

A signature related to lung adenocarcinoma prognosis: A study based on TCGA database

Jia Wang

Shandong University of Traditional Chinese Medicine

Xiaolu Zhang

Shandong University of Traditional Chinese Medicine

Xiaoming Zhang

Shandong University of Traditional Chinese Medicine

Yan Yao

Weifang Medical University

Xiaoran Ma

Shandong University of Traditional Chinese Medicine

Xiaowei Xu

Shandong University of Traditional Chinese Medicine

Jing Zhuang

Weifang Traditional Chinese Hospital

Lijuan Liu

Weifang Traditional Chinese Hospital

Peng Sun

Shandong University of Traditional Chinese Medicine

Jibiao Wu

Shandong University of Traditional Chinese Medicine

Changgang Sun (✉ scgdoctor@126.com)

Weifang Traditional Chinese Hospital <https://orcid.org/0000-0002-6648-3602>

Research article

Keywords: Lung adenocarcinoma; CeRNA network; Signature; Prognosis; Overall survival

Posted Date: January 22nd, 2020

DOI: <https://doi.org/10.21203/rs.2.21607/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The intrinsic molecular subtypes of lung adenocarcinoma (LUAD) impact clinical treatment decision-making, but the molecular mechanisms are still unclear. Therefore, we aimed to identify sensitive biomarkers to evaluate LUAD patient prognosis. **Methods:** Differentially expressed RNAs from LUAD patients were obtained from The Cancer Genome Atlas (TCGA) database and they were used to construct a competitive endogenous RNA (ceRNA) network. Based on the examination of clinical data, long noncoding RNAs (lncRNAs) and mRNAs in the network were selected by univariate and multivariate Cox regression analysis. Finally, functional enrichment analysis was used to reveal prognostic signatures based on the classification into high and low-risk groups, survival analysis, and an independence test. **Results:** The ceRNA network consisted of 21 mRNAs, 53 lncRNAs, and 8 miRNAs that were selected from the differentially expressed RNAs identified. Next, based on univariate and multivariate Cox regression analysis, a prognostic signature, including two mRNAs (HOXA10 and CBX2) and four lncRNAs (LINC00460, LINC00330, DGCR5, and C14orf132) was constructed. Eventually, survival analysis showed that significant differences in survival rates between high and low-risk groups and the area under the curve (AUC) for three-year survival was 0.714. Compared with clinical risk factors, including age, pathological stage, and TNM stage, our risk score had a higher prognostic value. **Conclusion:** By screening from a ceRNA network, we constructed a signature, including two mRNAs (HOXA10 and CBX2) and four lncRNAs (LINC00460, LINC00330, DGCR5, and C14orf132), that can be utilized as a prognostic biomarker in LUAD. This signature may provide options for clinical treatment.

1. Introduction

There are dozens of molecular pathways and mechanisms that have been proven to affect the occurrence and development of tumors[1]. Gene expression patterns identified in LUAD cover different functional pathways and exhibit variable outcomes for patients[2], which effect treatment choices[3]. The American Joint Committee on Cancer (AJCC) staging system is at the forefront of guiding clinical decision making and is, by far, the best way to predict LUAD patient prognosis. However, more than 20% of early-treatment patients ultimately develop cancer recurrence and metastasis[4]. Therefore, it is necessary to identify the molecular pathways involved in LUAD growth in order to design novel diagnostic strategies and targeted therapies.

Gene expression profiles play a unique role as predictors for prognosis and treatment[5]. The identification of risk-related biomarkers for LUAD will also provide an opportunity to understand cancer development and progression. Identification of sensitive biomarkers to evaluate LUAD prognosis at an early stage is essential for successful disease treatment.

Long non-coding RNAs (lncRNAs) and microRNAs (miRNAs), RNA subtypes without protein-coding functions, are widely expressed in the tissue of organisms[6]. In recent years, lncRNAs have been a hotspot for research because of their vital role in chromatin organization, transcription, RNA processing, and nuclear domains[7–11]. A growing number of studies have demonstrated that aberrant lncRNA

expression is related to prognosis or response to therapy in various types of cancer. For example, Wang et al. found that lncRNA miR503HG inhibits tumor metastasis as a prognostic indicator in liver cancer[12] and Xie et al. revealed that circulating lncRNAs are valuable, novel biomarkers for the diagnosis and prognosis of non-small cell lung cancer[13]. Consequently, there may be additional lncRNA prospects as biomarkers or therapeutic targets due to their genome-wide expression pattern in various tissues as well as their tissue-specific expression characteristics.

In 2011, Pandolfi proposed the competing endogenous RNA (ceRNA) hypothesis which suggests that lncRNAs and mRNAs compete for miRNA binding as a regulatory mechanism. Xia W. et al. revealed that the TWIST1-centered ceRNA network facilitates the development process of LUAD[14]. Through this mechanism, miRNAs bind to target mRNA molecules using miRNA response elements (MREs), which reduce the stability of target mRNAs and inhibit translation at the transcriptional level and, conversely miRNAs can be affected. Various types of RNA transcripts, including mRNAs, lncRNAs, and circular RNAs (circRNAs), compete for miRNA binding by mutating MREs as a means of mutual regulation. This confers new and broader biological functions for both mRNA and lncRNA. In addition, a previous study showed that a predictive signature combining the tumor immune microenvironment with chemosensitivity-related features can improve the prognosis and therapeutic outcome for patients with stage II to III gastric cancer[15]. Based on the RNA network and signature, we may find novel prognostic biomarkers or therapeutic targets for LUAD.

Taken together, we need to investigate the trends associated with the development of LUAD at the molecular level. In this study, we used the transcriptome sequencing data from LUAD patients in TCGA to identify differentially expressed genes that could be used to build related transcriptional regulatory networks, constructed a prognostic signature, and comprehensively analyzed the prognostic value of indicators.

2. Material And Methods

2.1 Data acquisition and differential expression analysis

The gene expression profiles were obtained from the Genomic Data Commons (GDC) database (<http://gdc-portal.nci.nih.gov/>), which contains all TCGA data for analysis. The data included RNA-seq data from 539 LUAD patient samples and 59 tumor-adjacent normal tissues, as well as all clinical information from these cases. Also, 521 tumor samples and 46 healthy samples from miRNA-seq data were downloaded. To explore possible lncRNA-mRNA-miRNA interactions, RNA-seq analysis was utilized. The edgeR package in R was used to identify differentially expressed genes or mRNAs (DEGs), lncRNAs (DELs), and miRNAs (DEMs) from the raw sequencing data in TCGA. Referencing previous literature, adjusted p values were used in the false discovery rate (FDR) to ensure more accurate results. The thresholds were set as $FDR < 0.01$ and $|\log_{2}FC \text{ (fold change)}| > 2$.

2.2 CeRNA network building

Based on the use of mircode (<http://www.mircode.org/index.php>), we predicted the interaction between DELs and DEMs for building a transcriptional regulatory network. Next, three online tools, including miRDB (<http://www.mirdb.org>), miRTarBase (<http://mirtarbase.mbc.nctu.edu.tw/php/index.php>), and TargetScan (http://www.targetscan.org/vert_72/), were used for predicting the target genes of significant DEMs. To obtain more accurate results, we constructed a Venn diagram to identify a common set from the three target gene sets as well as the intersection of the above results and 2296 DEGs. To build the lncRNA-mRNA-miRNA interaction network, the expression of lncRNAs and mRNAs was considered to have a negative correlation when the miRNAs were expressed. Cytoscape (version 3.7.0) was used to visualize the results and obtain the ceRNA network graph. The CentiScaPe2.2 APP was used to select central genes from the ceRNA network (degree > 16).

2.3 Prognosis-related RNA selection and prognostic signature construction

After excluding samples containing two or more tumor tissues and incomplete clinical data, we performed univariate Cox regression analysis of lncRNAs and mRNAs in the ceRNA network to screen for prognostic genes, with $p < 0.05$ considered significant. Then, multivariate Cox regression analysis was conducted to establish a prognostic signature for LUAD. The signature was defined as a collection of highly prognostic and connected genes within the network.

2.4 Functional enrichment analysis

DAVID (<https://david.ncifcr.gov/>), the database for annotation, visualization, and integrated discovery, is based on all genes in the human genome database and provides systematic and comprehensive annotation information of biological functions. We used DAVID to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses of mRNAs in the ceRNA network, with $p < 0.05$ considered significant. Visualization of the enrichment analysis results was performed using ggplot R software.

2.5 High and low-risk group classification and survival analysis

A risk score, which could better determine the performance of the lncRNAs and mRNAs in the signature, was calculated. It is the product of the Cox regression coefficient for each gene multiplied by the gene expression value. The calculation formula is as follows:

$$\text{Risk score} = \sum \beta \times \text{expression value}_{\text{gene}}$$

Here β refers to the regression coefficient of multivariate Cox analysis; $\text{Expression value}_{\text{gene}}$ is the gene expression value level of core lncRNAs and mRNAs.

The risk score median was regarded as the threshold. We divided the sample into the high-risk and low-risk groups. Survival curves, distribution of risk scores, and patient survival status were determined with

the R software package. By receiver-operating characteristic (ROC) curve analysis, the signature's performance was assessed, optimizing the critical expression threshold (condition: $AUC > 0.7$ and $p < 0.05$) and exploring the sensitivity of prognostic signatures for LUAD patient three-year survival rates. Furthermore, it was essential to use Kaplan–Meier plots (<http://kmplot.com/analysis/>) to explain LUAD patient overall survival in relationship to the prognostic signature gene expression, with $p < 0.05$ considered statistically significant.

2.6 Independent prognostic analysis and forestplot construction

For verifying the independence of the predictive signature, the prognostic values of clinical risk factors (age, pathological stage, TNM stage and risk score) were evaluated using the survival package in R and forestplots were constructed using R.

3. Results

3.1 DEGs, DELs, and DEMs

According to RNA-seq data from LUAD patients in the TCGA database, we found a total of 1085 DELs (883 upregulated and 202 downregulated) and 2296 DEGs (1781 upregulated and 515 downregulated) that were significant ($p < 0.01$, $|\logFC| > 2.0$, Fig. 1a, 1b). Meanwhile, 95 DEMs were also identified, consisting of 81 that were upregulated and 14 that were downregulated ($p < 0.01$ and $|\logFC| > 2.0$, Fig. 1c).

3.2 CeRNA network

To investigate possible lncRNA-mRNA-miRNA interactions in LUAD patients, we successfully predicted 92 DELs and 19 DEMs with interactions by MiCode. In addition, 33 target genes were later found and the results predicted matched with the 2296 DEGs. Based on the 92 DEL-DEM pairs and 22 DEM-DEG pairs, Cytoscape was used to build a visual network (Fig. 2), which involved 53 DELs (43 upregulated and 10 downregulated), 21 DEGs (15 upregulated and 6 downregulated), and 8 DEMs (5 upregulated and 3 downregulated) in Table 1. Within the network, the node degree indicates the number of links with other nodes and nodes with a higher degree usually play a pivotal role in the system being modeled by the network. Three miRNAs were identified, namely hsa-mir-144, hsa-mir-195, and hsa-mir-143 with degrees greater than 16.

Table 1
LncRNAs, miRNAs and mRNAs in the CeRNA network

	ID				Count
LncRNA	AGAP11 AGBL1 AC105206.1 C14orf132 LINC00472 LINC00470 C20orf197 AP002478.1 AC022148.1 MEG3 LINC00466 LINC00460 DSCAM - AS1 OR51B5	MIR137HG DLX6 - AS1 ERVH48 - 1 FNDC1 - IT1 DGCR5 FOXP1 - IT1 LINC00461 GRM5 - AS1 C2orf48 AC087269.1 AC020907.1 LINC00519 LINC00524	LINC00319 MUC2 C1orf220 ST7 - OT4 AL117190.1 STEAP2 - AS1 LINC00355 KIF25 - AS1 LINC00160 ABCA9 - AS1 LINC00330 NAV2 - AS2 RMST	EGOT AC112721.1 HNF1A - AS1 PVT1 C5orf64 LINC00518 LINC00221 AL391152.1 UCA1 CLDN10 - AS1 HOTAIR AC004832.1 ADAMTS9 - AS1	53
miRNA	hsa - mir - 372 hsa - mir - 144	hsa - mir - 195 hsa - mir - 31	hsa - mir - 200a hsa - mir - 143	hsa - mir - 182 hsa - mir - 96	8
mRNA	TMEM100 HOXA10 KCNQ5 TBX18 SALL1 PROK2	CEP55 E2F7 CHEK1 PSAT1 CLSPN	CDC25A CBX2 CCNE1 KIF23 DLC1	COL5A2 COL1A1 BDNF SELE SLC1A1	21

3.3 Prognosisrelated signature

In the above network, we discovered two DEGs and four DELs significantly related to LUAD prognosis using univariate Cox regression analysis ($p < 0.05$). Hereafter, multivariate Cox regression analysis was conducted on the two mRNAs (HOXA10 and CBX2) and the four lncRNAs (LINC00460, LINC00330, DGCR5, and C14orf132), identifying this combination as the optimal prognosisrelated signature (Table 2).

Table 2

Significant genes in univariate and multivariate Cox regression analysis.

Univariate cox regression analysis					
Gene	HR		z	pvalue	
HOXA10	1.070917205		2.013969099	0.044012778	
CBX2	0.856870437		-2.524037321	0.011601561	
LINC00460	0.918863897		-2.322112485	0.02022688	
LINC00330	0.878153532		-2.038758824	0.041474104	
DGCR5	0.906614988		-1.971151349	0.048706567	
C14orf132	0.842198535		-2.231767091	0.02563036	
Multivariate cox regression analysis					
Gene	coef	exp(coef)	se(coef)	z	p
HOXA10	0.0746	1.07746	0.03486	2.14	0.03235
CBX2	-0.1906	0.82647	0.05801	-3.285	0.00102
LINC00460	-0.08709	0.91659	0.0375	-2.323	0.0202
LINC00330	-0.09837	0.90631	0.06395	-1.538	0.12398
DGCR5	-0.08088	0.9223	0.05279	-1.532	0.12547
C14orf132	-0.15842	0.85349	0.07509	-2.11	0.03488

3.4 Functional enrichment analysis

It is important to understand that the functions of lncRNAs relate to their target genes rather than their protein-encoding capabilities. The functional enrichment analysis resulted in genes that were involved in protein digestion and absorption as well as the cell cycle ($p < 0.05$ and gene count > 2). Three GO terms were identified out of the 19 terms (Fig. 3), including biological process (BP), cellular component (CC), and the molecular function (MF), which were associated with protein binding (MF), nucleoplasm (CC), and DNA replication (BP), respectively.

3.5 High-low risk groups and survival analysis

For a better evaluation of the signature's ability to predict LUAD patient prognosis, a risk score was built, where the coefficients for the samples were weighted. The risk score was calculated as follows (Fig. 4a):

Risk score = $(0.07460 \times \text{expression level of HOXA10}) + (-0.19060 \times \text{expression level of CBX2}) + (-0.08709 \times \text{expression level of LINC00460}) + (-0.09837 \times \text{expression level of LINC00330}) + (-0.08088 \times \text{expression level of DGCR5}) + (-0.15842 \times \text{expression level of C14orf132})$

level of DGCR5) + (-0.15842 × expression level of C14orf132). With the median risk score set as the threshold, a total of 480 patients meeting the standard was equally divided into the high-risk and low-risk groups. The survival status of each group is shown in Fig. 4b. Meanwhile, Survival analysis showed that there was a significant difference in the five-year overall survival between the high-risk group and the low-risk group ($p = 5e-05$, Fig. 4c). Over time, the survival rate of the two groups gradually decreased, yet the survival time of the low-risk group was significantly higher than that of the high-risk group. The AUC suggested superior prognostic value since the ROC curve indicated that the AUC of the lncRNA-mRNA signature for predicting three-year survival was 0.714 (Fig. 4d). We assessed the link between gene expression in the prognostic signature and LUAD patient survival by means of a Kaplan-Meier curve and log-rank test, where $p < 0.05$ was considered significant. Subsequently, two mRNAs (HOXA10 and CBX2) and two lncRNAs (DGCR5 and C14orf132) were selected. In the ceRNA network of LUAD patients, we determined that upregulation of CBX2, DGCR5, HOXA10 (HR: 1.55%, 1.16%, and 1.49%, $p < 0.05$) and downregulation of C14orf132 (HR:0.7%, $p < 0.05$) (Fig. 5) were significantly correlated with survival time and risk-group determination. A high level of C14orf132 or low levels of CBX2, DGCR5, HOXA10 were associated with higher patient survival rates.

3.6 Independence test

We evaluated the prognostic value of clinical risk factors (age, pathological stage, and TNM stages) and risk score with the survival package in R. The results showed that risk score was the superior factor affecting the survival rate (Fig. 6).

4. Discussion

During the last few years, along with the emerging role of high-throughput sequencing technology, molecular characteristics associated with LUAD has been gradually recognized and are becoming the focus of various studies. In the present study, we selected 53 DELs, 21 DEGs, and 8 DEMs to build a ceRNA network that centered on three core miRNAs that were downregulated, including miR-195, miR-143, and miR-144. Then, a signature consisting of two mRNAs (HOXA10 and CBX2) and four lncRNAs (LINC00460, LINC00330, DGCR5, and C14orf132) was ultimately constructed. Our results showed that there were significant differences in the survival rates between the high and low-risk groups according to the risk scores of the identified signature. In addition, survival analysis revealed that the presence of HOXA10, CBX2, and DGCR5 as a risk factor led to a significant decrease in LUAD patient survival time, while C14orf132 was regarded as a protective factor, extending survival. Therefore, this signature is a strong predictor of overall survival.

Since miRNAs play a role in the molecular regulation of transcription [14], changes in these may impact prognosis. It has been widely recognized by scholars that miRNAs with high node degrees have shown an important inhibitory effect on LUAD and miR-195 can suppress the course of lung adenocarcinoma by regulating CD4 + T cell activation[16]. In LUAD, oncogenic networks and oncogenes guided by miR-143 were identified and may help with tumor suppressor effects, establish new prognostic indicators, and find

therapeutic targets[17]. The propagation, migration, and invasion of LUAD cells are obstructed by overexpression of the miR-144 family target, EZH2[18]. These studies all indicate that miRNAs in the ceRNA network may act as suppressors in LUAD. Meanwhile, these data provide strong clues that mRNAs and lncRNAs in the network are potential candidates for LUAD prognosis.

We relied on the intrinsic correlation between epigenetic characteristics and transcriptional regulation to identify an RNA-based signature as a reliable prognostic tool. We have also identified previous research that supports their importance in cancer. High expression of HOXA10 can promote tumor development in gastric cancer, hepatocellular carcinoma, and ovarian cancer[19–21]. More importantly, downregulated LINC00483 suppresses tumor cell invasion, migration, and epithelial to mesenchymal transition. It can also bind to miR-144 and encourage the radiosensitivity of LUAD by inhibiting the expression of HOXA10 [22], which promotes LUAD progression directly by enhancing Wnt/ β -catenin signaling[23]. Studies suggest that we can improve the clinical radiotherapy effect on LUAD by inhibiting the transcriptional regulation of HOXA10, thus prolonging survival. The other mRNA in the signature is CBX2, which is overexpressed in cancers, including breast cancer and hepatocellular carcinoma, and is significantly related to poor prognosis[24–25], indicating that its activity provides an advantages for the growth and development of cancer[26]. In addition, targeted therapy studies in non-small cell lung cancer clearly show that SMARCE1 inhibits EGFR expression, in part by modulating the level of the polycomb repressive complex component CBX2. Specifically, SMARCE1 interacts directly with SWI/SNF and the EGFR oncogenic signal, as an important regulator of the drug response to MET and ALK inhibitors in non-small cell lung cancer cells[27]. We speculate that CBX2 may have strong performance by regulating its own expression CBX2 and affecting target therapies. Moreover, survival analysis in the current study revealed that HOXA10 and CBX2 were risk factors for reduced survival and this was verified in the above studies.

Furthermore, lncRNAs act as ceRNA or miRNA sponges, representing an extensive form of gene expression regulation at the post-transcriptional level[28]. Out of four lncRNAs (LINC00460, LINC00330, DGCR5, and C14orf132) in the signature, C14orf132 and LINC00330 were found to be related to the prognosis of colorectal cancer and bladder cancer, respectively[29–30]. In this study, high expression of C14orf132 could prolong the survival of LUAD patients, but its specific mechanism of action needs to be explored. Via targeting of the miR-302c-5p/FOXA1 axis, LINC00460 can facilitate tumor growth in LUAD[31]. Similarly, LINC00460 has also been confirmed to promote the progression of multiple cancers through different target genes[32–34]. The last lncRNA, DGCR5, is slightly different because when it is highly expressed it exerts anticancer activity in various cancers, including gastric cancer, hepatocellular carcinoma, cervical cancer, and bladder cancer, suggesting a good prognosis[35–38]. Interestingly, DGCR5 promotes LUAD progression by inhibiting hsa-mir-22-3p[39], which means higher expression DGCR5 implies a worse prognosis for LUAD patients and this finding was consistent with our current survival analysis results. DGCR5 may be considered a unique prognostic indicator for LUAD. However, the underlying function remains unclear and is worth investigating.

By GO and KEGG functional enrichment analysis, the biological functions of the hub DEGs in the network were identified. Downregulated DEGs were mainly involved in the cell cycle, while upregulated DEGs were

linked with protein digestion and absorption. A previous study showed that cell cycle pathway disruption can cause cell cycle arrest, which is related to the prognosis of human cancers[40]. Moreover, the GO terms were related to protein binding, the nucleoplasm, and DNA replication. The first enrichment function, protein binding, is important during tumor metastasis when tumor cells interact with the microenvironment through binding of cell surface receptors to protein ligands[41]. The transcriptionally active genes, especially those involved in developmental regulation and the cell cycle, interact in the nucleus with the nuclear porin[42]. It is also known that DNA replication disorders cause genomic instability and confer genetic diversity during tumorigenesis[43], explaining how the above pathways may identify effective treatment strategies for LUAD.

For further verification of the significance of this study's findings, we compared our risk score results with clinical risk factors (age, pathological stage, and TNM stage), and determined that our risk score showed a superior effect on survival, although we cannot exclude the possibility that clinical risk factors affect patient prognosis. Different from previous studies, we constructed a ceRNA network centered on prognosis-related miRNAs and used univariate and multivariate analysis to select an effective mRNA-lncRNA signature from the network. We then used clinical risk factors to validate the independent predictive significance of the signature. We believed that the use of the various statistical tests may make the results more reliable. However, this study is not without flaws. First, this study was based on an online database, which has limitations. Second, this study is somewhat simple, so additional in-depth functional analysis of the four lncRNAs in the signature is necessary. Even so, our research may provide guidance for future studies, which may help in the selection of clinical treatment targets.

5. Conclusion

In conclusion, from a ceRNA network, we constructed a signature consisting of two mRNAs (HOXA10 and CBX2) and four lncRNAs (LINC00460, LINC00330, DGCR5, and C14orf132), which could be regarded as a prognostic biomarker in LUAD and may provide options for clinical treatment.

Abbreviations

LUAD: Lung adenocarcinoma; TCGA: The Cancer Genome Atlas; AJCC: American Joint Committee on Cancer; AUC: Area under the curve; ROC: receiver operating characteristic; GDC: Genomic Data Commons; MREs: miRNA response elements; DEGs: Differentially expressed genes or mRNAs; DELs: Differentially expressed lncRNAs; DEMs: Differentially expressed miRNAs; FDR: False discovery rate; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes.

Declarations

Acknowledgements

Not applicable.

Authors' Contributions

CS, JW, JW and XZ design research topic. JW, XZ, XZ, and YY obtained data on lung adenocarcinoma. JW, XZ, XM and XX build the related network and analyze the data. JW and XZ write the manuscript. JZ, LL, PS and CS revise the manuscript. All authors read and approved the final manuscript.

Funding

This work is supported by the grants from National Natural Science Foundation of China (81673799, 81703915, 81973677)

Availability of data and materials

All data analyzed in this study are from open data (freely available to anyone) at TCGA database: "<https://xenabrowser.net/datapages/>"

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no conflict of interests.

References

1. Ma T, Ma H, Zou Z, et al.: The Long Intergenic Noncoding RNA 00707 Promotes Lung Adenocarcinoma Cell Proliferation and Migration by Regulating Cdc42. *Cell. Physiol. Biochem.* 2018;45(4)
2. Wilkerson MD, Yin X, Walter V, et al.: Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation. *PLoS ONE* 2012;7(5)
3. Press RH, Zhang C, Cassidy RJ, et al.: Targeted sequencing and intracranial outcomes of patients with lung adenocarcinoma brain metastases treated with radiotherapy. *Cancer* 2018 09 01; 124:17
4. Hung JJ, Jeng WJ, Chou TY, et al.: Prognostic value of the new International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society lung adenocarcinoma classification on death and recurrence in completely resected stage I lung adenocarcinoma. *Ann Surg* 2013;258:1079-86.

5. Jeong E, Moon SU, Song M, et al.: Transcriptome modeling and phenotypic assays for cancer precision medicine. *Arch. Pharm. Res.* 2017 Aug;40(8)
6. Jing S, Yun-Hui L, Yan-Qiu Z, et al.: Integrated analysis of long non-coding RNA-associated ceRNA network reveals potential lncRNA biomarkers in human lung adenocarcinoma. *Int. J. Oncol.* 2016 Nov;49(5)
7. Schmitt, A.M, Chang, H.Y.: Long noncoding RNAs: at the intersection of cancer and chromatin biology. *Cold Spring Harb Perspect Med* 2017 Jul 05;7(7)
8. Quinodoz, S, Guttman, M.: Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. *Trends Cell Biol.* 2014 Nov;24(11)
9. Vance, K.W, Ponting, C.P.: Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends Genet.* 2014 Aug;30(8)
10. Bergmann, J.H, Spector, D.L.: Long non-coding RNAs: modulators of nuclear structure and function. *Curr. Opin. Cell Biol.* 2014 Feb;26
11. Chen, L.L.: Linking long noncoding RNA localization and function. *Trends Biochem. Sci.* 2016 09;41(9)
12. Wang H, Liang L, Dong Q, et al.: Long noncoding RNA miR503HG, a prognostic indicator, inhibits tumor metastasis by regulating the HNRNPA2B1/NF- κ B pathway in hepatocellular carcinoma. *Theranostics* 2018;8(10)
13. Xie Y, Zhang Y, Du L, et al.: Circulating long noncoding RNA act as potential novel biomarkers for diagnosis and prognosis of non-small cell lung cancer. *Mol Oncol* 2018 05;12(5)
14. Xia W, Mao Q, Chen B, et al.: The TWIST1-centered competing endogenous RNA network promotes proliferation, invasion, and migration of lung adenocarcinoma. *Oncogenesis* 2019 Oct 23;8(11)
15. Jiang Y, Xie J, Huang W, et al.: Tumor Immune Microenvironment and Chemosensitivity Signature for Predicting Response to Chemotherapy in Gastric Cancer. *Cancer Immunol Res* 2019 Dec;7(12)
16. Yuan C, Xiang L, Bai R, et al.: MiR-195 restrains lung adenocarcinoma by regulating CD4+ T cell activation via the CCDC88C/Wnt signaling pathway: a study based on the Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO) and bioinformatic analysis. *Ann Transl Med* 2019 Jun;7(12)
17. Sanada H, Seki N, Mizuno K, et al.: miR-143 Involvement of Dual Strands of (and) and Their Target Oncogenes in the Molecular Pathogenesis of Lung Adenocarcinoma. *Int J Mol Sci* 2019 Sep 11;20(18)
18. Liu C, Yang Z, Deng Z, et al.: Downregulated miR-144-3p contributes to progression of lung adenocarcinoma through elevating the expression of EZH2. *Cancer Med* 2018 11;7(11)
19. Song C, Han Y, Luo H, et al.: HOXA10 induces BCL2 expression, inhibits apoptosis, and promotes cell proliferation in gastric cancer. *Cancer Med* 2019 Sep;8(12)
20. Zhang Y, Chen J, Wu SS, et al.: HOXA10 knockdown inhibits proliferation, induces cell cycle arrest and apoptosis in hepatocellular carcinoma cells through HDAC1. *Cancer Manag Res* 2019;11

21. Zhang HY, Li JH, Li G, et al.: Activation of ARK5/miR-1181/HOXA10 axis promotes epithelial-mesenchymal transition in ovarian cancer. *Oncol. Rep.* 2015 Sep;34(3)
22. Yang QS, Li B, Xu G, et al.: Long noncoding RNA LINC00483/microRNA-144 regulates radiosensitivity and epithelial-mesenchymal transition in lung adenocarcinoma by interacting with HOXA10. *J. Cell. Physiol.* 2019 Jul;234(7)
23. Sheng K, Lu J, Zhao H.: ELK1-induced upregulation of lncRNA HOXA10-AS promotes lung adenocarcinoma progression by increasing Wnt/ β -catenin signaling. *Biochem. Biophys. Res. Commun.* 2018 06 27;501(3)
24. Jangal M, Lebeau B, Witcher M.: Beyond EZH2: is the polycomb protein CBX2 an emerging target for anti-cancer therapy? *Expert Opin. Ther. Targets* 2019 Jul;23(7)
25. Mao J, Tian Y, Wang C, et al.: CBX2 Regulates Proliferation and Apoptosis via the Phosphorylation of YAP in Hepatocellular Carcinoma. *J Cancer* 2019;10(12)
26. Zheng S, Lv P, Su J, et al.: Overexpression of CBX2 in breast cancer promotes tumor progression through the PI3K/AKT signaling pathway. *Am J Transl Res* 2019;11(3)
27. Papadakis AI, Sun C, Knijnenburg TA, et al.: SMARCE1 suppresses EGFR expression and controls responses to MET and ALK inhibitors in lung cancer. *Cell Res.* 2015 Apr; 25(4)
28. Tianshi Ma, Hongwei Ma, Zigu Zou, et al.: The Long Intergenic Noncoding RNA 00707 Promotes Lung Adenocarcinoma Cell Proliferation and Migration by Regulating Cdc4. *Cell Physiol Biochem* 2018;45:1566-1580
29. Liu H, Gu X, Wang G, et al.: Copy number variations primed lncRNAs deregulation contribute to poor prognosis in colorectal cancer. *Aging(Albany NY)* 2019 Aug 22; 11(16)
30. Zhu N, Hou J, Wu Y, et al.: Integrated analysis of a competing endogenous RNA network reveals key lncRNAs as potential prognostic biomarkers for human bladder cancer. *Medicine(Baltimore)* 2018 Aug; 97(35)
31. Ye JJ, Cheng YL, Deng JJ, et al.: lncRNA LINC00460 promotes tumor growth of human lung adenocarcinoma by targeting miR-302c-5p/FOXA1 axis. *Gene* 2019 Feb 15;685
32. Zhu Y, Yang L, Chong QY, et al.: Long noncoding RNA linc00460 promotes breast cancer progression by regulating the miR-489-5p/FGF7/AKT axis. *Cancer Manag Res* 2019;11
33. Wang F, Liang S, Liu X, et al.: LINC00460 modulates KDM2A to promote cell proliferation and migration by targeting miR-342-3p in gastric cancer. *Onco Targets Ther* 2018;11
34. Zhang Y, Liu X, Li Q, et al.: lncRNA LINC00460 promoted colorectal cancer cells metastasis via miR-939-5p sponging. *Cancer Manag Res* 2019;11
35. Xu Y, Zhang G, Zou C, et al.: Long noncoding RNA DGCR5 suppresses gastric cancer progression by acting as a competing endogenous RNA of PTEN and BTG1. *J. Cell. Physiol.* 2019 Jul;234(7)
36. Sun Y, Sun H.: Propofol exerts anticancer activity on hepatocellular carcinoma cells by raising lncRNA DGCR5. *J. Cell. Physiol.* 2019 Sep 19

37. Liu Y, Chang Y, Lu S, et al.: Downregulation of long noncoding RNA DGCR5 contributes to the proliferation, migration, and invasion of cervical cancer by activating Wnt signaling pathway. *J. Cell. Physiol.* 2019 Jul;234(7)
38. Fang C, He W, Xu T, et al.: Upregulation of lncRNA DGCR5 correlates with better prognosis and inhibits bladder cancer progression via transcriptionally facilitating P21 expression. *J. Cell. Physiol.* 2019 May;234(5)
39. Dong HX, Wang R, Jin XY, et al.: LncRNA DGCR5 promotes lung adenocarcinoma (LUAD) progression via inhibiting hsa-mir-22-3p. *J. Cell. Physiol.* 2018 05;233(5)
40. Park, M.T, S.J. Lee. Cell cycle and cancer. *J Biochem Mol Biol*, 2003. 36(1): p. 60-65.
41. Liu Z, Han X, Chen R, et al.: Microfluidic Mapping of Cancer Cell-Protein Binding Interaction. *ACS Appl Mater Interfaces* 2017 Jul 12;9(27)
42. Kalverda B, Pickersgill H, Shloma VV, et al.: Nucleoporins directly stimulate expression of developmental and cell-cycle genes inside the nucleoplasm. *Cell* 2010 Feb 05;140(3)
43. Kitao H, Iimori M, Kataoka Y, et al.: DNA replication stress and cancer chemotherapy. *Cancer Sci.* 2018 Feb;109(2)

Figures

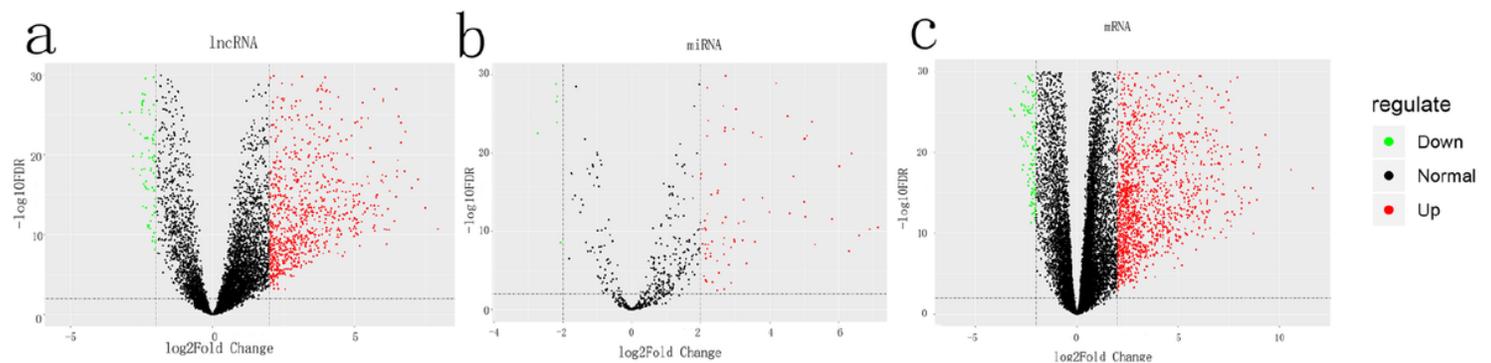


Figure 1

Volcano plots of differentially expressed (DE) RNAs. a. DE lncRNAs, b. DE miRNAs, and c. DE mRNAs in lung adenocarcinoma (LUAD). The red and green dots represent upregulated and downregulated lncRNAs, respectively. The cutoff criteria are an FDR < 0.01, $|\log_2 \text{fold change} (\log_2 \text{FC})| > 2$.

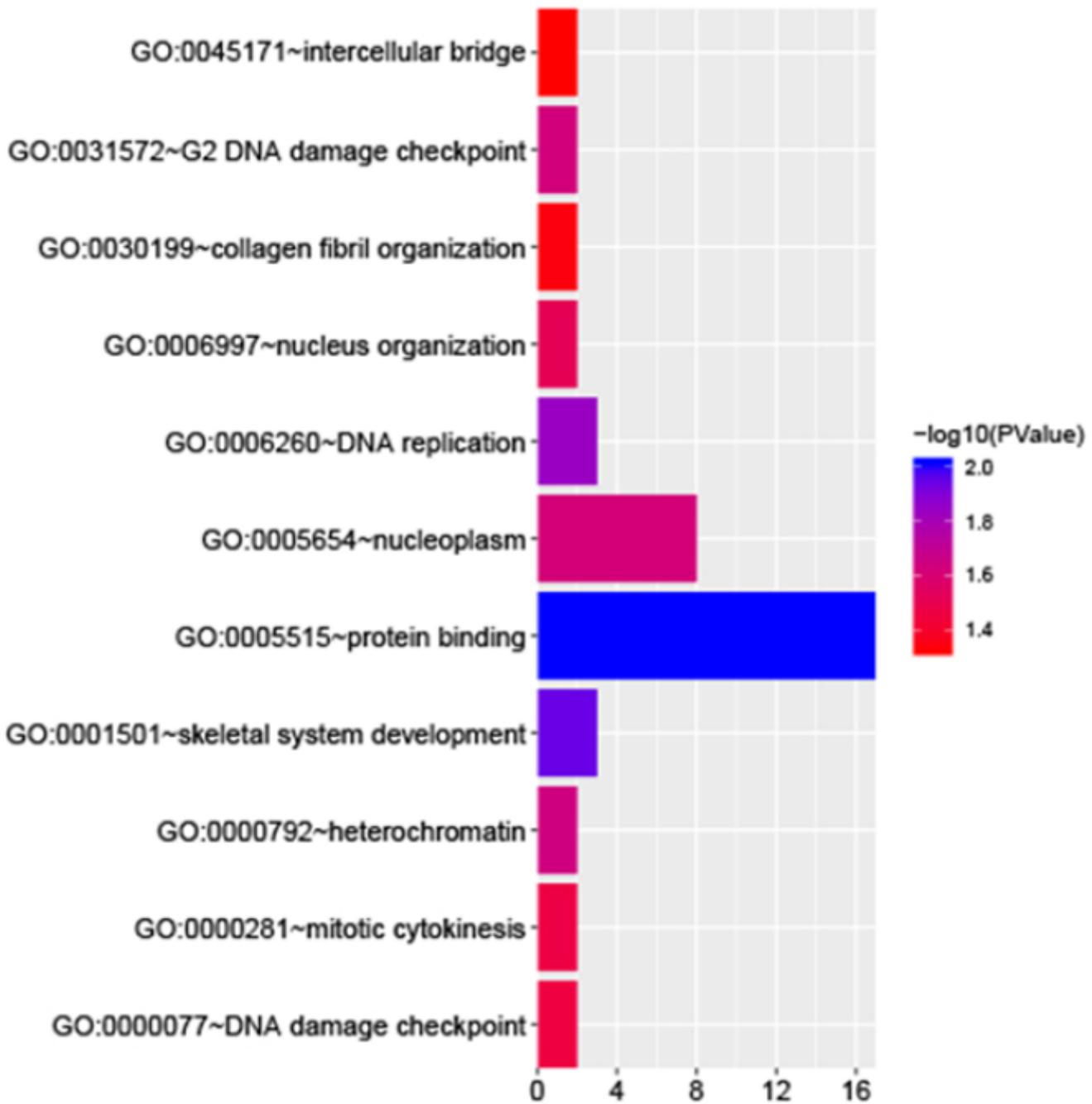


Figure 3

The histogram shows the GO terms of 21 mRNAs in the ceRNA network ($p < 0.05$). The correlation is more significant as the blue/red ratio increases.

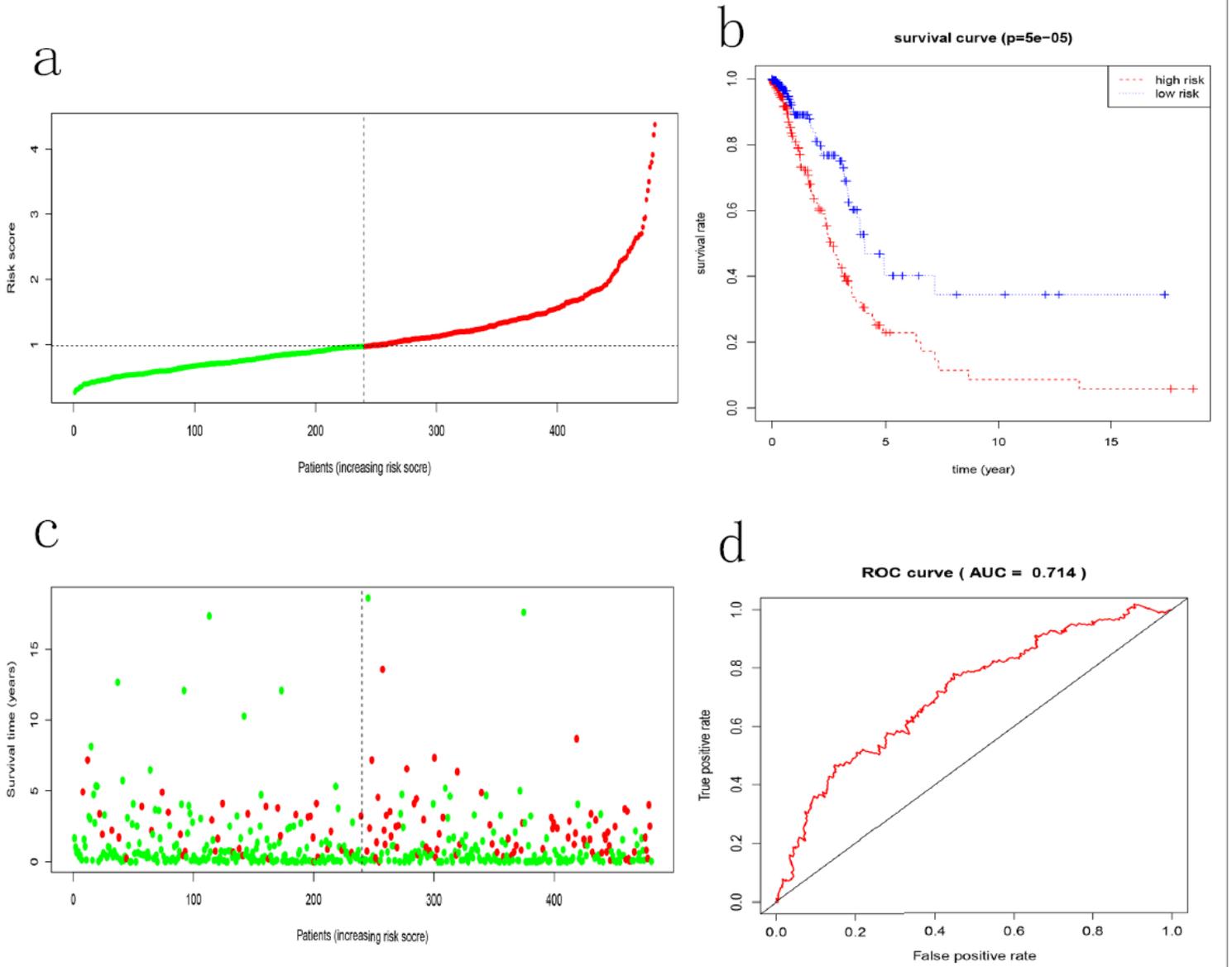


Figure 4

4a. Risk scores for all patients. 4b. Distribution of survival status of all patients. The red dots represent death and the green dots represent survival. 4c. Survival curve of the low-risk and high-risk groups based on the median risk score. 4d. The receiver operating characteristic (ROC) curve of the three-year survival rate.

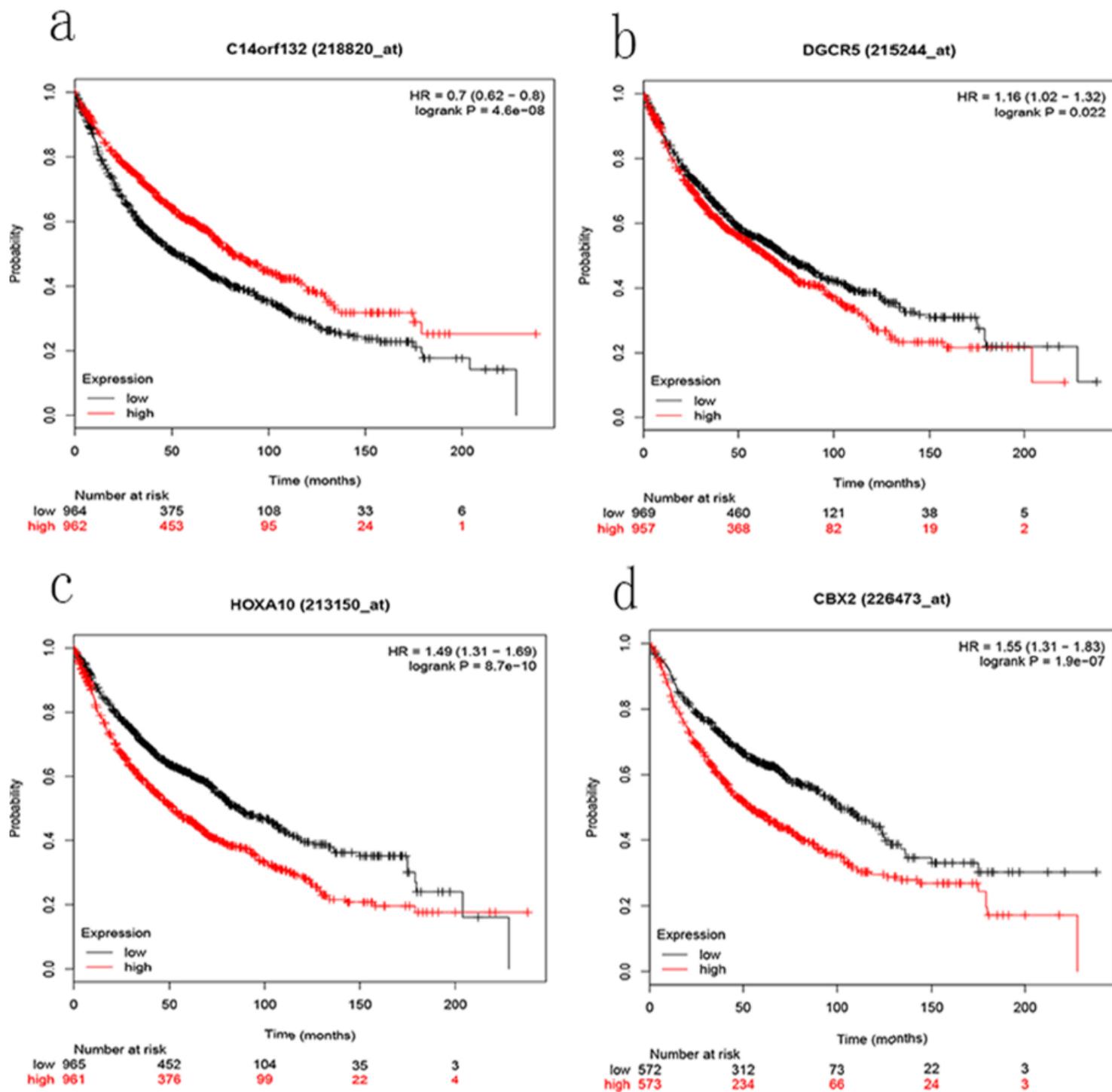


Figure 5

The Kaplan-Meier plot of overall patient survival in relation to four genes. a. C14orf132 b. DGCR5 c. HOXA10 d. CBX2.

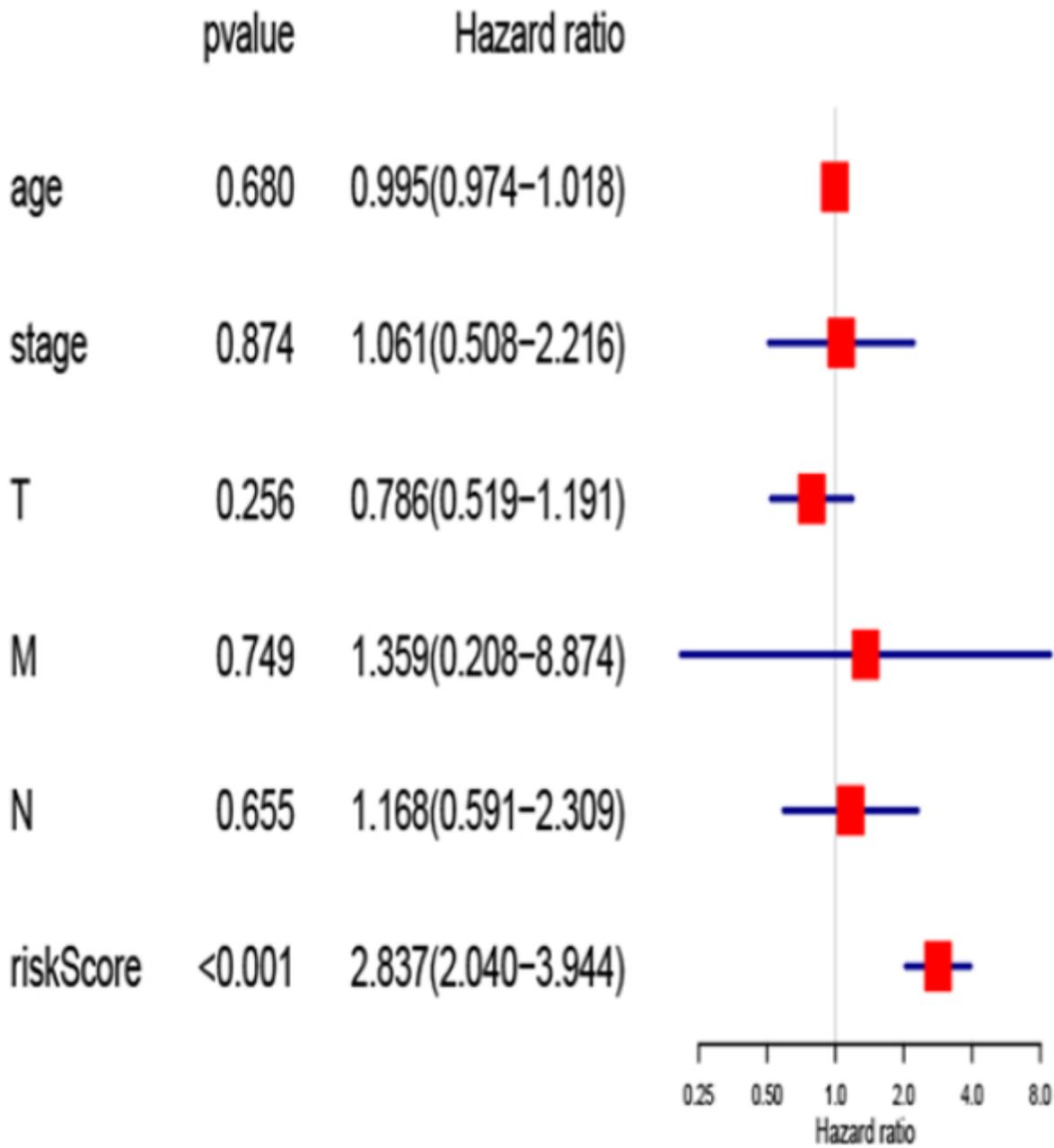


Figure 6

Survival analysis of LUAD patients with different risk factors