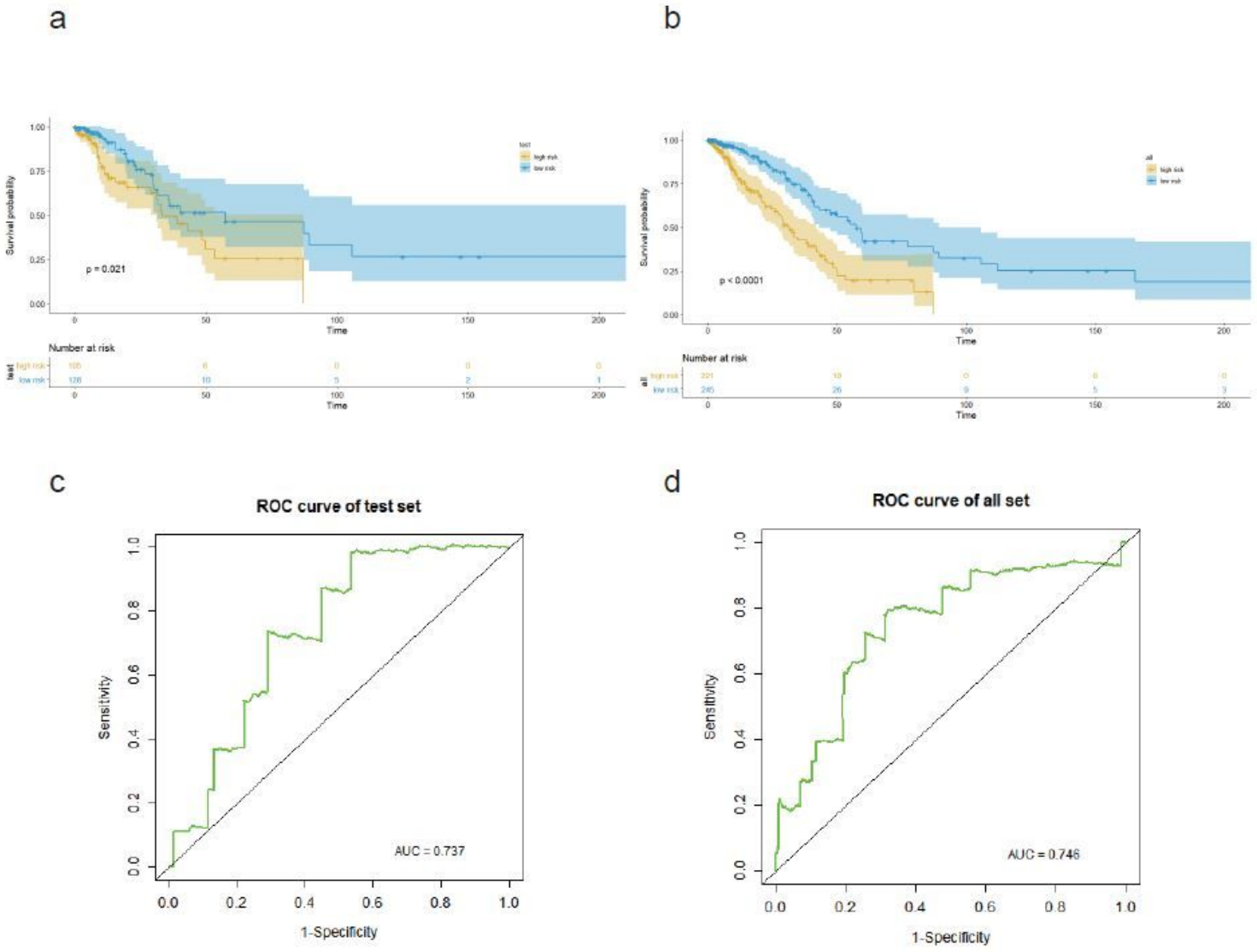

Figure 2

The six-lncRNA related risk score model predicts overall survival of patients with LUAD in the testing set and the entire set. (a) Kaplan-Meier plots of the survival of the LUAD patients using the six-lncRNA signature-related risk score model in the testing set . (b) Kaplan-Meier plots of the survival of the LUAD patients using the six-lncRNA signature-related risk score mode in the entire set.(c) ROC analysis of risk score for survival prediction in the testing set (n=233).(d) ROC analysis of risk score for survival prediction in the entire set (n=466).

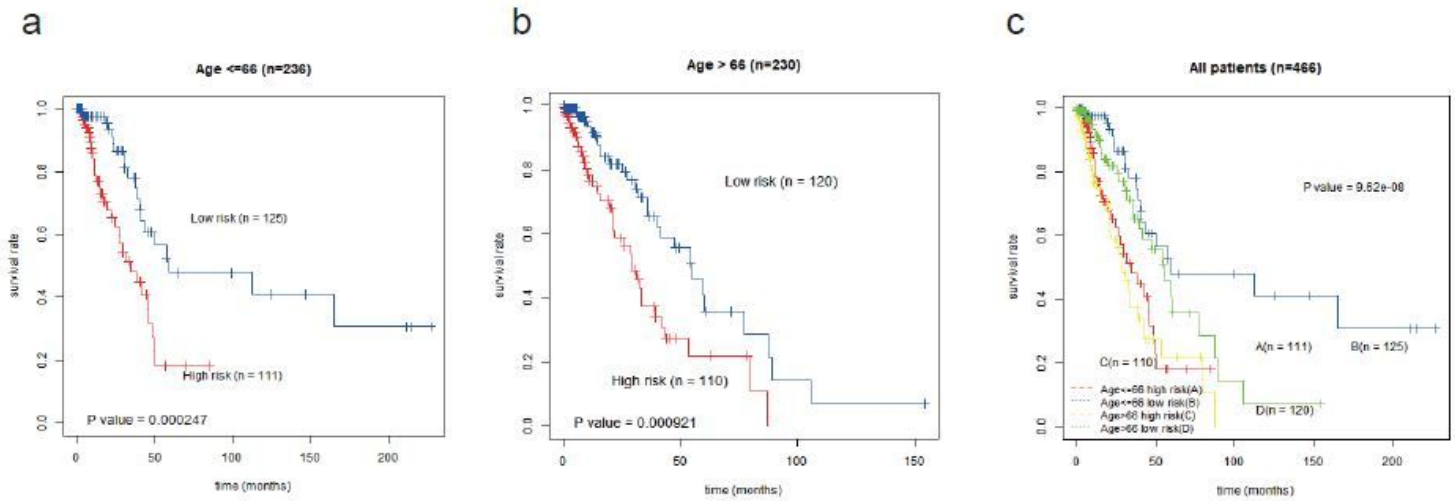

Figure 3

Stratification analyses of all patients adjusted to age using the six-lncRNA signature. (a) The Kaplan-Meier plot of the younger patients with LUAD (age <= 66, n=236). (b) The Kaplan-Meier plot of the elder patients with LUAD (age>66, n=230). (c) The Kaplan-Meier plot of the entire patients with LUAD(N=466). The number below the curve is the number of the patients in the high- and low-risk group. The P-value represents the differences between the curves from the results of two-sided log-rank tests.

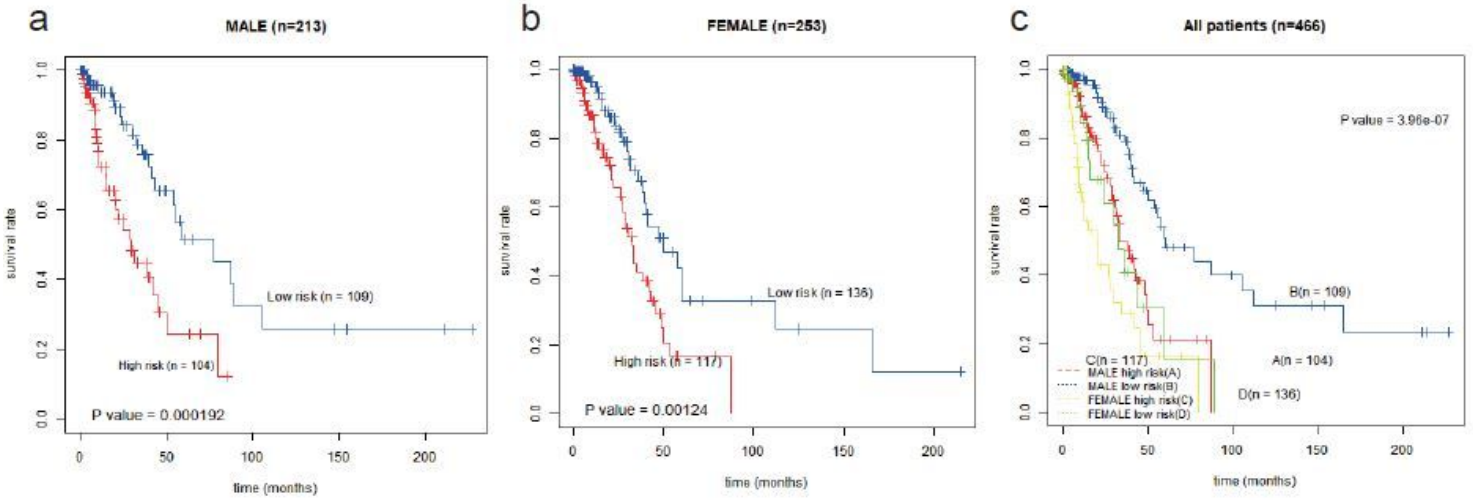

Figure 4

Stratification analyses of all patients adjusted to gender using the six-lncRNA signature. (a) The Kaplan-Meier plot of male with LUAD (MALE, n=213). (b) The Kaplan-Meier plot of female with LUAD (FEMALE, n=253). (c) The Kaplan-Meier plot of the entire patients with LUAD (N=466). The number below the curve is the number of the patients in the high- and low-risk group. The P-value represents the differences between the curves from the results of two-sided log-rank tests.

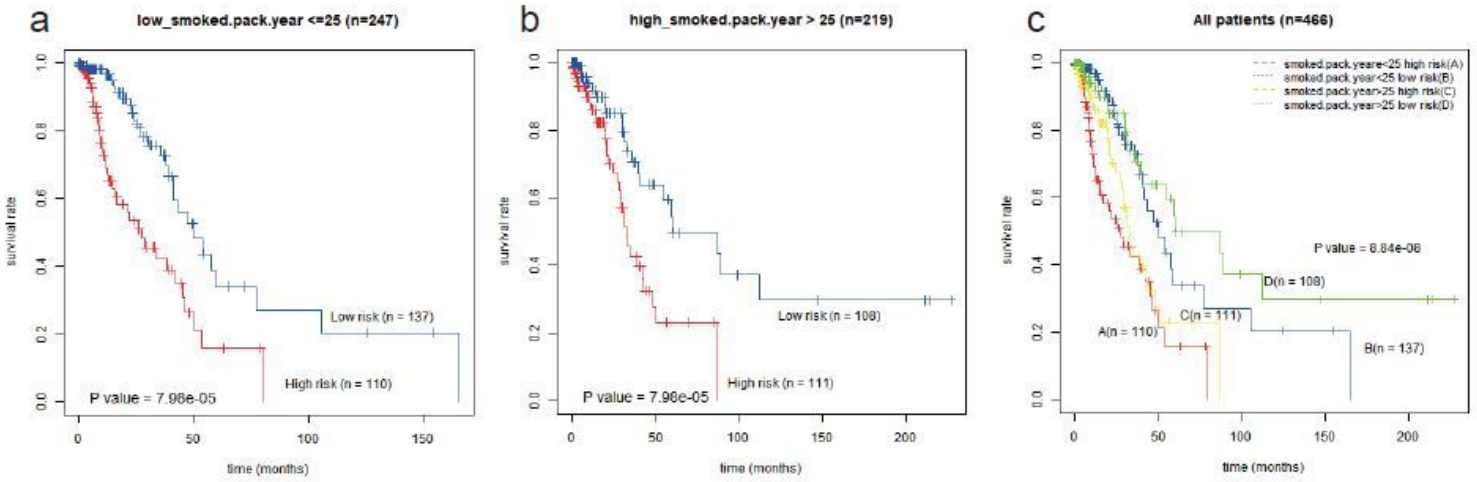

Figure 5

Stratification analyses of all patients adjusted to smoked pack-year using the six-lncRNA signature. (a) The Kaplan-Meier plot of the low smoked pack-year patients with LUAD (low_smoked.pack.year, n=247). (b) The Kaplan-Meier plot of the high smoked pack-year patients with LUAD (high_smoked.pack.year, n=219). (c) The Kaplan-Meier plot of the entire patients with LUAD (N=466). The number below the curve is the number of the patients in the high- and low-risk group. The P-value represents the differences between the curves from the results of two-sided log-rank tests.

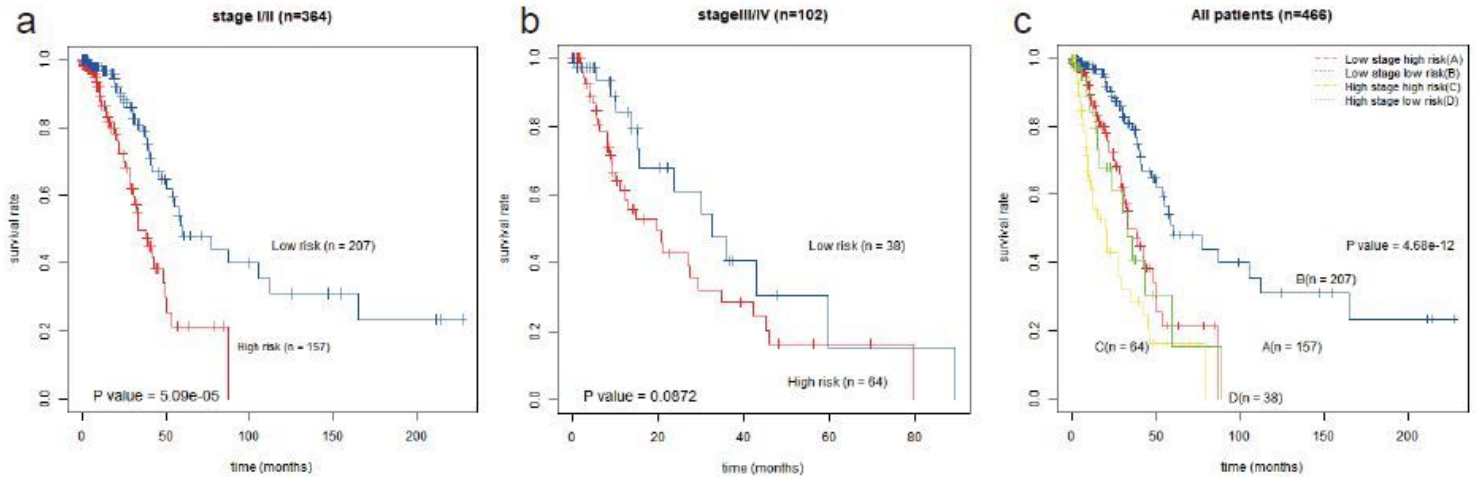

Figure 6

Stratification analyses of all patients adjusted to tumor stage using the six-lncRNA signature. (a) The Kaplan-Meier plot of the stage I/II patients with LUAD (stageI/II, n=364). (b) The Kaplan-Meier plot of the stage III/IV patients with LUAD (stage III/IV, n=102). (c) The Kaplan-Meier plot of the entire patients with LUAD (N=466). The number below the curve is the number of the patients in the high- and low-risk group. The P-value represents the differences between the curves from the results of two-sided log-rank tests.

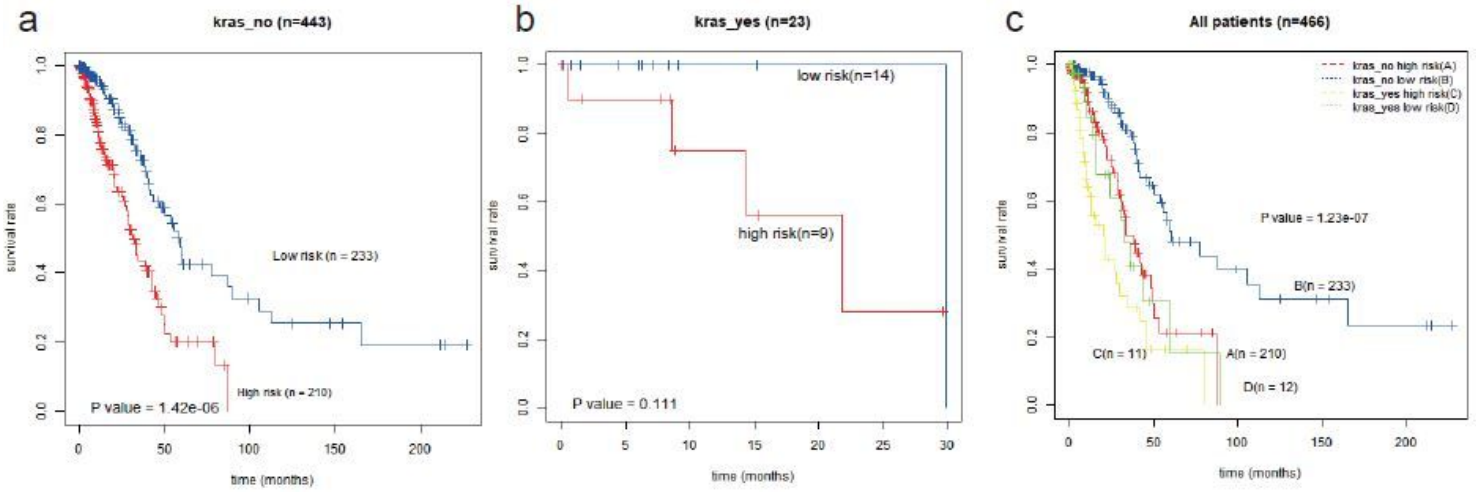

Figure 7

Stratification analyses of all patients adjusted to kras state using the six-lncRNA signature. (a) The Kaplan-Meier plot of the kras_no patients with LUAD (kras_no, n=443). (b) The Kaplan-Meier plot of the kras_yes patients with LUAD (kras_yes, n=23). (c) The Kaplan-Meier plot of the entire patients with LUAD (N=466). The number below the curve is the number of the patients in the high- and low-risk group. The P-value represents the differences between the curves from the results of two-sided log-rank tests.

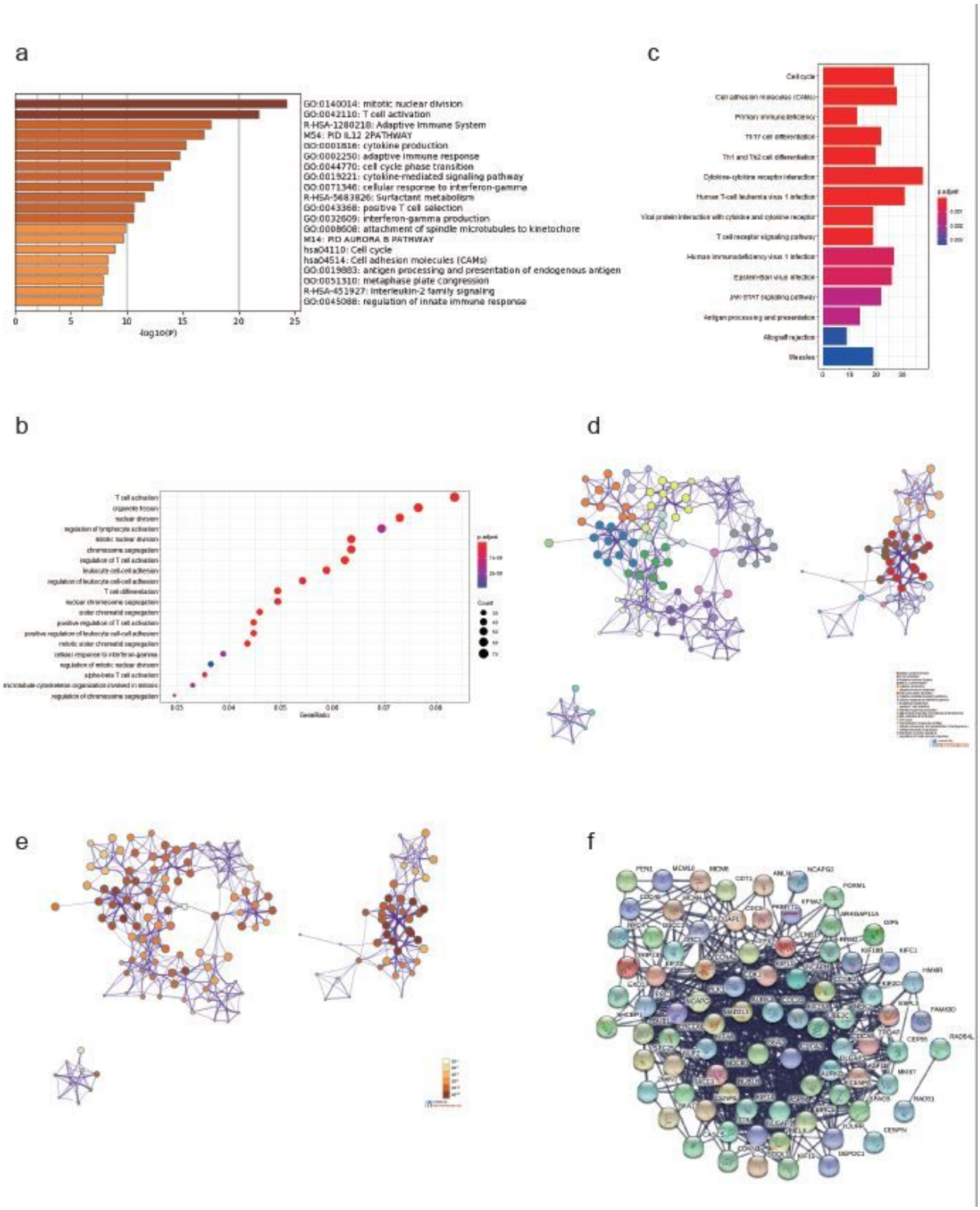

Figure 8

The results of functional enrichment analysis of the six lncRNA co-expressed protein-coding genes. (a) Bar graph of top 20 enriched terms , colored by p-values;(b) Dot plot of the functional enriched GO terms ; (c) Network of KEGG enriched terms colored by p-value, where terms containing more genes tend to have a more significant p-value ;(d) Network of enriched terms: colored by cluster ID, where nodes that share the same cluster ID are typically close to each other; (e) Network of enriched terms:colored by p-value,

where terms containing more genes tend to have a more significant p-value.□f□Gene–gene interaction network among the most significant module of six lncRNA co-expressed protein-coding genes.(STRING and Cytoscape).