Competing endogenous RNA network identifies mRNA biomarkers for overall survival of lung adenocarcinoma: Two novel on-line precision medicine predictive tools

Jinsong Lin 1, Shubiao Lu 2, Zhijian Jiang 2, Chongjing Hu 2, Zhiqiao Zhang 2

1 Department of Internal Medicine, The Affiliated Chencun Hospital of Shunde Hospital, Southern Medical University, Shunde, Guangdong, China
2 Department of Internal Medicine, The Affiliated Chencun Hospital of Shunde Hospital, Southern Medical University, Shunde, Guangdong, China

Corresponding Author: Zhiqiao Zhang
Email address: sdgrxjbk@smu.edu.cn

Background: Individual mortality risk predicted curve at the individual level can provide valuable information for directing individual treatment decision. The present study attempted to explore potential post-transcriptional biological regulatory mechanism related with overall survival of lung adenocarcinoma (LUAD) patients through competitive endogenous RNA (ceRNA) network and develop two precision medicine predictive tools for predicting the individual mortality risk curves for overall survival of LUAD patients.

Methods: Multivariable Cox regression analyses were performed to explore the potential prognostic indicators, which were used to construct a prognostic model for overall survival of LUAD patients. Time-dependent receiver operating characteristic (ROC) curves were used to assess the predictive performance of prognostic model.

Results: There were 494 LUAD patients in model cohort and 233 LUAD patients in validation cohort. Differentially expressed mRNAs, miRNAs, and lncRNAs were identified between LUAD tissues and normal tissues. A ceRNA regulatory network was constructed on previous differentially expressed mRNAs, miRNAs, and lncRNAs. Fourteen mRNA biomarkers were identified as independent risk factors by multivariate Cox regression and used to develop a prognostic model for overall survival of LUAD patients. The C-indexes of prognostic model in model group were 0.786 (95% CI: 0.744-0.828) , 0.736 (95% CI: 0.694-0.778) and 0.766 (95% CI: 0.724-0.808) for one year, two year and three year overall survival respectively. Two precision medicine predicted tools were developed for predicting individual mortality risk curves for LUAD patients.

Conclusion: The current study explored potential post-transcriptional biological regulatory mechanism and prognostic biomarkers for overall survival of LUAD patients. Two on-line precision medicine predictive tools were helpful to predict the individual mortality risk predicted curves for overall survival of LUAD patients. Smart Cancer Survival Predictive System could be used at https://zhangzhqiao2.shinyapps.io/Smart_cancer_predictive_system_9_LUAD_E1002/.
Competing endogenous RNA network identifies mRNA biomarkers for the overall survival of lung adenocarcinoma:

Two novel online precision medicine predictive tools

Jinsong Lin#, Shubiao Lu#, Zhijian Jiang, Chongjing Hu, Zhiqiao Zhang*

Department of Internal Medicine, The Affiliated Chencun Hospital of Shunde Hospital, Southern Medical University, Shunde, Guangdong, China

# These authors were the co-first authors.
* Corresponding author: Zhiqiao Zhang (E-mail: sdgrjkbk@163.com)
Abstract

Background: Individual mortality risk prediction curves at the individual level can provide valuable information for directing individualized treatment decisions. The present study attempted to explore potential posttranscriptional biological regulatory mechanisms related to the overall survival of lung adenocarcinoma (LUAD) patients through a competitive endogenous RNA (ceRNA) network and develop two precision medicine tools for predicting the individual mortality risk curves for the overall survival of LUAD patients.

Methods: Multivariable Cox regression analyses were performed to explore the potential prognostic indicators, which were used to construct a prognostic model for the overall survival of LUAD patients. Time-dependent receiver operating characteristic (ROC) curves were used to assess the predictive performance of the prognostic model.

Results: There were 494 LUAD patients in the model cohort and 233 LUAD patients in the validation cohort. Differentially expressed mRNAs, miRNAs, and lncRNAs were identified between LUAD tissues and normal tissues. A ceRNA regulatory network was constructed on previous differentially expressed mRNAs, miRNAs, and lncRNAs. Fourteen mRNA biomarkers were identified as independent risk factors by multivariate Cox regression and used to develop a prognostic model for the overall survival of LUAD patients. The C-indexes of the prognostic model in the model group were 0.786 (95% CI: 0.744-0.828), 0.736 (95% CI: 0.694-0.778), and 0.766 (95% CI: 0.724-0.808) for one-year, two-year and three-year overall survival, respectively. Two precision medicine prediction tools were developed for predicting individual mortality risk curves for LUAD patients.

Conclusion: The current study explored potential posttranscriptional biological regulatory mechanisms and prognostic biomarkers for the overall survival of LUAD patients. Two online precision medicine predictive tools were helpful to predict the individual mortality risk prediction curves for the overall survival of LUAD patients. Smart
Cancer Survival Predictive System could be used at

https://zhangzhiqiao2.shinyapps.io/Smart_cancer_predictive_system_9_LUAD_E1002/.

**Keywords:** Competitive endogenous RNA, lung adenocarcinoma, mRNA, overall survival, prognostic model
Background
Lung cancer is the most common cause of malignant tumours and tumour-related death (Bray et al. 2018). There were approximately 2.1 million newly diagnosed lung cancer patients and 1.8 million lung cancer deaths worldwide in 2018 (Bray et al. 2018). As the most prevalent type of non-small-cell lung cancer, lung adenocarcinoma (LUAD) is the leading cause of lung cancer-related mortality (Bray et al. 2018). The overall survival of patients with LUAD is extremely poor, with a 5-year overall survival (OS) rate of less than 20% (Lin et al. 2016). Although great progress has been made in cancer diagnosis and targeted therapy, the 5-year overall survival rate of patients with LUAD remains low (Qi et al. 2016). Therefore, establishing a reliable prognostic model for screening high-risk patients with poor overall survival is of great clinical significance for optimizing individualized treatments and improving the management of patients with cancer.

Several studies have reported potential molecular biological regulation mechanisms for different tumours (Huang et al. 2018; Shi et al. 2018; Zeng et al. 2018; Zhong et al. 2018). Salmena et al. proposed a posttranscriptional molecular biological regulation mechanism named competing endogenous RNA (ceRNA) (Salmena et al. 2011). Long noncoding RNAs (lncRNAs), as endogenous molecular sponges, can upregulate the expression of mRNA through competitive binding with miRNA response elements (Thomson & Dinger 2016). Some researchers have used a ceRNA regulatory network to explore the potential molecular biological regulation mechanism for the prognosis of patients with LUAD. (Li et al. 2018; Sui et al. 2016; Wang et al. 2018; Wang et al. 2019). Xing Li et al. developed a prognostic signature (without external validation) for the prognosis of lung cancer by using 29 mRNAs and 3 lncRNAs (Li et al. 2018). However,
this prognostic model is too complex to calculate and be used for clinical application. A nomogram is a graphical tool that can directly display the results, with the advantages of easy calculation and good interpretability (Cheng 2018; Song et al. 2018). Based on the prognostic nomogram, our research team has further developed precision medicine predictive tools for providing individual mortality risk prediction curves for different cancers (Cheng et al. 2019a; Zhang et al. 2019b; Zhang et al. 2019d). Therefore, the present study attempted to construct two precision medicine predictive tools for providing individualized mortality risk prediction curves for the OS of LUAD patients by using prognostic mRNA biomarkers through a ceRNA regulatory network.

**Materials and Methods**

**Protocol approval**

Data were collected as described in our previous studies (Cheng et al. 2019b; Zhang et al. 2019a; Zhang et al. 2019c). The study data in the current study were downloaded from The Cancer Genome Atlas (TCGA) database ([https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga](https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga)), UCSC Xena database ([https://xena.ucsc.edu/](https://xena.ucsc.edu/)), cBioPortal database ([http://www.cbioportal.org/](http://www.cbioportal.org/)), and Gene Expression Omnibus (GEO) database ([https://www.ncbi.nlm.nih.gov/geo/](https://www.ncbi.nlm.nih.gov/geo/)). The current study did not require ethics approval, as all study datasets were downloaded and analysed in accordance with the corresponding data policies of previous databases.

**The gene expression information of model cohort**

Data were collected as described in our previous studies (Zhang et al. 2020; Zhang et al. 2019c). The current study downloaded the original RNA expression information from TCGA database ([https://tcga-data.nci.nih.gov/docs/publications/tcga/](https://tcga-data.nci.nih.gov/docs/publications/tcga/)) on the Illumina HiSeq 2000 RNA Sequencing platform. The RNA symbols were determined according to GENCODE Version 29 ([https://www.gencodegenes.org/human/](https://www.gencodegenes.org/human/)). The RNA expression
dataset contained 8,848 lncRNAs and 21,138 mRNAs from 535 lung adenocarcinoma samples and 59 normal samples. The miRNA expression dataset was obtained from the UCSC Xena database (https://xena.ucsc.edu/). The miRNA expression dataset involved 1,881 miRNAs from 518 lung adenocarcinoma and 46 normal samples. According to the median of the original gene expression value, the original expression values were converted into “0” for lower expression and “1” for higher expression.

Differentially expressed analyses
Differentially expressed analyses between tumour samples and normal samples were performed by the edgeR package with R software (version 3.5.2) as described in our previous studies (Zhang et al. 2019c; Zhang et al. 2018). For differentially expressed analyses, a false discovery rate (FDR) less than 0.05 and a ratio value of 1.5 between tumour tissues and normal tissues were defined as the thresholds.

The clinical information of model cohort
Data were collected as described in our previous studies (Zhang et al. 2019c; Zhang et al. 2018). The clinical information was obtained from the cBioPortal database. Patients with LUAD with inadequate survival information or OS times less than one month (for living patients) were removed (n=19). Therefore, 503 LUAD patients were identified with adequate OS information. There were 494 patients with adequate gene detection information and survival information when performing intersection between clinical datasets on the gene detection dataset.

The corresponding information of validation cohort
Data were collected as described in our previous studies (Zhang et al. 2019c; Zhang et al. 2018). The current study downloaded GSE37745 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37745) and GSE50081
Statistical analysis

The statistical analyses were carried out as described in our previous studies (Zhang et al. 2020; Zhang et al. 2019c): continuous variables are presented as the mean ± standard deviation (SD) or median (first percentile, third percentile) as appropriate. Continuous variables were compared by t-test or Mann-Whitney U test as appropriate. Categorical variables were compared through the chi-squared test or Fisher’s exact test as appropriate. Time-dependent receiver operating characteristic (ROC) curves were used to assess the predictive performance of prognostic models. Decision curve analysis (DCA) is an assessment method to compare the predictive performance of prognostic models (Localio & Goodman 2012; Vickers et al. 2008; Vickers & Elkin 2006). The statistical analyses were carried out using SPSS Statistics 19.0 (SPSS Inc., USA) and R software (version 3.5.2) with the following packages: “GOplot”, "timeROC", "rms", "pROC", "survival", “clusterProfiler” and "glmnet", as described in our previous studies (Zhang et al. 2020; Zhang et al. 2019c). A P value < 0.05 was the threshold value for statistical significance.

Results

Study cohorts

In the model group (Doc S1), 182 patients (36.8%) out of 494 patients died during the follow-up period, while 127 patients (54.5%) out of 233 patients died in the verification group (Doc S2). The basic features of the two groups are compared in Table 1.
Differentially expressed analyses

Based on the given threshold, 3310 upregulated IncRNAs, 675 downregulated IncRNAs, 95 upregulated miRNAs, 30 downregulated miRNAs, 4913 upregulated mRNAs, and 1921 downregulated mRNAs were identified as differentially expressed RNAs between LUAD tissues and normal tissues. Based on previous differentially expressed mRNAs, 2982 mRNAs were identified as prognostic mRNA indicators by univariate Cox analyses.

Construction of a competing endogenous RNA network

To obtain the IncRNA-miRNA pairs, the miRcode database (http://www.mircode.org/) was searched according to the differentially expressed IncRNAs. To obtain the miRNA-mRNA pairs, TargetScan (http://www.targetscan.org/), miRDB (http://www.mirdb.org/), and miRTarBase (http://mirtarbase.mbc.nctu.edu.tw/php/index.php) were searched according to the previous IncRNA-targeted miRNAs. The miRNA-predicted mRNAs that could be searched in three databases were defined as miRNA-targeted mRNAs. Then, we determined the interaction between miRNA-targeted mRNAs and prognostic mRNAs to identify prognosis-associated miRNA-targeted mRNAs. Finally, twenty-two IncRNAs, twenty-nine miRNAs, and seventy-three mRNAs were used to build the ceRNA network for the overall survival of Patients with LUAD. The ceRNA regulatory network was depicted by Cytoscape software.

Functional enrichment analyses

To explore the biological functions of prognostic mRNAs, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed on prognostic mRNAs in the ceRNA network. The bar plot and dot plot of prognostic mRNAs are depicted in Fig. 2A and Fig. 2B. GO terms and KEGG pathways of prognostic mRNAs are depicted in Fig. 3A and Fig. 3B.
**Development of a prognostic nomogram**

The previous prognostic mRNAs were used to construct a prognostic model for overall survival. The information on these mRNAs is summarized in Table 2. The risk score of the prognostic model was calculated by using the following formula: risk score =

\[-0.7305 \times \text{DNAJC27} + 0.4776 \times \text{NPAS2} + \]

\[(0.3941 \times \text{PHKA1}) - (0.5537 \times \text{CDADC1}) + (0.4792 \times \text{PTGFRN}) + (0.3941 \times \text{DDIT4}) - (0.4133 \times \text{SCAMP5}) - (0.4367 \times \text{SPRY2}) + (0.3850 \times \text{CSE1L}) - (0.3940 \times \text{ELAVL4}) + (0.3293 \times \text{TRIM29}) + (0.3685 \times \text{LPGAT1}) - (0.3279 \times \text{TRPC3}) + (0.3257 \times \text{DCBLD2}) \]

A prognostic nomogram chart is presented in Fig. 4. Kaplan-Meier survival curves (Supplementary Fig. 1) demonstrated that these 14 prognostic mRNAs were significantly correlated with OS ($P<0.05$). The results of differential expression analysis of 14 enrolled mRNAs are listed in Table 3.

**Predictive performance of the prognostic model**

Based on the median value, LUAD patients in the model group could be classified into a high-risk subgroup and a low-risk subgroup. The overall survival rate (Fig. 5A) in the high-risk group was significantly poorer than that in the low-risk group ($P<0.001$).

Harrell's concordance indexes (C-indexes) of the prognostic signature for overall survival in the model group were 0.786 (95% CI: 0.744-0.828), 0.736 (95% CI: 0.694-0.778), and 0.766 (95% CI: 0.724-0.808) for 1-year, 2-year, and 3-year OS, respectively (Fig. 5B).

The scatter plot and the interaction distribution scatter plot are presented in Fig. 5C and Fig. 5D. Calibration curves are depicted in Supplementary Fig. 2A for 1 year, Supplementary Fig. 2B for 2 years, and Supplementary Fig. 2C for 3-year overall survival.

**External validation of the prognostic model**

Kaplan-Meier analysis (Fig. 6A) indicated that there was a significant difference in OS...
between the low-risk group and the high-risk group in the validation dataset \((P<0.001)\).

The C-indexes of the prognostic model for overall survival in the validation dataset were 0.948 (95% CI: 0.912-0.984), 0.936 (95% CI: 0.900-0.972), and 0.890 (95% CI: 0.854-0.926) for 1-year, 2-year, and 3-year overall survival, respectively (Fig. 6B). Calibration curves are depicted in Supplementary Fig. 3A for 1 year, Supplementary Fig. 3B for 2 years, and Supplementary Fig. 3C for 3-year overall survival.

**Independence assessment of the prognostic model**

Table 4 shows that the prognostic model was an independent influencing factor for OS after adjustment for confounding effects in the model dataset. In the validation dataset, multivariate Cox regression analyses demonstrated that the prognostic model was an independent influencing factor for OS. DCA is shown in Supplementary Fig. 4A for 1-year OS, Supplementary Fig. 4B for 2-year OS, and Supplementary Fig. 4C for 3-year OS. The clinical impact curve is presented in Supplementary Fig. 4D.

**Smart Cancer Survival Predictive System**

A precision medicine predictive tool named the Smart Cancer Survival Predictive System was developed for predicting the prognosis of LUAD patients. The Smart Cancer Survival Predictive System (Fig. 7) is available at [https://zhangzhiqiao2.shinyapps.io/Smart_cancer_predictive_system_9_LUAD_E1002/](https://zhangzhiqiao2.shinyapps.io/Smart_cancer_predictive_system_9_LUAD_E1002/).

The Smart Cancer Survival Predictive System could predict full-time mortality risk prediction curves for one specific patient (Fig. 7A). Fig. 7B demonstrates the mortality rate predicted percentage and 95% confidence interval at different user-defined time points.

**Gene Survival Analysis Screen System**
A second precision medicine predictive tool named the Gene Survival Analysis Screen System (Fig. 8) was developed to explore the survival features of prognostic mRNAs. The Gene Survival Analysis Screen System is available at https://zhangzhiqiao7.shinyapps.io/Gene_Survival_Analysis_Screening_System_9_LUA_D_E1002/. Fig. 8A depicts and compares the survival curves between two defined subgroups. Fig. 8B displays the results of univariate survival analysis for selected variables.

Discussion

The current ceRNA regulatory network depicts lncRNA-miRNA-mRNA regulatory pathways, which are helpful for understanding the biological regulatory mechanisms of OS in LUAD. The current study constructed and verified a fourteen-mRNA prognostic nomogram for OS. This fourteen-mRNA prognostic nomogram was suitable to screen LUAD patients with poor OS. Based on this fourteen-mRNA prognostic nomogram, we developed an online precision medicine predictive tool named the Smart Cancer Survival Predictive System, which can generate full-time mortality risk prediction curves for one specific patient at the individual level.

Several prognostic models have been built for predicting the prognosis of lung cancer patients (Xie & Xie 2019; Yan et al. 2018; Zuo et al. 2019). However, these previous prognostic models could only provide prognostic information for a special subgroup at the group level and failed to provide individual mortality risk prediction at the individual level. Our Smart Cancer Survival Predictive System was superior to the previous prognostic models for its special predictive ability in predicting individual mortality risk curves at the individual level. Meanwhile, the Smart Cancer Survival Predictive System could further provide the mortality rate predicted percentage and 95% confidence interval at different user-defined time points. These special predictive functions in smart cancer survival predictive systems are of great significance for improving individual
NPAS2 was associated with a favourable prognosis of LUAD patients (Qiu et al. 2019). CDADC1 was found to be associated with survival in bladder cancer (Shivakumar et al. 2017). DDIT4 promotes gastric cancer proliferation and tumorigenesis (Du et al. 2018). The knockdown or overexpression of SPRY2 promoted or suppressed the proliferation of prostate cancer cells (Gao et al. 2018). CSE1L was correlated with overall survival in patients with hepatocellular carcinoma (Zhang et al. 2019d). Ectopic expression of TRIM29 potentially contributes to metastasis and poor prognosis in patients with osteosarcoma (Zeng et al. 2017). Increased activity of TRPC3 channels is necessary for the development of ovarian cancer (Yang et al. 2009). DCBLD2 correlated with the overall survival in pancreatic ductal adenocarcinoma (Raman et al. 2018). The results of the current study were in good agreement with those of previous studies.

There are four main advantages of the current study: First, all operations are computed automatically in the background without manual calculation, which is convenient for use; Second, our tool can provide full-time individual mortality risk prediction through an individualized survival curve for a particular patient; Third, our tool can provide individual mortality risk prediction percentages and 95% CIs at specific time points (such as 12, 24, 36, 48, 60, and 72 months) through line graphs and tables; Fourth, the Gene Survival Analysis Screen System allows users to define different subgroups by themselves. Meanwhile, users are free to download, upload, and select the dataset for gene survival analysis. To the best of our knowledge, this online full-time individualized risk calculator is the first to provide full-time individual mortality risk prediction through a web calculator based on gene expression data for patients with lung cancer.

The shortcomings of our study were as follows: First, the genetic detection platforms of the model group and validation group were different and need to be taken into account when interpreting the results of the current study; Second, the performance of the
fourteen-mRNA prognostic nomogram in the verification group was better than that in the model group, which could be explained by the longer follow-up period and higher mortality in the verification group. The median survival times were 20.3 and 25.3 months ($P=0.006$) for patients who died in the model group and validation group, respectively, whereas the median survival times were 22.2 and 69.1 months ($P<0.001$) for living patients in the model group and validation group, respectively; Third, some important prognostic factors, such as surgical procedures and adjuvant therapies, were not included in the survival analysis. It is necessary to carry out multicentre, prospective, and large-sample clinical research to further study the clinical application value of fourteen-mRNA prognostic nomograms in different populations.

Conclusions

In conclusion, the current study explored potential posttranscriptional biological regulatory mechanisms and prognostic biomarkers for the overall survival of LUAD patients. Two online precision medicine predictive tools were developed and are helpful to predict the individual mortality risk prediction curves for the overall survival of LUAD patients. The Smart Cancer Survival Predictive System can be used at https://zhangzhiqiao2.shinyapps.io/Smart_cancer_predictive_system_9_LUAD_E1002/.

References

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, and Jemal A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68:394-424.

Cheng C, Wang Q, Zhu M, Liu K, and Zhang Z. 2019a. Integrated analysis reveals potential long non-coding RNA biomarkers and their potential biological functions for disease free survival in gastric cancer patients. Cancer Cell Int 19:123.

Cheng C, Wang Q, Zhu M, Liu K, and Zhang Z. 2019b. Integrated analysis reveals potential long non-coding RNA biomarkers and their potential biological functions for disease free survival in gastric cancer patients. Cancer Cell Int 19:123. 10.1186/s12935-019-0846-6

Cheng P. 2018. A prognostic 3-long noncoding RNA signature for patients with gastric cancer. J Cell Biochem
Du F, Sun L, Chu Y, Li T, Lei C, Wang X, Jiang M, Min Y, Lu Y, Zhao X, Nie Y, and Fan D. 2018. DDIT4 promotes gastric cancer proliferation and tumorigenesis through the p53 and MAPK pathways. Cancer Commun (Lond) 38:45.

Gao W, Hong Z, Huang H, Zhu A, Lin S, Cheng C, Zhang X, Zou G, and Shi Z. 2018. miR-27a in serum acts as biomarker for prostate cancer detection and promotes cell proliferation by targeting Sprouty2. Oncol Lett 16:5291-5298.

Huang Y, Xiang B, Liu Y, Wang Y, and Kan H. 2018. LncRNA CDKN2B-AS1 promotes tumor growth and metastasis of human hepatocellular carcinoma by targeting let-7c-5p/NAP1L1 axis. Cancer Lett 437:56-66.

Li X, Li B, Ran P, and Wang L. 2018. Identification of ceRNA network based on a RNA-seq shows prognostic lncRNA biomarkers in human lung adenocarcinoma. Oncol Lett 16:5697-5708. 10.3892/ol.2018.9336

Lin JJ, Cardarella S, Lydon CA, Dahlberg SE, Jackman DM, Janne PA, and Johnson BE. 2016. Five-Year Survival in EGFR-Mutant Metastatic Lung Adenocarcinoma Treated with EGFR-TKIs. J Thorac Oncol 11:556-565. 10.1016/j.jtho.2015.12.103

Localio AR, and Goodman S. 2012. Beyond the usual prediction accuracy metrics: reporting results for clinical decision making. Ann Intern Med 157:294-295. 10.7326/0003-4819-157-4-201208210-00014

Qi L, Li Y, Qin Y, Shi G, Li T, Wang J, Chen L, Gu Y, Zhao W, and Guo Z. 2016. An individualised signature for predicting response with concordant survival benefit for lung adenocarcinoma patients receiving platinum-based chemotherapy. Br J Cancer 115:1513-1519. 10.1038/bjc.2016.370

Qiu M, Chen YB, Jin S, Fang XF, He XX, Xiong ZF, and Yang SL. 2019. Research on circadian clock genes in non-small-cell lung carcinoma. Chronobiol Int 36:739-750.

Raman P, Maddipati R, Lim KH, and Tozeren A. 2018. Pancreatic cancer survival analysis defines a signature that predicts outcome. Plos One 13:e0201751.

Salmena L, Poliseno L, Tay Y, Kats L, and Pandolfo PP. 2011. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell 146:353-358.

Shi X, Zhao Y, He R, Zhou M, Pan S, Yu S, Xie Y, Li X, Wang M, Guo X, and Qin R. 2018. Three-IncRNA signature is a potential prognostic biomarker for pancreatic adenocarcinoma. Oncotarget 9:24248-24259. 10.18632/oncotarget.24443

Shivakumar M, Lee Y, Bang L, Garg T, Sohn KA, and Kim D. 2017. Identification of epigenetic interactions between miRNA and DNA methylation associated with gene expression as potential prognostic markers in bladder cancer. BMC Med Genomics 10:30.

Song W, Miao DL, and Chen L. 2018. Nomogram for predicting survival in patients with pancreatic cancer. Onco Targets Ther 11:539-545. 10.2147/ott.s154599

Sui J, Li YH, Zhang YQ, Li CY, Shen X, Yao WZ, Peng H, Hong WW, Yin LH, Pu YP, and Liang GY. 2016. Integrated analysis of long non-coding RNA associated ceRNA network reveals potential lncRNA biomarkers in human lung adenocarcinoma. Int J Oncol 49:2023-2036. 10.3892/ijo.2016.3716

Thomson DW, and Dinger ME. 2016. Endogenous microRNA sponges: evidence and controversy. Nat Rev Genet 17:272-283.

Vickers AJ, Cronin AM, Elkin EB, and Gonen M. 2008. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak 8:53. 10.1186/1472-6947-8-53
385 Vickers AJ, and Elkin EB. 2006. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 26:565-574. 10.1177/0272989x06295361

386 Wang X, Ding Y, Da B, Fei Y, and Feng G. 2018. Identification of potential prognostic long noncoding RNA signatures based on a competing endogenous RNA network in lung adenocarcinoma. *Oncol Rep* 40:3199-3212.

388 Wang X, Ding Y, Da B, Fei Y, and Feng G. 2018. Identification of a putative competitive endogenous RNA network for lung adenocarcinoma using TCGA datasets. *PeerJ* 7:e6809.

392 Xie H, and Xie C. 2019. A Six-Gene Signature Predicts Survival of Adenocarcinoma Type of Non-Small-Cell Lung Cancer Patients: A Comprehensive Study Based on Integrated Analysis and Weighted Gene Coexpression Network. *Biomed Res Int* 2019:4250613. 10.1155/2019/4250613

396 Yan H, Xin S, Ma J, Wang H, Zhang H, and Liu J. 2018. A three microRNA-based prognostic signature for small cell lung cancer overall survival. *J Cell Biochem*. 10.1002/jcb.28159

398 Yang SL, Cao Q, Zhou KC, Feng YJ, and Wang YZ. 2009. Transient receptor potential channel C3 contributes to the progression of human ovarian cancer. *Oncogene* 28:1320-1328.

400 Zeng J, Cai X, Hao X, Huang F, He Z, Sun H, Lu Y, Lei J, Zeng W, Liu Y, and Luo R. 2018. LncRNA FUNDC2P4 down-regulation promotes epithelial-mesenchymal transition by reducing E-cadherin expression in residual hepatocellular carcinoma after insufficient radiofrequency ablation. *Int J Hyperthermia* 34:802-811.

402 Zeng SX, Cai QC, Guo CH, Zhi LQ, Dai X, Zhang DF, and Ma W. 2017. High expression of TRIM29(ATDC) contributes to poor prognosis and tumor metastasis by inducing epithelial-mesenchymal transition in osteosarcoma. *Oncol Rep* 38:1645-1654.

406 Zhang Z, He T, Huang L, Ouyang Y, Li J, Huang Y, Wang P, and Ding J. 2019a. Two precision medicine predictive tools for six malignant solid tumors: from gene-based research to clinical application. *J Transl Med* 17:405. 10.1186/s12967-019-02151-8

409 Zhang Z, Li J, He T, and Ding J. 2020. Bioinformatics Identified 17 Immune Genes as Prognostic Biomarkers for Breast Cancer: Application Study Based on Artificial Intelligence Algorithms. *Front Oncol* 10:330.

411 Zuo S, Wei M, Zhang H, Chen A, Wu J, Wei J, and Dong J. 2019. A robust six-gene prognostic signature for bladder cancer cells assurance guidelines. *J buon* 23:776-781.
prediction of both disease-free and overall survival in non-small cell lung cancer. *J Transl Med* 17:152. 10.1186/s12967-019-1899-y

**Declarations**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

All authors agreed to publication.

**Availability of data and materials**

The study data are available at Supplemental documents of the current manuscript.

**Competing interests**

The authors declare no conflict of interest.

**Funding**

Guangdong Provincial Health Department (A2016450 and B2018237)

**Authors’ contributions**

Conceptualization: ZZ, JL, CH, SL, and ZJ. Data curation: ZZ, JL, and ZJ. Formal analysis: ZZ, SL, and ZJ. Funding acquisition: ZZ. Investigation: ZZ, SL, and ZJ. Methodology: ZZ, JL, and ZJ. Project administration: ZZ, SL, and ZJ. Resources: ZZ; Software: ZZ. Supervision: ZZ, SL, CH, and ZJ. Validation: ZZ, JL, CH, and SL. Visualization: ZZ, JL, and ZJ. Writing-original draft: ZZ, JL, CH, SL, and ZJ. Writing-
Acknowledgements

We would like to express sincere thanks to Dr. Gary S Collins (University of Oxford), Dr Manali Rupji (Emory University), and Mrs Qingmei Liu for inspirations, suggestions, and assistance in the development of our precision medicine tools.

Figure legends

Figure 1. Competitive endogenous RNA network chart: Red triangles represent 22 lncRNAs; blue triangles represent 29 miRNAs; green circles represent 73 mRNAs.

Figure 2. (A) The bar plot of GO terms for prognostic mRNAs; (B) The dot plot of GO terms for prognostic mRNAs.

Figure 3. (A) Chord chart of GO terms for prognostic mRNAs; (B) KEGG pathways for prognostic mRNAs.

Figure 4. The prognostic nomogram for overall survival.

Figure 5. (A) Survival curves in the model group; (B) Time-dependent receiver operating characteristic curves in the model group; (C) The distribution of prognostic model scores in the model group; (D) The overall survival status and overall survival time in the model group.

Figure 6. (A) Survival curves in the validation group; (B) Time-dependent receiver operating characteristic curves in the validation group.

Figure 7. Home page of Smart Cancer Survival Predictive System: (A) Individual mortality risk predictive curve display page; (B) Different time-point individual mortality risk prediction display page.

Figure 8. Home page of the Gene Survival Analysis Screen System: (A) Survival curve plot display page; (B) Univariate analysis table display page.

Abbreviations:

LUAD: lung adenocarcinoma; TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus; ROC: receiver operating characteristic; OS: overall survival;
IncRNA: long noncoding RNA; miRNA: microRNA; mRNA: messenger RNA; HR: hazard ratio; CI: confidence interval; AJCC: American Joint Committee on Cancer; SD: standard deviation; DCA: decision curve analysis; ceRNA: competitive endogenous RNA.

**ORCID**

Zhiqiao Zhang http://orcid.org/0000-0003-4631-8818

**Supporting information**

Supplementary Figure 1. Survival curves of prognostic genes

Supplementary Figure 2. Calibration curve of prognostic model in model group

Supplementary Figure 3. Calibration curve of the prognostic model in the validation group

Supplementary Figure 4. The decision curve analysis of prognostic model
Table 1 (on next page)

Table 1. The clinical features of lung adenocarcinoma patients in the model group and validation group.
Table 1. The clinical features of lung adenocarcinoma patients in the model group and validation group

|                                      | Model group (n=494) | Validation group(n=233) | P value |
|--------------------------------------|---------------------|--------------------------|---------|
| Death [n(%)]                         | 182(36.8)           | 127(54.5)                | <0.001  |
| Survival time for living patients(mean ± SD, month) | 22.2(15.6,37.6)     | 69.1(59.3,88.0)          | <0.001  |
| Survival time for dead patients (mean ± SD, month) | 20.3(9.8,36.8)      | 25.3(12.4,48.4)          | 0.006   |
| Age (mean ± SD, year)                | 65.3±10.0           | 66.1±9.9                 | 0.318   |
| Male [n(%)]                          | 229(46.4)           | 111(47.6)                | 0.746   |
| AJCC Stage (IV)                      | 25                  | 4                        | <0.001  |
| AJCC Stage (III)                     | 80                  | 13                       |         |
| AJCC Stage (II)                      | 116                 | 54                       |         |
| AJCC Stage (I)                       | 273                 | 162                      |         |
| AJCC Stage (NA)                      | 0                   | 0                        |         |
| AJCC PT (T1)                         | 229                 | NA                       | NA      |
| AJCC PT (T0)                         | 265                 | NA                       | NA      |
| AJCC PT (NA)                         | 0                   | NA                       | NA      |
| AJCC PN (N4)                         | 18                  | NA                       | NA      |
| AJCC PN (N3)                         | 44                  | NA                       | NA      |
| AJCC PN (N2)                         | 264                 | NA                       | NA      |
| AJCC PN (N1)                         | 165                 | NA                       | NA      |
| AJCC PN (N0)                         | 3                   | NA                       | NA      |
| AJCC PN (NA)                         | 0                   | NA                       | NA      |
| AJCC PM (M3)                         | 2                   | NA                       | NA      |
| AJCC PM (M2)                         | 69                  | NA                       | NA      |
| AJCC PM (M1)                         | 92                  | NA                       | NA      |
| AJCC PM (M0)                         | 302                 | NA                       | NA      |
| AJCC PM (NA)                         | 11                  | NA                       | NA      |

Note: Continuous variables were compared by t-test or Mann-Whitney U test as appropriate; Categorical variables were compared by chi-squared test or Fisher’s exact test as appropriate; NA, missing data; SD: standard deviation; AJCC: American Joint Committee on Cancer.
Table 2. The model information of prognostic mRNAs in univariate and multivariable Cox regression analyses
Table 2. The model information of prognostic mRNAs in univariate and multivariable Cox regression analyses

| Gene                 | Univariate analysis | Multivariate analysis |
|----------------------|---------------------|-----------------------|
|                      | HR                  | 95% CI                | P-value | coefficient | HR     | 95% CI   | P-value |
| DNAJC27 (High/Low)   | 0.594               | 0.441-0.800           | <0.001  | -0.731      | 0.482  | 0.336-0.691 | <0.001 |
| NPAS2 (High/Low)     | 1.613               | 1.202-2.164           | <0.001  | 0.478       | 1.612  | 1.166-2.229 | 0.004  |
| PHKA1 (High/Low)     | 1.367               | 1.020-1.832           | 0.036   | 0.394       | 1.483  | 1.032-2.131 | 0.033  |
| CDADC1 (High/Low)    | 0.699               | 0.521-0.936           | 0.016   | -0.554      | 0.575  | 0.394-0.839 | 0.004  |
| PTGFRN (High/Low)    | 1.356               | 1.012-1.818           | 0.042   | 0.479       | 1.615  | 1.120-2.329 | 0.010  |
| DDIT4 (High/Low)     | 1.634               | 1.217-2.193           | <0.001  | 0.398       | 1.489  | 1.086-2.042 | 0.013  |
| SCAMP5 (High/Low)    | 0.737               | 0.550-0.988           | 0.041   | -0.413      | 0.662  | 0.474-0.924 | 0.015  |
| SPRY2 (High/Low)     | 0.714               | 0.533-0.957           | 0.024   | -0.437      | 0.646  | 0.455-0.918 | 0.015  |
| CSE1L (High/Low)     | 1.342               | 1.001-1.798           | 0.049   | 0.385       | 1.470  | 1.024-2.110 | 0.037  |
| ELAVL4 (High/Low)    | 0.707               | 0.527-0.950           | 0.021   | -0.394      | 0.674  | 0.496-0.918 | 0.012  |
| TRIM29 (High/Low)    | 1.391               | 1.037-1.864           | 0.028   | 0.329       | 1.390  | 1.010-1.912 | 0.043  |
| LPGAT1 (High/Low)    | 1.410               | 1.052-1.890           | 0.022   | 0.369       | 1.446  | 1.017-2.054 | 0.040  |
| TRPC3 (High/Low)     | 0.681               | 0.506-0.915           | 0.011   | -0.328      | 0.720  | 0.525-0.988 | 0.042  |
| DCBLD2 (High/Low)    | 1.346               | 1.005-1.802           | 0.046   | 0.326       | 1.385  | 1.001-1.915 | 0.049  |

Note: HR, hazard ratio; CI, confidence interval. The medians of mRNA expression values were used as cut-off values to stratify mRNA expression values into high expression group (as value 1) and low expression group (as value 0).
Table 3 (on next page)

Table 3. Results of differential expression analysis of the enrolled mRNAs
Table 3. Results of differential expression analysis of the enrolled mRNAs

| Symbol   | F           | P value | FC  | logCPM | FDR  |
|----------|-------------|---------|-----|--------|------|
| PHKA1    | 207.115     | <0.001  | 3.252 | 4.717  | <0.001 |
| CDADC1   | 125.913     | <0.001  | 0.666 | 2.854  | <0.001 |
| LPGAT1   | 116.687     | <0.001  | 2.218 | 6.843  | <0.001 |
| PTGFRN   | 104.576     | <0.001  | 2.304 | 6.604  | <0.001 |
| DNAJC27  | 95.382      | <0.001  | 0.621 | 2.815  | <0.001 |
| SCAMP5   | 66.600      | <0.001  | 2.384 | 4.159  | <0.001 |
| CSE1L    | 63.349      | <0.001  | 1.694 | 6.671  | <0.001 |
| NPAS2    | 52.396      | <0.001  | 2.416 | 4.793  | <0.001 |
| TRPC3    | 44.057      | <0.001  | 0.429 | -0.488 | <0.001 |
| SPRY2    | 42.334      | <0.001  | 0.629 | 5.171  | <0.001 |
| ELAVL4   | 41.979      | <0.001  | 3.749 | -0.076 | <0.001 |
| DDIT4    | 34.794      | <0.001  | 2.249 | 7.135  | <0.001 |
| DCBLD2   | 30.157      | <0.001  | 2.262 | 7.216  | <0.001 |
| TRIM29   | 7.324       | 0.007   | 2.108 | 5.400  | 0.010 |
Table 4 (on next page)

Table 4. Univariate and multivariable Cox regression analyses for independence assessment of prognostic model

Table 4. Univariate and multivariable Cox regression analyses for independence assessment of prognostic model
Table 4. Univariate and multivariable Cox regression analyses for independence assessment of prognostic model

|                        | Univariate analysis | Multivariate analysis |     |
|------------------------|---------------------|-----------------------|-----|
|                        | HR                  | 95% CI                | P-value | coefficient | HR | 95% CI | P-value |
| Model group (n=494)    |                     |                       |       |             |    |        |         |
| Age(≥60/<60 years)     | 1.003               | 0.722-1.392           | 0.987 | 0.078       | 1.062 | 0.776-1.507 | 0.643 |
| Gender (Male/Female)   | 1.048               | 0.783-1.403           | 0.754 | 0.077       | 1.080 | 0.805-1.449 | 0.609 |
| AJCC stage (IV,III/II,I) | 2.599             | 1.906-3.545           | <0.001 | 0.800       | 2.225 | 1.622-3.051 | <0.001 |
| Prognostic model (High/Low) | 2.705         | 1.986-3.684           | <0.001 | 0.895       | 2.447 | 1.786-3.352 | <0.001 |
| Validation group(n=233)|                     |                       |       |             |    |        |         |
| Age(≥60/<60 years)     | 0.876               | 0.595-1.289           | 0.501 | -0.271      | 0.762 | 0.512-1.136 | 0.183 |
| Gender (Male/Female)   | 1.308               | 0.921-1.857           | 0.134 | 0.268       | 1.308 | 0.911-1.877 | 0.146 |
| AJCC stage (IV,III/II,I) | 0.988             | 0.517-1.888           | 0.974 | 0.354       | 1.424 | 0.738-2.750 | 0.292 |
| Prognostic model (High/Low) | 5.109           | 3.428-7.348           | <0.001 | 1.651       | 5.214 | 3.531-7.701 | <0.001 |

Note: AJCC, the American Joint Committee on Cancer; HR, hazard ratio; CI, confidence interval. The median of Prognostic model scores was used as the cut-off value to stratify gastric cancer patients into high risk group and low risk group.
Figure 1

Figure 1. Competitive endogenous RNA network chart: Red triangles represented 22 lncRNAs; Blue triangles represented 29 miRNAs; green circles represented 73 mRNAs.
Figure 2

Figure 2. (A) The barplot of GO terms for prognostic mRNAs; (B) The dotplot of GO terms for prognostic mRNAs
Figure 3

Figure 3. (A) Chord chart of GO terms for prognostic mRNAs; (B) KEGG pathways for prognostic mRNAs.
Figure 4

Figure 4. The prognostic nomogram for overall survival.
Figure 5

Figure 5. (A) Survival curves in model group; (B) Time-dependent receiver operating characteristic curves in model group; (C) The distribution of prognostic model score in model group; (D) The overall survival status and overall survival time in model group.
Figure 6

Figure 6. (A) Survival curves in validation group; (B) Time-dependent receiver operating characteristic curves in validation group.
Figure 7

Figure 7. Home page of Smart Cancer Survival Predictive System: (A) Individual mortality risk predictive curve display page; (B) Different time-point individual mortality risk prediction display page.
Figure 8

Figure 8. Home page of Gene Survival Analysis Screen System: (A) Survival curve plot display page; (B) Univariate analysis table display page.
Step 1: Select dataset
Using testing dataset (.xlsx) or upload a new dataset (.xlsx)?

Step 2: Select subgroup
Subgroup

Patient gender

Tumor stage

Download testing dataset (.csv)

Kaplan-Meier survival curve
Survival status variable
OS

Survival time variable
OS_MONTH

Define time unit

Step 1: Select dataset
Using testing dataset (.xlsx) or upload a new dataset (.xlsx)?

Step 2: Select subgroup
Subgroup

Patient gender

Tumor stage

Download testing dataset (.csv)

Univariate survival analysis table
Show 10 entries

Variable Category Number Hazard Ratio (95% confidence interval)
TSPAN6 1 362 0.85 (0.86-1.063)

Number

0 364

Age Low 363 0.647 (0.678-1.059)

High 363

Gender 1 339 1.156 (0.924-1.444)

0 387

Stage 2 170 2.099 (1.619-2.733)

3 93 2.653 (1.94-3.628)

4 29 2.364 (1.429-3.91)

1 434

Showing 1 to 10 of 10 entries

PeerJ reviewing PDF | (2020:08:52379:2:0:NEW 14 Apr 2021)