

# Lower expression of LINC00092 in lung adenocarcinoma might mean poorer prognosis

## A study based on data mining and bioinformatics

Xin Wang, PhD<sup>a,\*</sup>, Xianwei Wang, Dr<sup>b</sup>, Guichuan Huang, PhD<sup>c</sup>, Yi Chen, PhD<sup>c</sup>

### Abstract

The mechanisms that underlie long non-coding RNA 00092 (LINC00092) in lung adenocarcinoma (LUAD) remain unclear. In this study, by mining the Cancer Genome Atlas and Gene Expression Omnibus databases and using bioinformatics tools, we try to elucidate the function of LINC00092 in LUAD.

The Cancer Genome Atlas and gene expression Omnibus microarray datasets were used to analyze and evaluate the expression of LINC00092 in LUAD and its clinical significance. Clinical samples were collected and the relative expression level of LINC00092 were identified by quantitative real time polymerase chain reaction. The LINC00092 related genes were identified by Multi Experiment Matrix, The Atlas of ncRNA in Cancer and the database of RNA-Binding Protein specificities. The predicted genes were then sent to the Gene Ontology enrichment and the Kyoto Encyclopedia of Genes and Genomes pathway analysis.

The expression of LINC00092 was significantly decreased in LUAD tissues compared to non-tumor tissues (standard mean difference = -1.10, 95% confidence interval: -1.87 to -0.32,  $P < .001$ , random). Low expression of LINC00092 was associated with the poor overall survival (hazard ratio = 1.32, 95% confidence interval: 1.08–1.62,  $P < .05$ , fixed) and high pathological stage ( $P < .05$ ). The relative expression level of LINC00092 in clinical samples were significantly lower in LUAD tissues compared with adjacent normal tissues. ( $P < 0.05$ ) 61 LINC00092 related genes were identified; the Kyoto Encyclopedia of Genes and Genomes analysis showed that the most significant signaling pathways were: NF- $\kappa$ B, HIF-1 and ErbB signaling pathways.

In this study, we found that the decrease of LINC00092 expression was involved in LUAD tumorigenesis and metastasis, and the depletion of LINC00092 was associated with a poor prognosis in patients with LUAD. The mechanisms that underlie LINC00092 in LUAD might be related to the NF- $\kappa$ B, HIF-1 and ErbB signaling pathways.

**Abbreviations:** CC = cellular component, CI = confidence interval, GEO = gene expression Omnibus, GO = gene ontology, HR = hazard ratio, KEGG = the Kyoto Encyclopedia of Genes and Genomes, LINC00092 = long non-coding RNA 00092, LUAD = lung adenocarcinoma, MEM = multi experiment matrix, PPI = protein-to-protein interaction, RBPDB = the database of RNA-binding protein specificities, TANRIC = the Atlas of ncRNA in cancer, TCGA = the Cancer Genome Atlas.

**Keywords:** bioinformatics, lung adenocarcinoma, long non-coding RNAs, long non-coding RNA 00092

Editor: Martin S. Staeger.

The authors have no funding and conflicts of interest to disclose.

All samples collection and analyzing have been approved by the ethics institute of the People's Hospital of Deyang City and all the informed consents of the participants have been obtained.

<sup>a</sup> Department of Respiratory Medicine, <sup>b</sup> Department of Pathology, People's Hospital of Deyang City, Deyang, Sichuan Province, <sup>c</sup> Department of Respiratory Medicine, the First Affiliated Hospital of Chongqing Medical University, Chongqing, Peoples R China.

\* Correspondence: Xin Wang, Department of Respiratory Medicine, People's Hospital of Deyang City, 137 North Taishan Road, Deyang 618000, Sichuan Province, Peoples R China (e-mail: wang1980xin@163.com).

Copyright © 2020 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Wang X, Wang X, Huang G, Chen Y. Lower expression of LINC00092 in lung adenocarcinoma might mean poorer prognosis: a study based on data mining and bioinformatics. *Medicine* 2020;99:44(e23012).

Received: 13 October 2019 / Received in final form: 19 January 2020 /

Accepted: 2 October 2020

<http://dx.doi.org/10.1097/MD.00000000000023012>

## 1. Introduction

Lung cancer is the most frequent cause of cancer-related deaths worldwide. About 1.8 million people were diagnosed with lung cancer, and 1.6 million people died as a result of the disease every year. Lung adenocarcinoma (LUAD), as the most common histologic type of lung cancer, accounts for almost half of the lung cancers.<sup>[1,2]</sup> Despite the development of various detection measures and clinical treatments, LUAD remains a major public health problem. Lots of patients with LUAD are diagnosed with advanced stage or distant metastasis when they initially seek medical treatment. The overall cure rate and 5-year survival rate of LUAD is still slim.<sup>[3,4]</sup>

Long non-coding RNAs (lncRNAs) are a class of non-coding RNAs that have no protein-coding capacity and are more than 200 nucleotides in length. lncRNAs play an important role in regulating gene expression at various levels, such as the regulation of the targeted gene transcription, post-transcriptional function as well as epigenetic regulation of genes.<sup>[5–7]</sup> Recent studies have reported that lncRNAs have been involved in the process of occurrence, progression and treatment of various cancers, such as breast cancer, leukemia, colorectal cancer, gastric cancer, and liver cancer.<sup>[8–11]</sup> Therefore, the discovery and

identification of novel functional lncRNAs as candidates for diagnosis and intervention of LUAD is of great significance.

Public databases such as the Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) have been proved valuable tools because they allow us to access large amounts of high-throughput data and related clinical data. Those data are informative and far from being fully applied. By analyzing these data, we can identify numerous prognostic biomarkers for various malignancies. In this study, we tried to use TCGA and GEO databases as source of relevant data and clinical information, and followed by using multiple public bioinformatics tools to explore the potential role of long non-coding RNA 00092 (LINC00092) in LUAD tumorigenesis and prognosis.

## 2. Materials and methods

### 2.1. Data acquired from TCGA

The relevant LINC00092 data in LUAD tumor tissues and normal tissues were obtained from the TCGA database (<https://portal.gdc.cancer.gov/>) up to December 2019. Primary Site: bronchus and lung; Program: TCGA; Project: TCGA-LUAD; Disease Type Key words adenomas and adenocarcinomas, acinar cell neoplasms, cystic, mucinous and serous neoplasms. The RNA expression data were classified as Level 3. All RNA-Seq and clinical data were chosen and archive-building function was performed. All quantification files with “.txt” were selected, and all files in the data and clinical sections were examined.

### 2.2. Data acquired from GEO

GEO databases, including GEO datasets (<http://www.ncbi.nlm.nih.gov/gds/>) and GEO profiles (<http://www.ncbi.nlm.nih.gov/geo/profiles/>), were also used to gain the relevant LINC00092 data. The Key words “lung cancer”, “non-small cell lung cancer”, “NSCLC”, “lung adenocarcinoma(s)”, “lncRNA”, “long non-coding RNA” and “LINC00092” were used as search terms. The following criteria were strictly observed:

- 1) study type: Expression profiling by array;
- 2) datasets were obtained from Homo sapiens;
- 3) datasets included LUAD tissues and non-tumor tissues;
- 4) for analyzing of expression, the datasets included more than 5 samples in each group; for analyzing of hazard ratio (HR), the datasets included no less than 20 samples;
- 5) datasets provided overall survival (OS) information;
- 6) the expression data of LINC00092 from the cases and the controls were provided or could be calculated;
- 7) the non-small cell lung cancer (NSCLC) datasets should contain a clearly defined subset of adenocarcinomas.

Potentially relevant datasets were further screened by reviewing titles, summary and overall design. Relevant full articles and samples information were evaluated in order to identify studies that met the eligible criteria. Data for eligible datasets and series were up to December 2019.

### 2.3. Confirmation by quantitative real time polymerase chain reaction

According to the results of TCGA and GEO database analysis, clinical tissue samples of LUAD (n=20) and their normal adjacent (n=20) were obtained. All samples collection and analyzing have been approved by the ethics institute of the

**Table 1**

**Primer sequences used for quantitative real time polymerase chain reaction.**

Name	Sequence 5' to 3'
LINC00092-F	CGATTGGTTTGCCTTTG
LINC00092-R	CAGGTGGATAAGATGGGAGGAG
GAPDH-F	CACTCCTCCACCTTTGACGC
GAPDH-R	CTGTGCTGTAGCCAAATTCGT

LINC00092 = long non-coding RNA 00092.

People's Hospital of Deyang City and all the informed consents of the participants have been obtained. Total RNA was extracted by RNeasy FFPE Kit (Cat. No. 73504, QIAGEN, Shanghai, China). Complementary DNAs were generated using the QuantiNova Reverse Transcription Kit (Cat. No. 205411, QIAGEN, Shanghai, China). Quantitative real-time PCR were conducted by using the QuantiNova SYBR Green PCR Kit (Cat. No. 208054, QIAGEN, Shanghai, China) on GENTIER 96E PCR system (Tianlong, Xian, China). Relative expression level of LINC00092 were analyzed by using the  $2^{-\Delta\Delta Ct}$  method. The PCR primers were designed and synthesized by Takara (TAKARA, Dalian, China). The human GAPDH was used as endogenous controls for lncRNA expression in this analysis. All PCR primers are listed in Table 1.

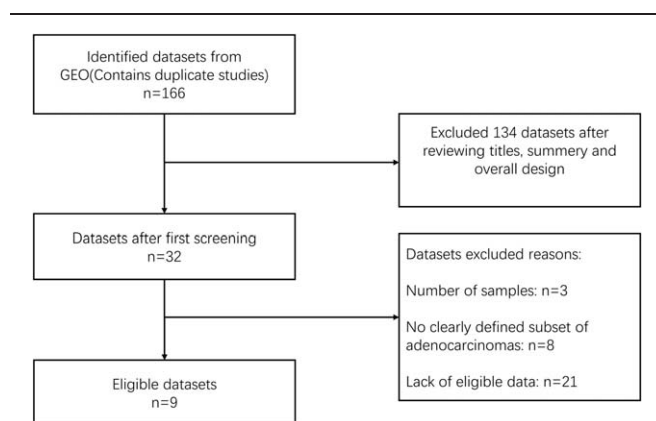
### 2.4. Statistical analysis

LINC00092 expression data were presented as mean  $\pm$  standard deviation, and SPSS v25 software (IBM Corporation, Armonk, NY, USA) was used for data analysis. Independent-samples *t* test was employed for the analysis of LINC00092 expression between 2 groups. To evaluate the relationship between LINC00092 expression and OS in LUAD patients, the Kaplan–Meier method and log-rank *P*-value were calculated with the analysis tool SPSS Statistics 25. The cutoff value of LINC00092 expression was determined by its median value. Survival curves were plotted with GraphPad Prism v8.0 (GraphPad Software, Inc., La Jolla, CA, USA) and also harvest from the analysis tools, which can be accessed online at OncoLnc (<http://www.oncolnc.org/>).<sup>[12]</sup> GraphPad Prism 8 was also used to complete the box plot and symbols & lines plot of clinical samples.

Meta-Analyses were accomplished by using STATA software package v15.0 (Stata Corporation, College Station, TX, USA). All data obtained from TCGA and GEO datasets were estimated as standard mean difference with a 95% Confidence Interval (95% CI). HR and 95% CI were calculated to represent the prognosis of LUAD with expression of LINC00092. Effect sizes were pooled with a random or fixed effects model. HR>1 represented a poor prognosis for the group with low LINC00092 expression and would be considered to be statistically significant if the 95% CI did not overlap 1. *P*<.05 was considered to a significant difference.

### 2.5. Genes that related to LINC00092 in LAUD

Multi Experiment Matrix (MEM) (<http://biit.cs.ut.ee/mem/>) is a web bioinformatic tool, it can identify lncRNAs related genes by gathering publicly available gene expression datasets from the ArrayExpress (<http://www.ebi.ac.uk/arrayexpress/>) database.<sup>[13]</sup> The Atlas of ncRNA in Cancer (TANRIC) (<https://ibl.mdander>



**Figure 1.** The process of data identification and acquisition. GEO= gene expression Omnibus.

son.org/tanric/\_design/basic/index.html) is an open-access web resource for interactive exploration of lncRNAs in cancer.<sup>[14]</sup> These 2 bioinformatic tools along with the database of RNA-Binding Protein specificities (RBPDB) (<http://rbpdb.ccrb.utoronto.ca/>),<sup>[15]</sup> a collection of experimental observations of RNA binding sites, were used to determine genes that related to LINC00092 in LUAD. Repeat genes from any 2 of those 3 sites were included as LINC00092 related genes in LUAD.

**Table 2**

**Basic information about enrolled datasets.**

Datasets	Information				Number of enrolled samples
	LINC00092 expression data in tumor samples	LINC00092 expression data in non-tumor samples	OS information		
GSE 50081	Y	N	Y		127
GSE 29013	Y	N	Y		30
GSE 3141	Y	N	Y		58
GSE 118370	Y	N	N		12
GSE 27262	Y	Y	N		50
GSE 40791	Y	Y	N		194
GSE 19188	Y	Y	Y		110
GSE 30219	Y	Y	Y		99
GSE 31210	Y	Y	Y		246
TCGA	Y	Y	Y		592

GEO= gene expression Omnibus, LINC00092 = long non-coding RNA 00092, N=no, OS=overall survival, TCGA=the cancer genome atlas, Y=yes.

**Table 3**

**Characteristics of long non-coding RNA 00092 expression profiling datasets obtained from gene expression Omnibus and the cancer genome atlas.**

Datasets	No of tumor samples	No of normal samples	Tumor tissue		Normal tissue	
			Mean	SD	Mean	SD
GSE 118370	6	6	6.9911	0.8016	7.7549	1.0297
GSE27262	25	25	0.8644	0.1472	1.1048	0.1244
GSE40791	94	100	3.0318	0.2528	3.0886	0.1640
GSE19188	45	65	0.0208	0.1857	-0.0059	0.2015
GSE30219	85	14	4.3188	0.2017	4.6061	0.2679
GSE31210	226	20	72.7700	25.9322	109.4831	29.4056
TCGA	533	59	6566.9673	7211.8636	22631.9661	8569.0259

GEO= gene expression Omnibus, TCGA=the cancer genome atlas, SD=standard deviation.

## 2.6. The comprehensive analysis of LINC00092 related genes

Gene ontology (GO) analysis, including 3 major groups: biological processes, cellular component (CC) and molecular functions, were performed by using DAVID website (<https://david.ncifcrf.gov/>, version 6.8).<sup>[16]</sup> Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was implemented in KOBAS website (<http://kobas.cbi.pku.edu.cn/>, version 3.0).<sup>[17]</sup> The protein-to-protein interaction (PPI) network analysis was conducted by STRING website (<https://string-db.org/cgi/input.pl>).<sup>[18]</sup>

## 3. Results

### 3.1. Basic information, expression and HRs of LINC00092 in LUAD

According to the inclusion criteria, 9 GEO series (GSE) in the GEO database were obtained for this study. The evaluation process of the datasets was shown in Figure 1. With the combination of TCGA and GEO data, a total number of 1518 LUAD samples were enrolled. Sample sizes ranged from 12 to 592 cases. Detailed information was shown in Tables 2–4 and Figure 2.

Combined with TCGA and GEO data, Meta-analysis revealed that the expression of LINC00092 was significantly down-regulated in LUAD tumor tissues compared with non-tumor tissues (standard mean difference = -1.10, 95% CI: -1.87 to -0.32,  $P < .001$ , random) (Fig. 3A). The low expression of LINC00092 was associated with poorer OS in LUAD patients (HR = 1.32, 95% CI: 1.08–1.62,  $P < .05$ , fixed) (Fig. 3B).

### 3.2. Analysis of clinicopathological parameters based on TCGA

The analysis of LINC00092 expression in TCGA between 2 groups showed that the expression level of LINC00092 was lower in tumor tissues ( $6566.9673 \pm 7211.8636$ ), compared with normal tissues ( $22631.9661 \pm 8569.0259$ ;  $P < .001$ ). Additionally, in the pathologic T group, LINC00092 expression was significantly downregulated in T2-T4 group ( $5909.8924 \pm 6891.3360$ ), which represented for bigger tumor size, compared with the group T1 ( $8004.4016 \pm 7789.4820$ ;  $P = .003$ ). Furthermore, the expression of LINC00092 was significantly down-regulated in pathologic stage III-IV group ( $5365.8887 \pm 6987.8984$ ), compared with the group I-II ( $6949.9821 \pm 7346.8671$ ;  $P = .044$ ). However, LINC00092 was not associated

**Table 4****Analysis of overall survival in datasets obtained from gene expression Omnibus and the cancer genome atlas.**

Datasets	No of samples	No of high expression	No of low expression	HR	95% CI	P-value
GSE50081	127	84	83	0.9686	0.5586–1.679	.9095
GSE29013	30	15	15	1.544	0.3738–6.382	.5326
GSE30219	85	43	42	0.9577	0.5234–1.753	.8886
GSE31210	226	113	113	3.193	1.645–6.199	.0015
GSE19188	40	20	20	0.5986	0.2666–1.344	.2136
GSE3141	58	29	29	1.306	0.6518–2.615	.4489
TCGA	506	253	253	1.45	1.084–1.940	.0123

CI=Confidence Interval, GEO= gene expression Omnibus, HR=Hazard Ratio, TCGA=the cancer genome atlas.

with age, gender and “pathologic N”. Clinicopathological characteristics are presented in Table 5. Survival curves found a marked difference between high and low LINC00092 expression group ( $P < .05$ ). (Fig. 4).

### 3.3. Relative expression level of LINC00092 in clinical samples

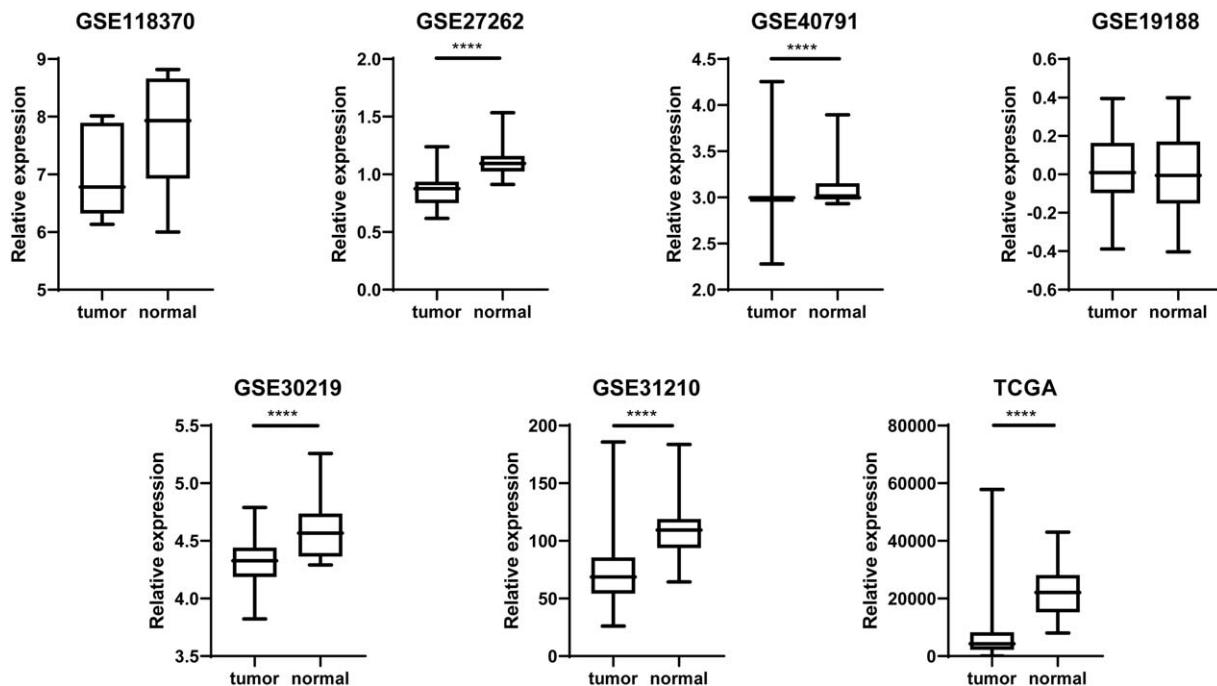
In the present study, we compared the expression of LINC00092 in 20 cases of LUAD tissues and their matched adjacent normal tissues. Total RNA was isolated from the tissues, and quantitative real-time PCR was performed. Relative expression level of LINC00092 were analyzed by using the  $2^{-\Delta\Delta C_t}$  method. Our data showed that the relative expression level of LINC00092 were markedly decreased in LUAD tissues compared to their adjacent normal tissues (Fig. 5A). A line connects 2 points from LUAD to normal with an increased trend, suggesting that the expression level of LINC00092 were significantly lower in LUAD tissues than in the matched normal tissues (Fig. 5B).

### 3.4. Identification of LINC00092 related genes in LUAD

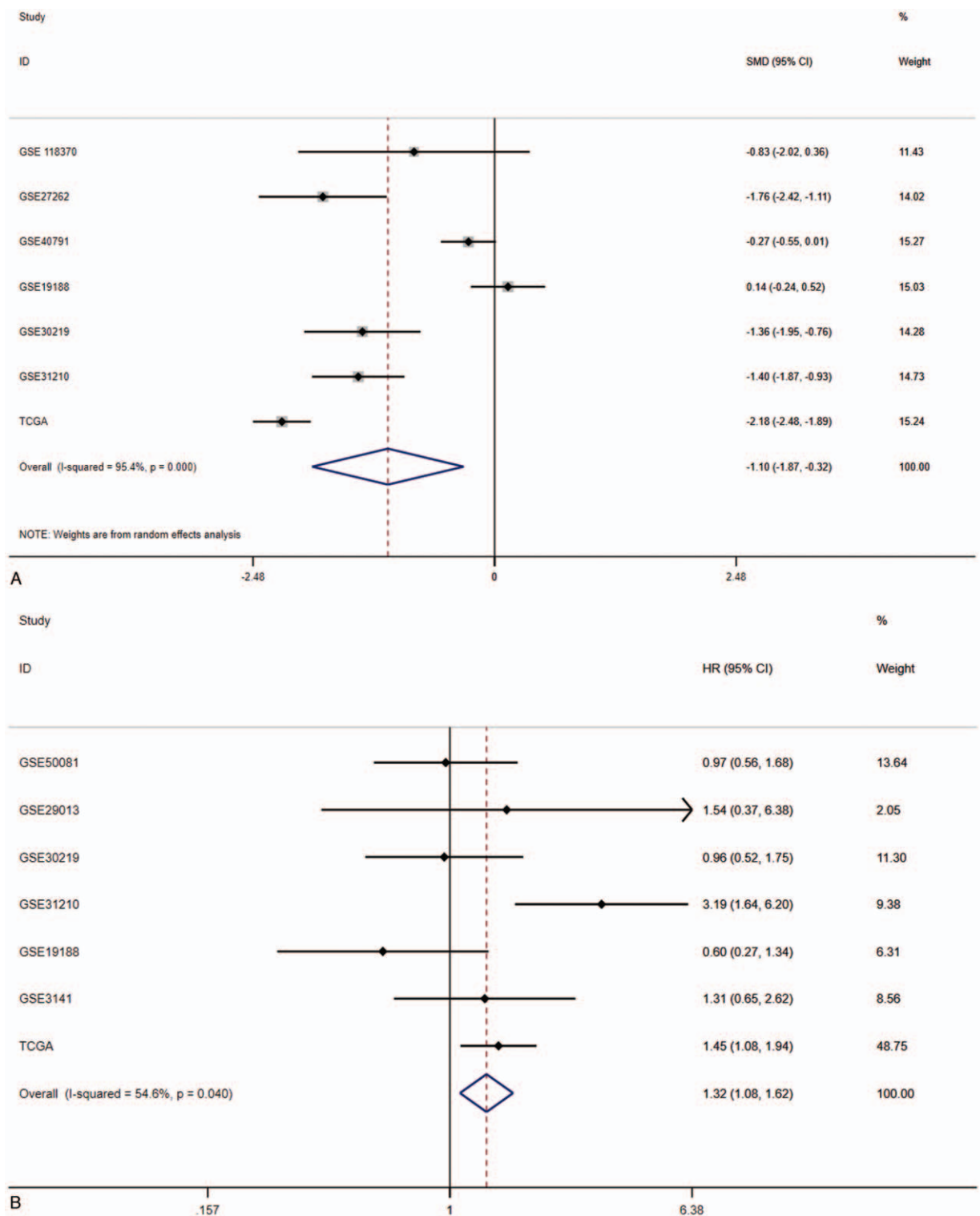
Three public bioinformatic tools were used to identify genes related to LINC00092: MEM, TANRIC, and RBPDB. The number of genes identified from MEM was 7160, when the score was set at 10-4. A total number of 78 genes related to LINC00092 were identified from TANRIC and another 39 genes were recognized by RBPDB. After calculating out genes repeated from any 2 of those 3 sites, 61 genes were eventually used as genes associated with LINC00092 in LUAD.

### 3.5. GO enrichment analysis

DAVID was enrolled for GO analysis. There were 14 significant annotations of the biological process (DIRECT) ( $P < .05$ ), including mRNA processing, positive regulation of phosphatidylinositol 3-kinase activity, B cell receptor signaling pathway, negative regulation of translation, MAPK cascade, and so on. When comes to the GO analysis about CC (DIRECT), 4 significant terms were observed ( $P < .05$ ). They



**Figure 2.** Expression levels of long non-coding RNA 00092 in lung adenocarcinoma tumor tissues and non-tumor tissues were measured in a total of 7 datasets. \*\*\*\* $P < .001$ . TCGA=the cancer genome atlas.



**Figure 3.** Meta-analysis based on the cancer genome atlas and gene expression omnibus datasets. Notes: (A) forest plots to assess the expression of long non-coding RNA 00092 in tumor tissues and non-tumor tissues. (B) Forest plots to assess the effect of downregulated long non-coding RNA 00092 expression on overall survival in patients with lung adenocarcinoma. Weights are from random effects analysis. GEO = gene expression Omnibus, LUAD = Lung adenocarcinoma, OS = overall survival, SMD = standard mean difference, TCGA = the cancer genome atlas.

are cytoplasmic stress granule, extracellular exosome, growth cone, and cell cortex. As to the molecular function (DIRECT) of GO analysis, 7 significant terms were procured ( $P < .05$ ). They were contributed to cytokine activity, RNA and poly (A) RNA

binding, receptor signaling protein tyrosine kinase activity, Ras guanyl-nucleotide exchange factor activity, etc. The outcomes showed in Table 6 and Figure 6. The top 5 results were listed according to  $P$ -value.

**Table 5**  
**Expression of long non-coding RNA 00092 and clinicopathological parameters in lung adenocarcinoma obtained from the cancer genome atlas.**

Clinicopathological Features	N	LINC00092 Relative expression		t	P-value
		Mean	SD		
Tissue				-13.867	<.001
Tumor	533	6566.9673	± 7211.8636		
Normal	59	22631.9661	± 8569.0259		
Age (yr)				-1.259	.209
<60	138	5851.8384	± 6603.9768		
≥60	358	6746.5712	± 7274.2121		
Gender				-1.121	.263
Male	239	6223.8628	± 7441.7324		
Female	276	6943.2861	± 7102.7066		
AJCC pathologic T				2.988	.003
T1	172	8004.4016	± 7789.4820		
T2–T4	343	5909.8924	± 6891.3360		
AJCC pathologic N				1.731	.084
NO NX	343	6997.2460	± 6712.3287		
N1–N3	171	5821.2346	± 8242.6709		
AJCC pathologic M				-0.187	.852
M0 MX	486	6583.2281	± 7307.5118		
M1	25	6861.5708	± 6412.0837		
Pathologic stage				2.022	.044
I–II	397	6949.9821	± 7346.8671		
III–IV	110	5365.8887	± 6987.8984		

AJCC=American Joint Committee on Cancer, LUAD=lung adenocarcinoma, SD=standard deviation, TCGA=the cancer genome atlas.

### 3.6. KEGG pathway analysis

KOBAS was responsible for the KEGG pathway analysis, there were 30 signaling pathways that were statistically significant ( $P < .05$ ). The top 20 items were listed in Figure 7 according to corrected  $P$ -value. The signaling pathways were mainly enriched in the NF-kappa B signaling pathway, HIF-1 signaling pathway and ErbB signaling pathway, etc.

### 3.7. PPI interaction of the 61 genes

The protein-to-protein network analysis was conducted by STRING website (Fig. 8). In the networks, the nodes represent for proteins, and each color corresponds to a cluster. The line thickness is an indicator for the strength of the evidence. The edges indicate the predicted functional associations, which are colored by the types of predicted associations using different evidence. The red lines, green lines, blue lines, purple lines, yellow lines, light blue lines, and black lines indicate fusion evidence, neighborhood evidence, co-occurrence evidence, experimental evidence, text mining evidence, database evidence, and co-expression evidence, respectively. As showed in this PPI network, several genes demonstrated higher connectivity, including SYK, ERBB4, NCL, IGF2BP1, PRKCB, WDR12, and CD40LG, etc.

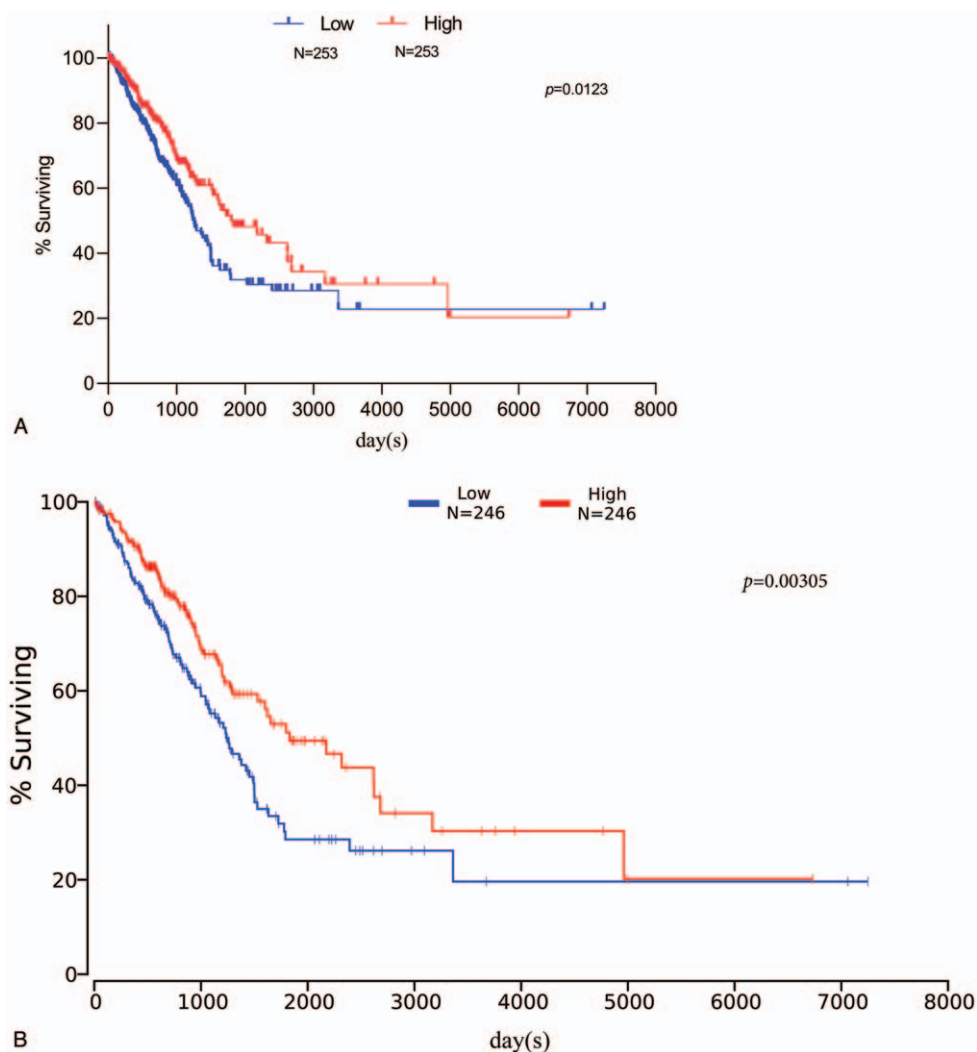
## 4. Discussion

The development of LUAD involves molecular mechanisms, and many studies have shown that lncRNA plays a key role in this process.<sup>[19–21]</sup> Therefore, exploration of tumor-associated lncRNAs and its related signaling pathways is of great significance, which can provide a sufficient theoretical basis for elucidating the potential mechanisms of LUAD patients and finding effective therapeutic targets. The TCGA and GEO

databases provided a large amount of open raw data from multiple experiments on multiple platforms, making big data analysis possible.

In this paper, we attempted to investigate the potential roles of LINC00092 in LUAD by analyzing the data excavated from TCGA and GEO databases, and we conformed the expression level of LINC00092 in in-house real tumor samples. Our results indicated that the expression of LINC00092 was significantly decreased in LUAD tissues compared with non-tumor tissues. Then the HR of meta-analysis established that low LINC00092 expression was a high HR indicator. Thereafter, analysis of TCGA data showed that the decline in LINC00092 expression was not related to gender and age, but was associated with pathologic T and pathological stage. As to the pathologic M, an indicator for remote metastasis, statistical analysis showed that there was no difference between 2 groups. The reason for this phenomenon might contributed to “MX”. “MX” means that the remote metastasis status cannot be assessed. Among the 511 eligible “pathologic M” data, there were 140 cases of “MX”, accounting for more than 27.3%. Simply categorizing “MX” as “M0” or “M1” might result in false positive or false negative results. Therefore, it has not been clarified whether the expression of LINC00092 is related to remote metastasis. We also found that patients with lower expression of LINC00092 exhibited poorer OS compared to patients with higher expression of LINC00092. Taken together, our findings suggested that the absence of LINC00092 played a crucial role in tumorigenesis and prognosis in patients with LUAD.

At present, there is still no related research on the molecular mechanism of LINC0092 in LUAD patients. So, the second step of this study is to explore the potential mechanism about LINC00092 in LUAD. Here, we enrolled 3 biological databases: MEM, TANRIC, and RBPDB. Those 3 biological tools belong to

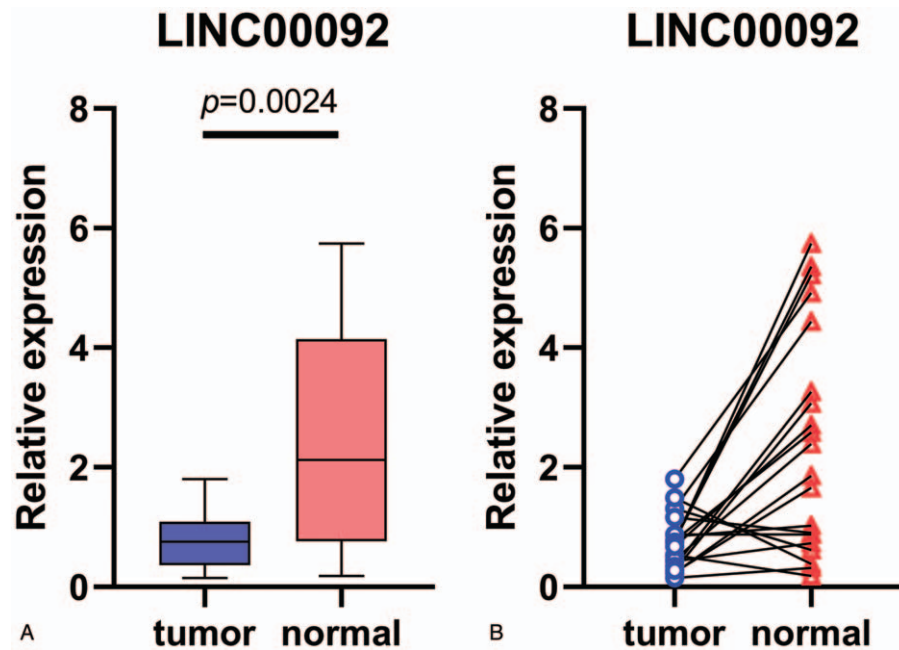


**Figure 4.** Kaplan–Meier survival curves for patients with high and low long non-coding RNA 00092 expression. Notes: (A) survival curve based on the cancer genome atlas data and plotted by GraphPad Prism 8. (B) Survival curve came from the OncoLnc website (<http://www.oncolnc.org/>).

3 different kinds of category. MEM recognizes lncRNA-related genes based on publicly available gene expression data sets from ArrayExpress database. TANRIC characterizes the expression profiles of lncRNAs in large patient cohorts of 20 cancer types, while the RBPDB database predicts genes associated with lncRNA based on the binding properties of the protein. We used these tools separately and conjointly to analyze the outcomes. Hence, 61 LINC00092 related genes were harvested. The next, we performed GO and KEGG pathway analysis to identify biological functions enriched among those LINC00092-related genes. The results of GO analysis demonstrate that LINC00092 is closely related to cellular RNA related processing and phosphorylating activities. In addition, the CC of LINC00092 related genes is mainly associated with cytoplasmic stress granule and exosome.

KEGG pathway analysis found that multiple signaling pathways may be regulated by LINC00092. Among these signaling pathway, NF- $\kappa$ B, HIF-1 and ErbB signaling pathways were significant. These signaling pathways have been proved to be clearly involved in the process of tumorigenesis. NF- $\kappa$ B signaling

pathway is activated in over 60% of lung cancers. Aberrant activation of NF- $\kappa$ B signaling in LUAD cell lines and tumor tissues could be involved in tumorigenesis and metastasis.<sup>[22–24]</sup> In addition to regulating cell proliferation, angiogenesis, and apoptosis, NF-kappaB is also involved in mechanism of epithelial-mesenchymal transition,<sup>[25]</sup> which is critical for cancer cell invasion and metastasis. Activation of HIF-1 signaling is considered as tumor-initiating factor. It may involve in tumor angiogenesis, energy metabolism, cell proliferation, invasion, and metastasis.<sup>[26,27]</sup> HIF-1alpha antagonist can inhibit progression and spread of LUAD.<sup>[28]</sup> The ErbB receptor tyrosine kinase family contains 4 cell surface receptors: ErbB1/EGFR/HER1, ErbB2/HER2, ErbB3/HER3, and ErbB4/HER4.<sup>[29]</sup> Besides its involvement in tumorigenesis and metastasis through Akt, MAPK and many other pathways, ErbB also appears as a kinase and transcriptional regulatory molecule in the nucleus. In many forms of malignancy, members of the ErbB family and some of their ligands are often overexpressed, amplified, or mutated.<sup>[30]</sup> Therefore, we suppose that LINC092 inhibits tumorigenesis and progression by modulating these pathways in LUAD.



**Figure 5.** Relative expression level of long non-coding RNA 00092 in clinical samples. Notes: (A) box plot to uncover the relative expression level of long non-coding RNA 00092 between lung adenocarcinoma and normal tissues. (B) Symbols & lines plot suggesting that the relative expression level of long non-coding RNA 00092 was lower in lung adenocarcinoma tissues than in the matched normal tissues. LUAD=Lung adenocarcinoma.

Finally, the following limitations should be considered in our analysis:

- (1) the inclusive data and clinical samples were limited. We believe that as the data grows and becomes more sophisticated, there may be more useful results;
- (2) This study was based on bioinformatics analysis, and the genes involved were predicted. Therefore, we still need further experiments to evaluate and verify the contribution of LINC00092 in LUAD.

## 5. Conclusion

In conclusion, based on the data we obtained from TCGA and GEO datasets and the information we harvested from real tumor samples, we conclude that the expression of LINC00092 might play a pivotal role in the initiation and progression of LUAD. LINC00092 was detected at low expression levels in LUAD tissues and was linked to high pathologic stage. Moreover, the lower expressions of LINC00092 in LUAD indicated poorer OS. The possible signaling pathways for LINC00092 contributing in

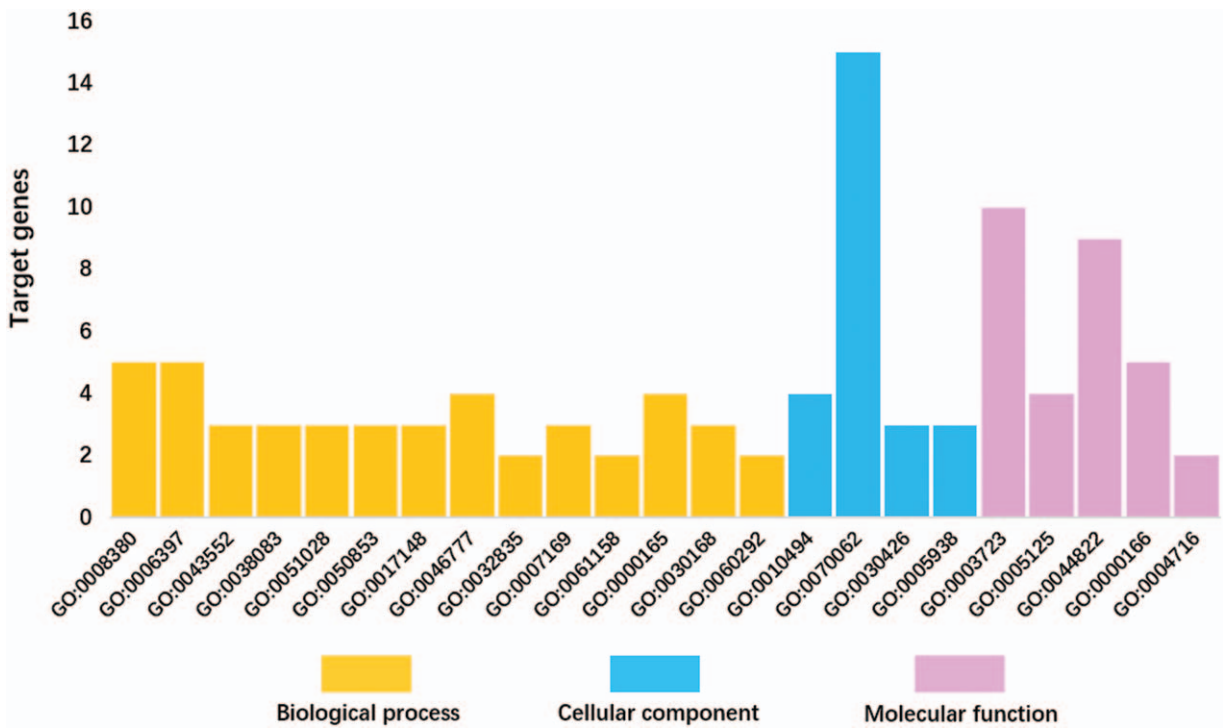
**Table 6**

### The gene ontology analysis of predicted target genes of long non-coding RNA 00092.

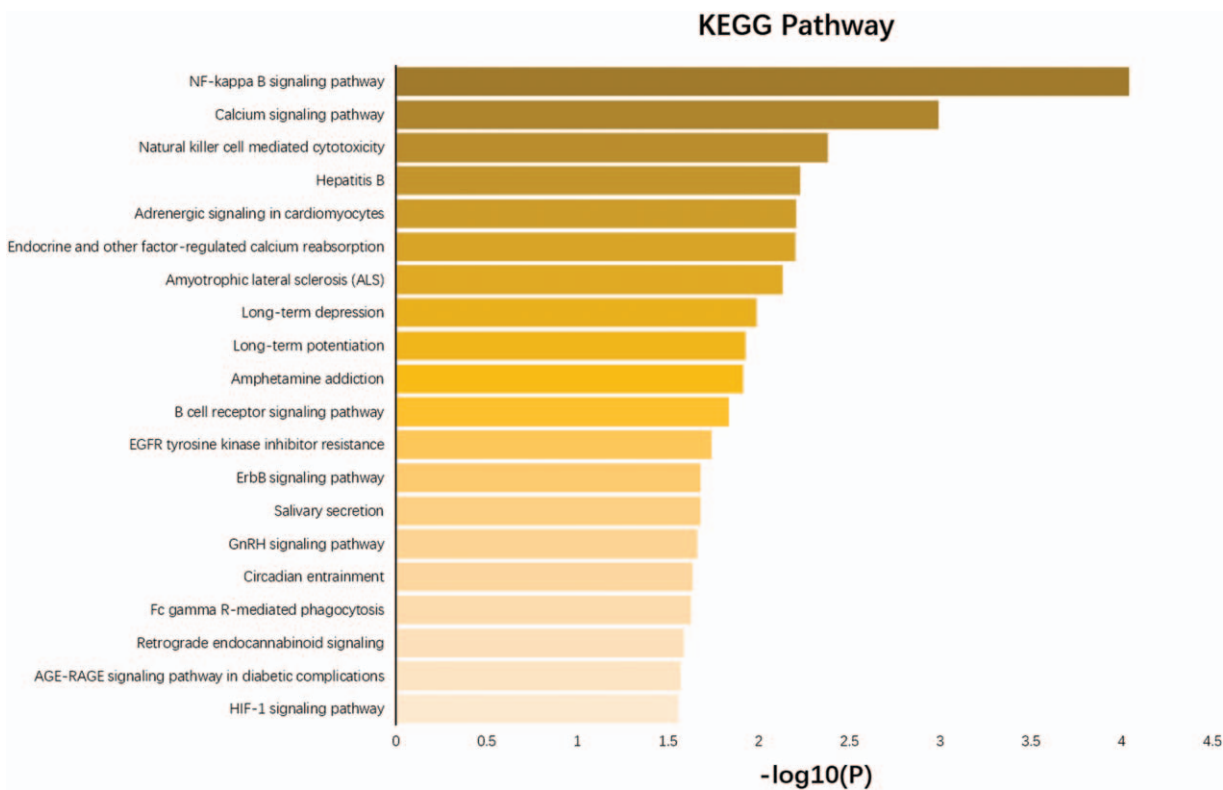
GO ID	Term	Count	P value
Biological process			
GO:0008380	RNA splicing	5	1.49 X10 <sup>-3</sup>
GO:0006397	mRNA processing	5	1.96 X10 <sup>-3</sup>
GO:0043552	Positive regulation of phosphatidylinositol 3-kinase activity	3	3.82 X10 <sup>-3</sup>
GO:0038083	Peptidyl-tyrosine autophosphorylation	3	6.31 X10 <sup>-3</sup>
GO:0051028	mRNA transport	3	8.62 X10 <sup>-3</sup>
Cellular component			
GO:0010494	Cytoplasmic stress Granule	4	1.46E X10 <sup>-4</sup>
GO:0070062	Extracellular exosome	15	2.31E X10 <sup>-2</sup>
GO:0030426	Growth cone	3	4.33E X10 <sup>-2</sup>
GO:0005938	Cell cortex	3	4.81E X10 <sup>-2</sup>
Molecular function			
GO:0003723	RNA binding	10	2.82 X10 <sup>-5</sup>
GO:0005125	cytokine activity	4	1.53 X10 <sup>-2</sup>
GO:0044822	Poly(A) RNA binding	9	1.71 X10 <sup>-2</sup>
GO:0000166	Nucleotide binding	5	1.94 X10 <sup>-2</sup>
GO:0004716	Receptor signaling protein tyrosine kinase activity	2	2.92 X10 <sup>-2</sup>

In this table, the top 5 terms of the Biological process analysis and Molecular function analysis are listed. Cellular component just has 4 results. ( $P < .05$ ). GO= gene ontology.

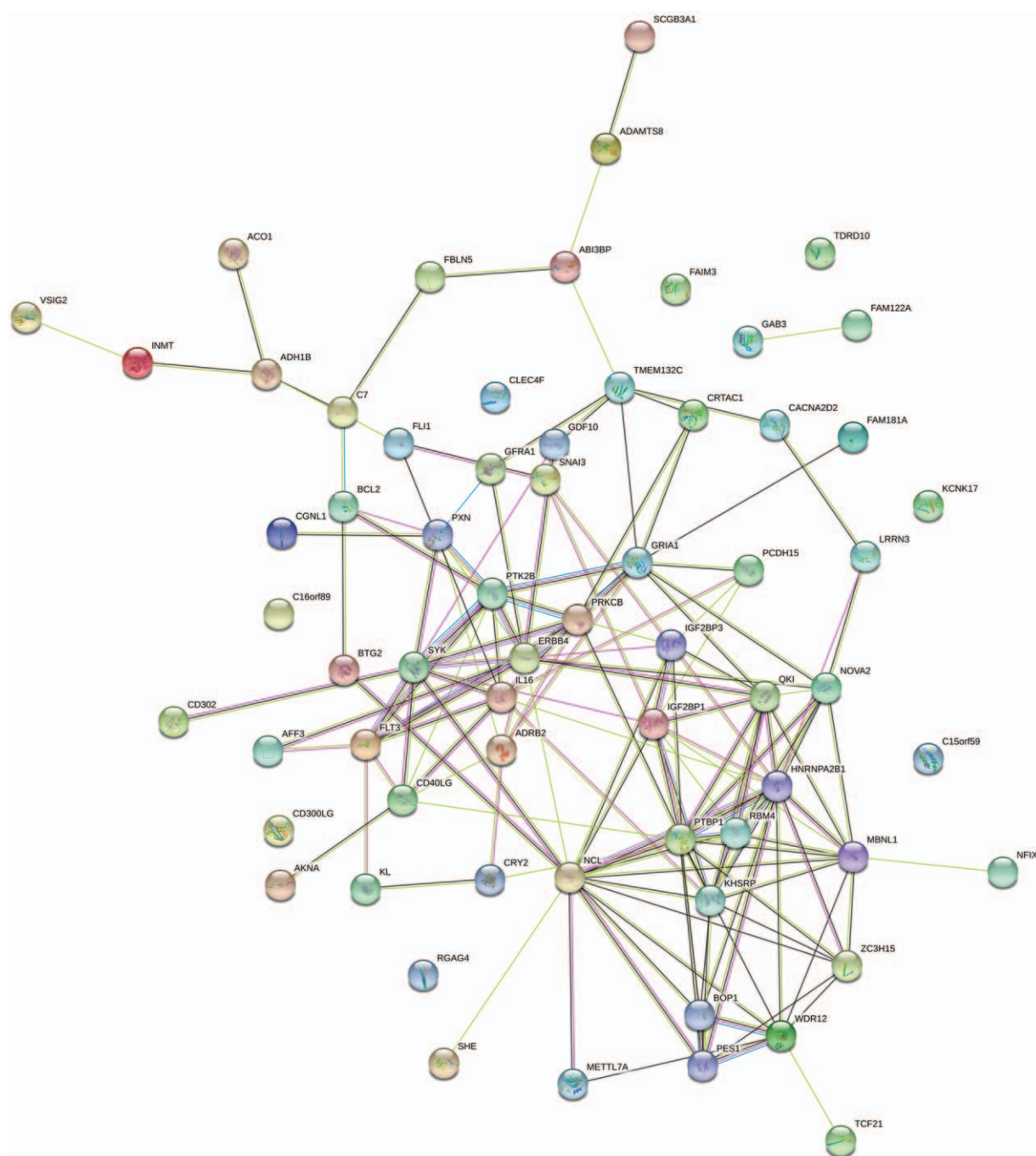




**Figure 6.** GO enrichment analysis of long non-coding RNA 00092 related genes in lung adenocarcinoma. Note: long non-coding RNA 00092 related genes were divided into 3 functional groups: biological process, cellular component, and molecular function. GO= gene ontology, LUAD=lung adenocarcinoma.



**Figure 7.** Pathway analysis of the predicted target genes of long non-coding RNA 00092. Note: Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis. Top 20 pathways were listed in this graph ( $P < .05$ ).



**Figure 8.** Protein-protein interactions of the long non-coding RNA 00092 related genes.

LUAD were NF- $\kappa$ B, HIF-1 and ErbB signaling pathways. Overall, this study might provide a potential approach to molecular diagnosis, evaluation of prognosis, and targeted therapy of LUAD in the future.

### Author contributions

Xin Wang designed the study and revised the article. Xianwei Wang collected the clinical samples and completed the experiments of clinical samples. Xin Wang reviewed the literature and

wrote the article. Guichuan Huang reviewed the literature and revised the article. All authors read and approved the final manuscript.

**Conceptualization:** Xin Wang.

**Data curation:** Xin Wang.

**Formal analysis:** Xin Wang.

**Investigation:** Xianwei Wang.

**Methodology:** Xianwei Wang, Guichuan Huang.

**Software:** Yi Chen.

**Writing – original draft:** Xin Wang.

## References

- [1] Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;136:E359–86.
- [2] Siegel RL, Miller KD, Jemal A. Cancer statistics. *CA Cancer J Clin* 2015;65:5–29.
- [3] Hirsch FR, Scagliotti GV, Mulshine JL, et al. Lung cancer: current therapies and new targeted treatments. *Lancet (London, England)* 2017;389:299–311.
- [4] Chen Z, Fillmore CM, Hammerman PS, et al. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat Rev Cancer* 2014;14:535–46.
- [5] Luo S, Lu JY, Liu L, et al. Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. *Cell Stem Cell* 2016;18:637–52.
- [6] Iyer MK, Niknafs YS, Malik R, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* 2015;47:199–208.
- [7] Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. *Cell* 2013;152:1298–307.
- [8] Tao H, Yang JJ, Zhou X, et al. Emerging role of long noncoding RNAs in lung cancer: current status and future prospects. *Respir Med* 2016;110:12–9.
- [9] Castillo J, Stueve TR, Marconett CN. Intersecting transcriptomic profiling technologies and long non-coding RNA function in. *Oncotarget* 2017;8:81538–57.
- [10] Lin C, Yang L. Long noncoding RNA in cancer: wiring signaling circuitry. *Trends Cell Biol* 2018;28:287–301.
- [11] Inamura K. Major tumor suppressor and oncogenic non-coding RNAs: clinical relevance in lung cancer. *Cells* 2017;6:12–26.
- [12] Anaya JJPCS. OncoLnc: Linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. *Peer J Computer Science* 2016;2:e67–79.
- [13] Adler P, Kolde R, Kull M, et al. Mining for coexpression across hundreds of datasets using novel rank aggregation and visualization methods. *Genome Biol* 2009;10:R139–149.
- [14] Li J, Han L, Roebuck P, et al. TANRIC: an interactive open platform to explore the function of lncRNAs in cancer. *Cancer Res* 2015;75:3728–37.
- [15] Cook KB, Kazan H, Zuberi K, et al. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res* 2011;39:D301–8.
- [16] Dennis G, Sherman BT, Hosack DA, et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* 2003;4:P3–13.
- [17] Wu J, Mao X, Cai T, et al. KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Res* 2006;34:W720–4.
- [18] Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2015;43:D447–52.
- [19] Dong S, Qu X, Li W, et al. The long non-coding RNA, GASS, enhances gefitinib-induced cell death in innate EGFR tyrosine kinase inhibitor-resistant lung adenocarcinoma cells with wide-type EGFR via down-regulation of the IGF-1R expression. *J Hematol Oncol* 2015;8:43–55.
- [20] Peng Z, Wang J, Shan B, et al. The long noncoding RNA LINC00312 induces lung adenocarcinoma migration and vasculogenic mimicry through directly binding YBX1. *Mol Cancer* 2018;17:167–72.
- [21] Wang C, Gu Y, Zhang E, et al. A cancer-testis non-coding RNA LIN28B-AS1 activates driver gene LIN28B by interacting with IGF2BP1 in lung adenocarcinoma. *Oncogene* 2019;38:1611–24.
- [22] De Simone V, Franzè E, Ronchetti G, et al. Th17-type cytokines, IL-6 and TNF- $\alpha$  synergistically activate STAT3 and NF- $\kappa$ B to promote colorectal cancer cell growth. *Oncogene* 2015;34:3493–503.
- [23] Karin M, Cao Y, Greten FR, et al. NF- $\kappa$ B in cancer: from innocent bystander to major culprit. *Nat Rev Cancer* 2002;2:301–10.
- [24] Schmid JA, Birbach A. IKK $\beta$  kinase beta (IKK $\beta$ /IKK2/IKK $\beta$ )—a key molecule in signaling to the transcription factor NF- $\kappa$ B. *Cytokine Growth Factor Rev* 2008;19:157–65.
- [25] Huber MA, Beug H, Wirth T. Epithelial-mesenchymal transition: NF- $\kappa$ B takes center stage. *Cell Cycle (Georgetown, Tex)* 2004;3:1477–80.
- [26] Chang CC, Lin MT, Lin BR, et al. Effect of connective tissue growth factor on hypoxia-inducible factor 1 $\alpha$  degradation and tumor angiogenesis. *J Natl Cancer Inst* 2006;98:984–95.
- [27] Galanis A, Pappa A, Giannakakis A, et al. Reactive oxygen species and HIF-1 signalling in cancer. *Cancer Lett* 2008;266:12–20.
- [28] Jacoby JJ, Erez B, Korshunova MV, et al. Treatment with HIF-1 $\alpha$  antagonist PX-478 inhibits progression and spread of orthotopic human small cell lung cancer and lung adenocarcinoma in mice. *J Thorac Oncol* 2010;5:940–9.
- [29] Avraham R, Yarden Y. Feedback regulation of EGFR signalling: decision making by early and delayed loops. *Nat Rev Mol Cell Biol* 2011;12:104–17.
- [30] Yarden Y, Pines G. The ERBB network: at last, cancer therapy meets systems biology. *Nat Rev Cancer* 2012;12:553–63.