



# A Novel Risk-Score Model With Eight MiRNA Signatures for Overall Survival of Patients With Lung Adenocarcinoma

Jun Wu<sup>1†</sup>, Yuqing Lou<sup>2†</sup>, Yi-Min Ma<sup>1</sup>, Jun Xu<sup>3</sup> and Tielu Shi<sup>1,4\*</sup>

<sup>1</sup>Center for Bioinformatics and Computational Biology, And the Institute of Biomedical Sciences, School of Life Sciences, East China Normal University, Shanghai, China, <sup>2</sup>Department of Pulmonary Medicine, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China, <sup>3</sup>Department of Emergency Medicine, The First Hospital of Anhui Medical University, Hefei, China, <sup>4</sup>Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University and Capital Medical University, Beijing, China

## OPEN ACCESS

### Edited by:

Wen-Lian Chen,  
Shanghai University of Traditional  
Chinese Medicine, China

### Reviewed by:

Fan Yang,  
Jiangxi Science and Technology  
Normal University, China  
Rongzhong Huang,  
Second Affiliated Hospital of  
Chongqing Medical University, China

### \*Correspondence:

Tielu Shi  
tielushi@yahoo.com

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Human and Medical Genomics,  
a section of the journal  
Frontiers in Genetics

Received: 14 July 2021

Accepted: 08 October 2021

Published: 12 November 2021

### Citation:

Wu J, Lou Y,  
Ma Y-M Xu J and Shi T (2021) A Novel  
Risk-Score Model With Eight MiRNA  
Signatures for Overall Survival of  
Patients With Lung Adenocarcinoma.  
Front. Genet. 12:741112.  
doi: 10.3389/fgene.2021.741112

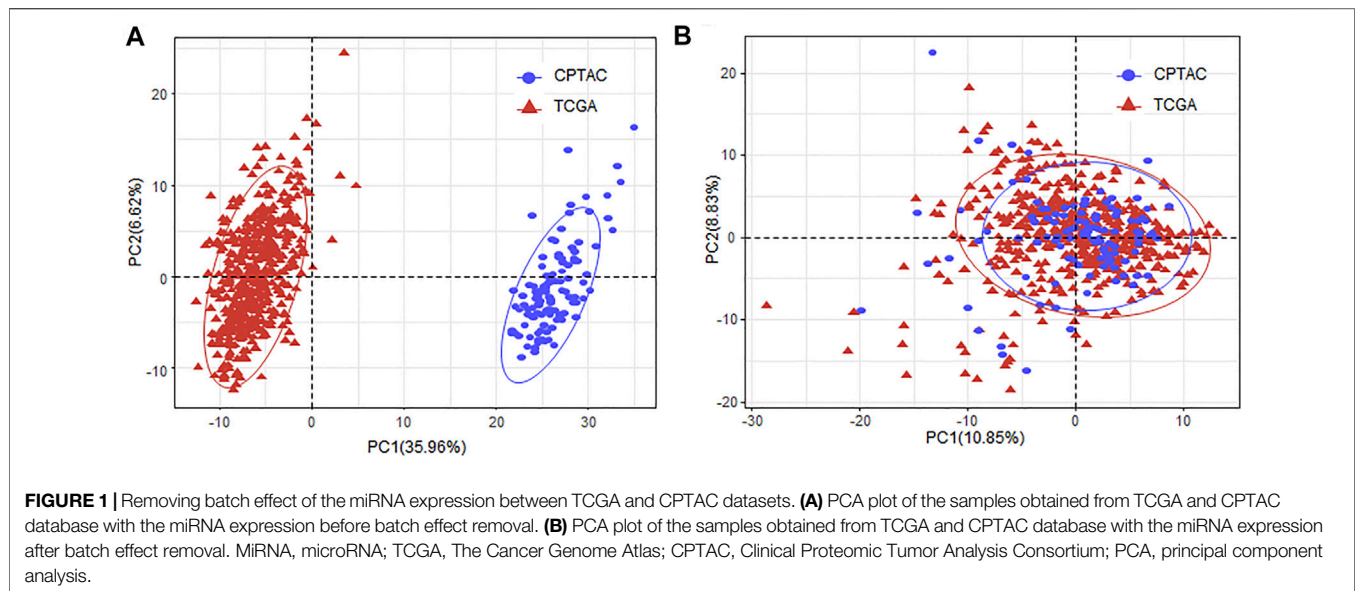
Lung adenocarcinoma (LUAD) is the most common subtype of lung cancer with heterogeneous outcomes and diverse therapeutic responses. To classify patients into different groups and facilitate the suitable therapeutic strategy, we first selected eight microRNA (miRNA) signatures in The Cancer Genome Atlas (TCGA)-LUAD cohort based on multi-strategy combination, including differential expression analysis, regulatory relationship, univariate survival analysis, importance clustering, and multivariate combinations analysis. Using the eight miRNA signatures, we further built novel risk scores based on the predefined cutoff and beta coefficients and divided the patients into high-risk and low-risk groups with significantly different overall survival time ( $p$ -value  $< 2 \times 10^{-16}$ ). The risk-score model was confirmed with an independent dataset ( $p$ -value =  $4.71 \times 10^{-4}$ ). We also observed that the risk scores of early-stage patients were significantly lower than those of late-stage patients. Moreover, our model can also provide new insights into the current clinical staging system and can be regarded as an alternative system for patient stratification. This model unified the variable value as the beta coefficient facilitating the integration of biomarkers obtained from different omics data.

**Keywords:** lung adenocarcinoma, microRNA signature, risk-score model, overall survival time, treatment response

## INTRODUCTION

Lung cancer, which is one of the most common and severe types of cancer, remains the leading cause of cancer incidence and mortality worldwide in both males and females (Siegel et al., 2019). Lung adenocarcinoma (LUAD) is the most prevalent histological subtype of lung cancer, with an increasing incidence over the past few decades (Ferlay et al., 2010). The traditional clinical staging system for LUAD, which is based on anatomical information, appears to be inadequate for prognosis evaluation or treatment choices now due to the heterogeneity among patients.

With the rapid advance of molecular biology, many diagnostic and prognostic biomarkers have been identified for various cancers (Wang et al., 2017a; Wang et al., 2017b; Cheng et al., 2019; Huang et al., 2020a; Sheng et al., 2020). With the use of these biomarkers, the traditional tumor classes can be further divided into new subtypes, which may benefit from different therapeutic strategies (Li et al., 2019; Sherafatian and Arjmand, 2019; Lathwal et al., 2020). Besides that, most targeted agents (e.g.,



cetuximab, gefitinib, and tamoxifen) are effectively only if their respective targets are mutated or differentially expressed (Sun et al., 2017; Yang et al., 2020).

MicroRNAs (miRNAs) are small non-protein-coding RNAs, which can negatively regulate gene expression by binding to their selective messenger RNAs (mRNAs), thereby influencing various biological progresses, such as cellular differentiation, cell-cycle control, and apoptosis (Bentwich, 2005; Cheng et al., 2005; Novello et al., 2013). MiRNAs are reported to be differentially expressed in various human cancers and act as both tumor suppressors and oncogenes (Volinia et al., 2006; Cui et al., 2020). For some certain types of cancer, the miRNAs are proved to be more effective in cancer classification than mRNAs (Miska, 2007), and the miRNAs are also used as signatures for prognosis prediction. Yu et al. identified five miRNAs significantly associated with patient relapse and survival based on 117 non-small cell lung cancer (NSCLC) patients (Yu et al., 2008). Li et al. also identified eight miRNAs as signatures for survival prediction in LUAD (Li et al., 2014). Similarly, Hess et al. provided a five-miRNA signature, which is a strong and independent prognostic factor for disease recurrence and survival of patients with HPV-negative head and neck squamous cell carcinoma (HNSCC) (Hess et al., 2019). All these results showed that miRNAs are powerful potential signatures for prognosis prediction. However, there were very few overlaps between these miRNA signatures identified by different groups. Moreover, most studies just focused on the miRNA or mRNA expression level independently and ignored the negatively regulative relationship between miRNAs and mRNAs.

In this study, based on the miRNA expression, gene expression profiles and clinical information of 516 LUAD samples from The Cancer Genome Atlas (TCGA) (The Cancer Genome Atlas Research Network, 2014), we built the miRNA–gene negative regulation pairs to ensure that the candidate miRNAs influence biological progress of these samples. Then, we screened

eight miRNA signatures through differential expression analysis, regulatory relationship filtering, univariate survival analysis, importance clustering, and multivariate combination selection. Based on the eight miRNA signatures, we built a risk-score model to group the patients as high-risk and low-risk. The model performance was further proved using an independent dataset. We demonstrated that the model can also be used for stratification of patients in the same tumor stage.

## RESULTS

### Data Collection

The gene expression, miRNA expression, and clinical data of TCGA-LUAD were download from UCSC Xena (<http://xena.ucsc.edu>) (Goldman et al., 2017). Besides that, we also downloaded the miRNA expression and related clinical data of LUAD from the Clinical Proteomic Tumor Analysis Consortium (CPTAC)-3 database (Edwards et al., 2015) using the R/Biconductor package “TCGAbiolinks” as the independent validation data (Colaprico et al., 2016; Mounir et al., 2019). Only the primary solid tumor (TP) and solid tissue normal (NT) samples were selected. Patients with less than 30 days of overall survival (OS) were excluded to avoid the possible unrelated causes of death. The details of the samples are shown in **Table 1**.

As the miRNA expression was obtained from different databases, we applied ComBat (Leek et al., 2012) to remove the batch effect (**Figures 1A,B**).

### Differential Gene Expression Analysis

The count data of gene expression were used to perform the differential expression analysis. The genes with adjusted  $p$ -value of less than  $1 \times 10^{-3}$  and absolute  $\log_2$  fold change  $\geq 1$  were regarded as significantly differentially expressed. As a result, a total of 4,522 (64.11%) upregulated and 2,531 (35.89%) downregulated genes

**TABLE 1** | Number of samples obtained from different databases.

TCGA-LUAD				CPTAC-LUAD	
Gene expression		MiRNA expression		MiRNA expression	
TP	NT	TP	NT	TP	NT
510	58	510	45	111	102

Note. OS, overall survival; TCGA, The Cancer Genome Atlas; LUAD, lung adenocarcinoma; CPTAC, Clinical Proteomic Tumor Analysis Consortium; miRNA, microRNA; TP, primary solid tumor; NT, solid tissue normal.

(Figure 2A). The Gene Ontology (GO) term and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis results showed that these differentially expressed genes (DEGs) were enriched in 842 biological processes (BPs), 161 molecular functions (MFs), 137 cellular components (CCs), and 44 KEGG pathways (Figure 2B; Supplement Table S1).

## MicroRNA Signature Identification Based on Multi-Strategy

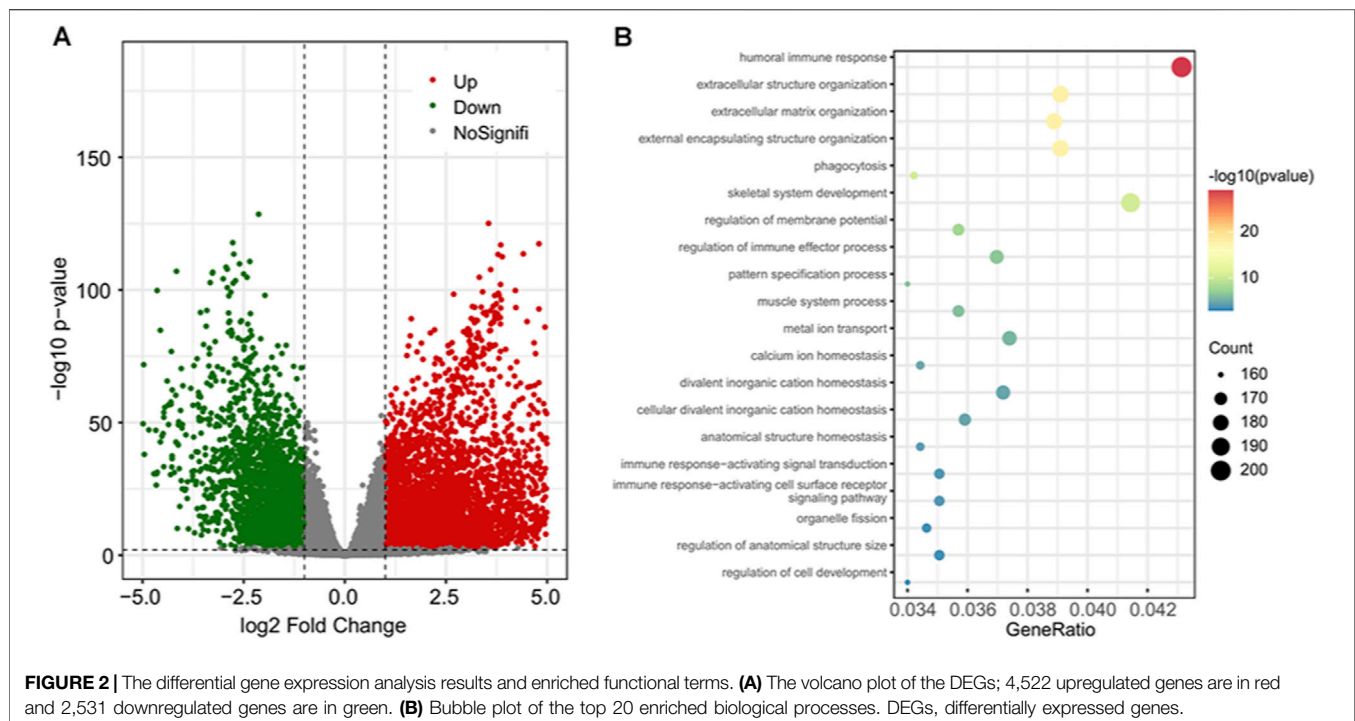
Using the negative regulation criterion and the information retrieved from three verified miRNA-target databases, we obtained 2,284 miRNA-gene pairs consisting of 228 miRNAs and 1,199 target genes. To examine the function term and effects of these miRNA regulators, we performed GO term and pathway enrichment analysis for these 1,199 target genes. The results showed that there were 924 genes functionally enriched in 700 BPs, 30 MFs, and 53 CCs with adjusted  $p$ -value of less than 0.05 (Supplement Table S2). Additionally,

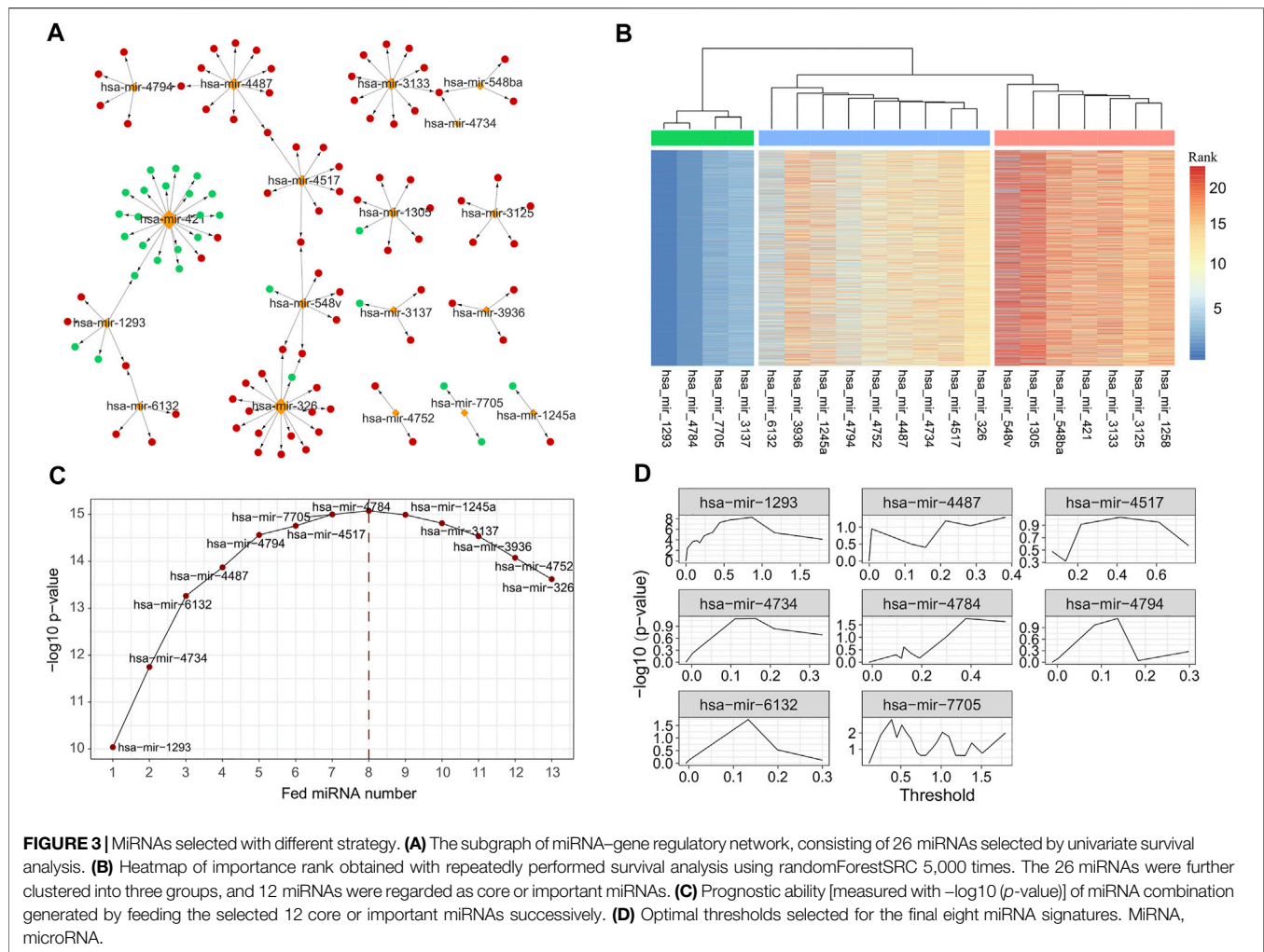
there were 163 genes enriched in 16 KEGG pathways, such as cell cycle, cellular senescence, and p53 signaling pathway (Supplement Table S2). By limiting the target genes as these functional enriched genes, we simplified the miRNA-gene regulation network consisting of 221 miRNA and 924 genes (Figure 3A).

We next performed the univariate survival analysis using the Cox proportional-hazards model with the 161 miRNA regulators. The results showed that 20 miRNAs of LUAD patients can be divided into two groups with significantly different OS (adjusted  $p$ -value of less than 0.05, Supplement Table S3). To further ensure the robustness of these miRNAs, we repeatedly performed survival analysis using randomForestSRC 5,000 times and measured the importance of the 21 miRNAs accordingly. With the variable importance rank matrix (see Methods), we clustered the 21 miRNAs into three groups using hierarchical cluster analysis (Figure 3B), and 13 miRNAs that ranked top in most of the repeats were selected for the downstream analysis.

To further select the optimal combination of the miRNA signatures, we performed multivariate survival analysis by adding the 13 miRNAs into the Cox regression model using greedy strategy (Figure 3C). By doing so, we observed that when the number of the miRNA signatures reached eight, the performance was no longer improved. Thus, we selected eight miRNAs (hsa-mir-1293, hsa-mir-4734, hsa-mir-6132, hsa-mir-4487, hsa-mir-4794, hsa-mir-4517, hsa-mir-7705, and hsa-mir-4784) as the miRNA signatures to build the risk-score prediction model.

For each of the miRNA signatures, we divided LUAD patients into two groups according to the miRNA expression with different thresholds and evaluated the discrimination validity





using log-rank test and Kaplan–Meier test (Figure 3D). The optimal threshold and the  $\beta$  coefficients for each miRNA signature were saved for the model building (see *Methods*).

## Performance Evaluation for the Risk-Score Model

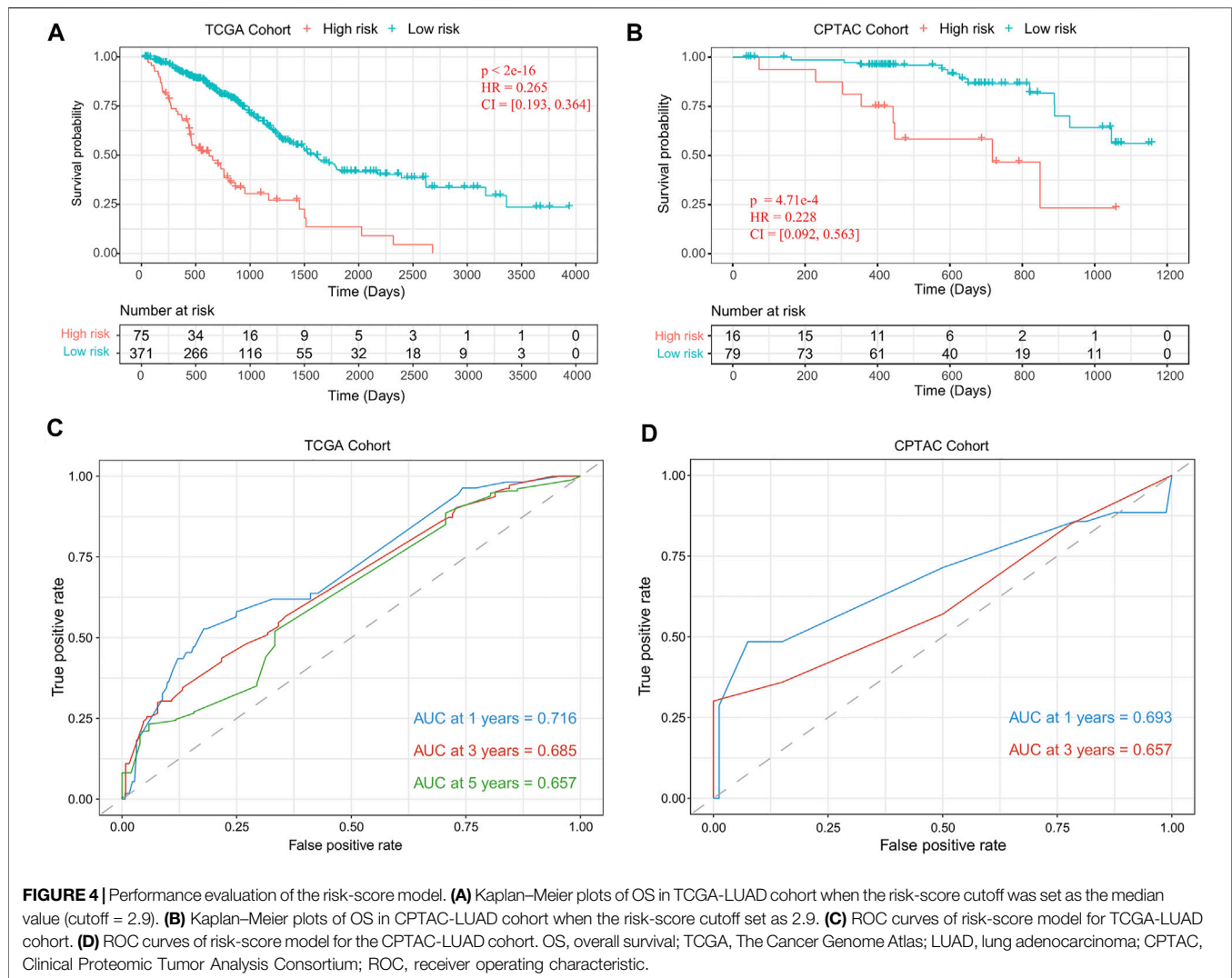
Using the risk-score model, we estimated the risk score for each LUAD patient and divided the LUAD cohort into high-risk and low-risk groups by defining the cutoff as the median risk score (cutoff = 2.9). The Kaplan–Meier survival analysis results showed that the OS time was significantly different between the patients in these two groups ( $p\text{-value} = 1.43 \times 10^{-18}$ , Figure 4A). We also evaluated the performance with the independent validation dataset (CPTAC-LUAD). The risk score of the patient in the CPTAC-LUAD dataset were estimated, and then the CPTAC-LUAD patients were divided into high-risk and low-risk groups with the cutoff determined by TCGA-LUAD dataset. The Kaplan–Meier survival analysis results showed that the OS time was significantly different between the CPTAC-LUAD patients in these two groups ( $p\text{-value} = 4.71 \times 10^{-4}$ , Figure 4B).

To further assess the prognostic power of proposed method, time-dependent receiver operating characteristic (ROC) curves were used to compare the specificity and sensitivity for the predicted results of TCGA-LUAD cohort (1 year, 0.716; 3 years, 0.685; 5 years, 0.657; Figure 4C) and CPTAC-LUAD cohort (1 year, 0.693; 3 years, 0.657; Figure 4D). The ROC curves and area under the ROC curve (AUC) showed high consistency of this risk-score model.

## The Prognostic Ability of the Risk-Score Model Within Different Clinical Groups

To further validate the prognostic ability of the risk-score model, we test the enrichment of low- and high-risk patients in the groups divided by different clinical indicators, such as age, gender, and clinical stages (Stages I–IV). We found that there was no significant difference of the risk score between the male and female patients ( $p\text{-value} = 0.133$ ), and the risk score also did not show significant correlation with the patient age ( $R = -0.079$ ,  $p\text{-value} = 0.1$ , Figure 5A). For the clinical stages, we found that the risk score of patients in Stage II and Stage III were significantly higher than that of patients in Stage I (Stage II:  $p\text{-value} = 1.2 \times 10^{-5}$ ,





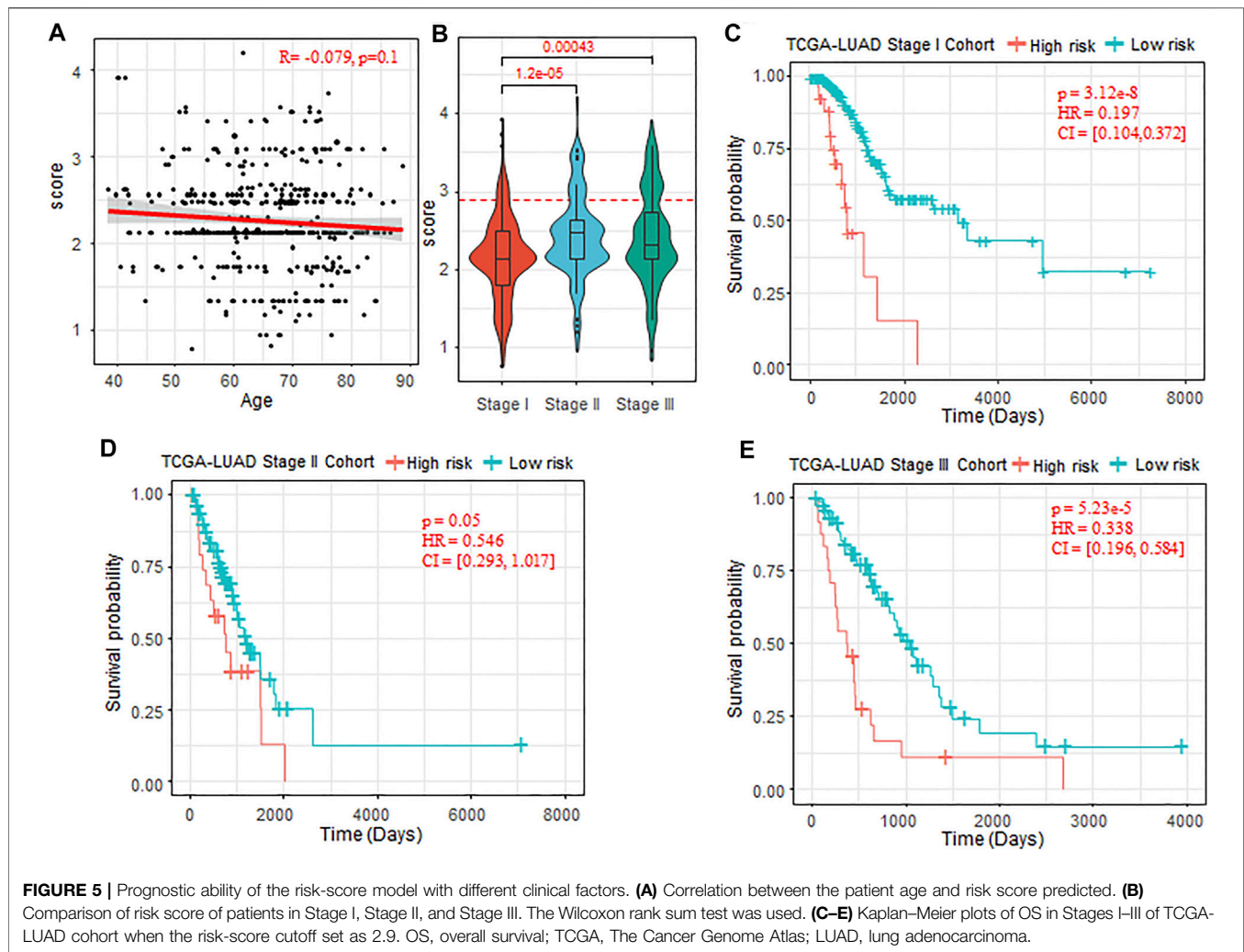
Stage III:  $p$ -value =  $4.3 \times 10^{-4}$ , **Figure 5B**). The low-risk patients were significantly enriched in early stage (Wilcoxon rank sum test  $p$ -value <  $2.2 \times 10^{-16}$ ). The clinical staging system is the most acknowledged clinicopathological factor for prognostication and therapy determination of LUAD, which are limited because the prognoses within the same clinical stage vary widely (Mlecnik et al., 2011). To further investigate the potentiality of the risk-score model, we tested the difference of OS between the low- and high-risk patients within the same clinical stage. The results showed that, for Stage I, Stage II, and Stage III, OS time was significantly shorter in the high-risk cohort compared with the low-risk cohort (Stage I,  $p$ -value =  $3.12 \times 10^{-8}$ ; Stage II,  $p$ -value = 0.05; Stage III,  $p$ -value =  $5.23 \times 10^{-5}$ ; **Figures 5C–E**).

## Treatment Response for the Groups Divided by the Risk-Score Model

To further evaluate the clinical benefit of the risk-score model, we extracted the treatment information for the LUAD

patients, and 155 patients received different types treatment and 297 patients without any treatment information. Patients who received more than two types of therapy (e.g., patients received both chemotherapy and immunotherapy) were excluded for the follow-up analysis. As the patients who received chemotherapy were enriched in Stage II–Stage IV (Fisher's exact test  $p$ -value =  $2.87 \times 10^{-24}$ ), we test the effectiveness of the chemotherapy on the patients in Stage II–Stage IV. The results showed that chemotherapy can improve prognosis to some extent ( $p$ -value = 0.09, **Figure 6A**).

We also observed that, in all the patients who received chemotherapy, the patients regarded as low-risk also benefited more from the chemotherapy than the high-risk chemotherapy ( $p$ -value =  $1.5 \times 10^{-4}$ , **Figure 6B**). In chemotherapy drugs specifically, we also observed that carboplatin can significantly prolong the OS of low-risk patients ( $p$ -value = 0.02, **Figure 5C**), but it has no benefit in the high-risk patients ( $p$ -value = 0.94, **Figure 5D**).



## METHODS

### Data Preprocessing

The quantile normalization procedure is applied to the gene and miRNA expression separately and filter out the genes and miRNAs with the expression value 0 across more than 90% of the samples. We also applied the ComBat (Leek et al., 2012) to remove the batch effect between the data in TCGA dataset and CPTAC dataset. The DESeq2 (Love et al., 2014) was used to perform the differential expression analysis between the tumor and normal samples using the raw count data. Genes with Benjamini and Hochberg adjusted  $p$ -value of less than  $1 \times 10^{-3}$  and fold change larger than 2 were regarded as significantly DEGs.

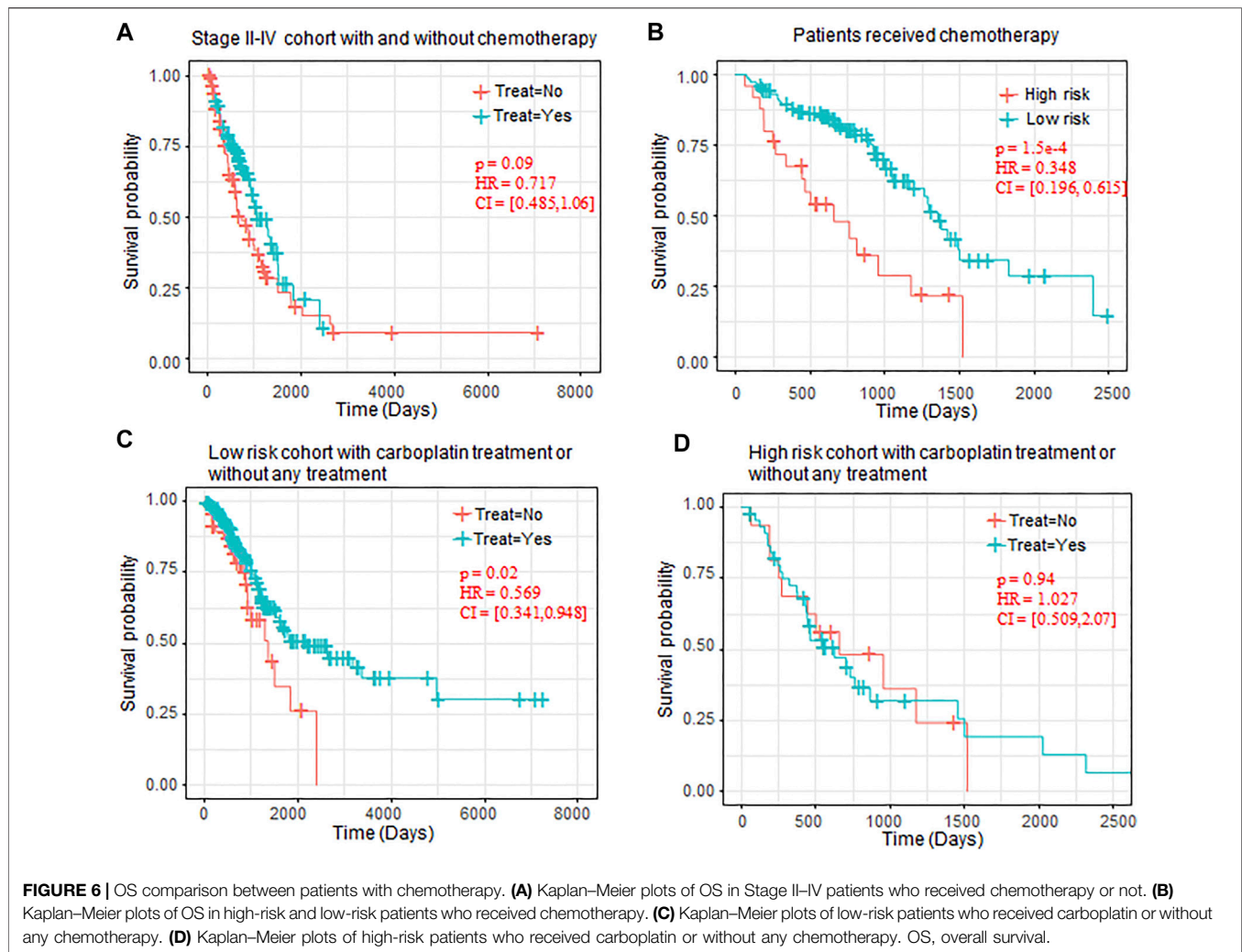
### Building the MicroRNA–Messenger RNA Negative Regulation Pairs

To obtain the relationship between miRNA and their target gene (mRNAs), we extracted the regulator factor miRNA of DEGs

from three verified miRNA–target databases (miRecords (Xiao et al., 2009), miRTarBase (Huang et al., 2020b), and TarBase (Karagkouni et al., 2018)) using the “multiMiR” R package (Ru et al., 2014). These regulatory relationships were further refined based on the negative regulated relationship that one miRNA and its target genes were negatively related. Spearman’s correlation test was applied to each miRNA–gene pair among 504 TP samples with both miRNA expression value and mRNA expression value available, and only the pairs with negative correlation coefficient and adjusted  $p$ -value  $< 0.01$  remained.

### MicroRNA Signature Selection

The procedure takes four steps to accomplish the miRNA signature selection. We first performed the functional enrichment analysis for the DEGs using the R/Biconductor package “clusterProfiler” (Yu et al., 2012), and functional terms with adjusted  $p$ -value of less than 0.05 were regarded as significantly enriched. We retained the miRNAs targeting the genes enriched in any functional terms. Next, we performed OS analysis for each of the remaining miRNAs, and the miRNAs with log-rank  $p$ -value of less than 0.05



remained. To further refine the miRNA signatures, we evaluated the extent to which each miRNA contributes to predicting survival using the metric of variable importance using the `vimp` function from the R package “`randomForestSRC`” (Ishwaran et al., 2020). We calculated variable importance using random permutation of the variable approach. To ensure robustness, we repeated this step 5,000 times, and a rank matrix for the miRNAs was obtained based on the calculated variable importance. Using the rank matrix, we divided these miRNAs into three groups (including important miRNAs, secondary miRNA, and meaningless miRNAs) using R function `hclust` with the default parameters. The miRNAs regarded as important or secondary were selected as candidate miRNA signatures and ranked according to the median of the 5,000 ranks of the miRNA. Finally, we performed the multivariate survival analysis using the Cox regression model by feeding the candidate miRNA signatures in sequence. The miRNAs that reduced the prognostic ability of the model were excluded. Ultimately, the rest of the miRNAs were regarded as the signatures.

## Building Risk-Score Estimator

For each miRNA signature, we calculated the optimal threshold that can divide the patients into the high-risk or low-risk group with the most significant OS time difference, and the beta ( $\beta$ ) coefficient for each miRNA signature was also calculated with the optimal threshold. The risk score of a patient can be defined as follows:

$$\text{Risk score} = \sum_i s_i$$

and  $s_i$  represents the risk score for a certain miRNA  $i$ , which was calculated as follows:

$$s_i = \begin{cases} |\beta_i|, & \text{if } \beta_i < 0 \text{ and miRNA expression lower than the related optimal threshold} \\ \beta_i, & \text{if } \beta_i > 0 \text{ and miRNA expression higher than the related optimal threshold} \\ 0, & \text{else} \end{cases}$$

## Statistical Analysis

Time-dependent ROC curve and AUC were generated with R package “`timeROC`” (Blanche, 2015). Survival analysis and

univariate and multivariate Cox regression analyses were performed with R package “survival” (Therneau and Lumley, 2010). The Kaplan–Meier curves were plot with R package “survminer” (Kassambara et al., 2017). Heatmap was drawn with R package “pheatmap” (Kolde and Kolde, 2015). The *p*-values of each variable were corrected using the Benjamini and Hochberg (BH) method (Benjamini and Hochberg, 1995).

## DISCUSSION

In this study, we have identified eight miRNA signatures associated with the OS of LUAD using both the miRNA expression and gene expression profiles obtained from TCGA-LUAD dataset. With these miRNA signatures, we built a novel risk-score model using both the optimal cutoff and corresponding beta coefficients; otherwise, the miRNA expression is used directly. This model divides LUAD patients into two groups (high-risk and low-risk) with significantly different OS times. The performance was proved to be consistent in both the training set (TCGA-LUAD) and independent validation set (CPTAC-LUAD).

Through consulting literature materials, we found that all the eight miRNAs were reported to be associated with various types of cancer, including lung cancer. Additionally, personalized cancer medicine is a clinical approach that strives to customize therapies based upon the genetic profiles of individual patient tumors. Our results further proved that stratification of LUAD patients is also important to the treatment and response to therapy. However, we also noted that the clinical information, such as treatment response, in TCGA database is mainly rough, and the results in this study need further investigation in the future.

Most importantly, as built based on the optimal threshold and corresponding beta coefficients, the proposed risk-score model was fit for different types of data, including both qualitative and quantitative. This risk-score model provided a new insight into the multi-omics data integration for prognosis.

## REFERENCES

- Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodological)* 57, 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x
- Bentwich, I. (2005). A Postulated Role for microRNA in Cellular Differentiation. *FASEB j.* 19, 875–879. doi:10.1096/fj.04-3609hyp
- Blanche, P., Dartigues, J. F., and Jacqmin-Gadda, H. (2013). Estimating and Comparing Time-Dependent Areas Under Receiver Operating Characteristic Curves for Censored Event Times with Competing Risks. *Stat. Med.* 32, 5381–5397.
- Cheng, A. M., Byrom, M. W., Shelton, J., and Ford, L. P. (2005). Antisense Inhibition of Human miRNAs and Indications for an Involvement of miRNA in Cell Growth and Apoptosis. *Nucleic Acids Res.* 33, 1290–1297. doi:10.1093/nar/gki200
- Cheng, Y., Wang, K., Geng, L., Sun, J., Xu, W., Liu, D., et al. (2019). Identification of Candidate Diagnostic and Prognostic Biomarkers for Pancreatic Carcinoma. *Ebiomedicine* 40, 382–393. doi:10.1016/j.ebiom.2019.01.003
- Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., et al. (2016). TCGAAbiolinks: an R/Bioconductor Package for Integrative Analysis of TCGA Data. *Nucleic Acids Res.* 44, e71. doi:10.1093/nar/gkv1507
- Cui, X., Liu, Y., Sun, W., Ding, J., Bo, X., and Wang, H. (2020). Comprehensive Analysis of miRNA-Gene Regulatory Network with Clinical Significance in Human Cancers. *Sci. China Life Sci.* 63, 1201–1212. doi:10.1007/s11427-019-9667-0
- Edwards, N. J., Oberti, M., Thangudu, R. R., Cai, S., McGarvey, P. B., Jacob, S., et al. (2015). The CPTAC Data Portal: A Resource for Cancer Proteomics Research. *J. Proteome Res.* 14, 2707–2713. doi:10.1021/pr501254j
- Ferlay, J., Shin, H.-R., Bray, F., Forman, D., Mathers, C., and Parkin, D. M. (2010). Estimates of Worldwide burden of Cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* 127, 2893–2917. doi:10.1002/ijc.25516

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

## ETHICS STATEMENT

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

## AUTHOR CONTRIBUTIONS

JW, YL, and TS conceived the study. JW and Y-MM performed the algorithm development and downstream bioinformatics analysis. JW and YL wrote the manuscript. JX and TS revised the manuscript. All authors read and approved the final manuscript.

## FUNDING

This work was supported by the National Natural Science Foundation of China grants (Nos. 31801118, 31671377), Shanghai Municipal Science and Technology Major Project (Grant No. 2017SHZDZX01), Beihang University and Capital Medical University Plan (BHME-201904), the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, ECNU, and the Nurture projects for basic research of Shanghai Chest Hospital (No. 2020YNJCM06).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.741112/full#supplementary-material>



- Goldman, M., Craft, B., Zhu, J. C., and Haussler, D. (2017). The UCSC Xena System for Cancer Genomics Data Visualization and Interpretation. *Cancer Res.* 77, 2584. doi:10.1158/1538-7445.AM2017-2584.
- Hess, J., Unger, K., Maihofer, C., Schüttrumpf, L., Wintergerst, L., Heider, T., et al. (2019). A Five-MicroRNA Signature Predicts Survival and Disease Control of Patients with Head and Neck Cancer Negative for HPV Infection. *Clin. Cancer Res.* 25, 1505–1516. doi:10.1158/1078-0432.ccr-18-0776
- Huang, H. Y., Lin, Y. C., Li, J., Huang, K. Y., Shrestha, S., Hong, H. C., et al. (2020). miRTarBase 2020: Updates to the Experimentally Validated microRNA-Target Interaction Database. *Nucleic Acids Res.* 48, D148–D154. doi:10.1093/nar/gkz896
- Huang, S., Yang, J., Fong, S., and Zhao, Q. (2020). Artificial Intelligence in Cancer Diagnosis and Prognosis: Opportunities and Challenges. *Cancer Lett.* 471, 61–71. doi:10.1016/j.canlet.2019.12.007
- Ishwaran, H., Kogalur, U. B., and Kogalur, M. U. B. (2020). *Package 'randomForestSRC'*.
- Karagkouni, D., Paraskevopoulou, M. D., Chatzopoulos, S., Vlachos, I. S., Tastsoglou, S., Kanellos, I., et al. (2018). DIANA-TarBase V8: a Decade-Long Collection of Experimentally Supported miRNA-Gene Interactions. *Nucleic Acids Res.* 46, D239–D245. doi:10.1093/nar/gkx1141
- Kassambara, A., Kosinski, M., Biecek, P., Fabian, S., and survminer (2017). *Drawing Survival Curves Using ggplot2*. R. Package Version 0.3.1.
- Kolde, R., and Kolde, M. R. (2015). Package 'pheatmap'. R. Package 1, 790.
- Lathwal, A., Kumar, R., Arora, C., and Raghava, G. P. S. (2020). Identification of Prognostic Biomarkers for Major Subtypes of Non-small-cell Lung Cancer Using Genomic and Clinical Data. *J. Cancer Res. Clin. Oncol.* 146, 2743–2752. doi:10.1007/s00432-020-03318-3
- Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The Sva Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments. *Bioinformatics* 28, 882–883. doi:10.1093/bioinformatics/bts034
- Li, X., Shi, Y., Yin, Z., Xue, X., and Zhou, B. (2014). An Eight-miRNA Signature as a Potential Biomarker for Predicting Survival in Lung Adenocarcinoma. *J. Transl. Med.* 12, 159. doi:10.1186/1479-5876-12-159
- Li, X., Lu, C., Lu, Q., Li, C., Zhu, J., Zhao, T., et al. (2019). Differentiated Super-enhancers in Lung Cancer Cells. *Sci. China Life Sci.* 62, 1218–1228. doi:10.1007/s11427-018-9319-4
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8
- Miska, E. A. (2007). MicroRNA Expression Profiles Classify Human Cancers. *Cytometry B-Clinical Cytometry* 72b, 126. doi:10.1002/cyto.b.v72b.6
- Mecnik, B., Bindea, G., Pagès, F., and Galon, J. (2011). Tumor Immunosurveillance in Human Cancers. *Cancer Metastasis Rev.* 30, 5–12. doi:10.1007/s10555-011-9270-7
- Mounir, M., Lucchetta, M., Silva, T. C., Olsen, C., Bontempi, G., Chen, X., et al. (2019). New Functionalities in the TCGAbiolinks Package for the Study and Integration of Cancer Data from GDC and GTEx. *Plos Comput. Biol.* 15, e1006701. doi:10.1371/journal.pcbi.1006701
- Novello, C., Pazzaglia, L., Cingolani, C., Conti, A., Quattrini, I., Manara, M. C., et al. (2013). miRNA Expression Profile in Human Osteosarcoma: Role of miR-1 and miR-133b in Proliferation and Cell Cycle Control. *Int. J. Oncol.* 42, 667–675. doi:10.3892/ijo.2012.1717
- Ru, Y., Kechris, K. J., Tabakoff, B., Hoffman, P., Radcliffe, R. A., Bowler, R., et al. (2014). The multiMiR R Package and Database: Integration of microRNA-Target Interactions along with Their Disease and Drug Associations. *Nucleic Acids Res.* 42, e133. doi:10.1093/nar/gku631
- Sheng, R., Li, X., Wang, Z., and Wang, X. (2020). Circular RNAs and Their Emerging Roles as Diagnostic and Prognostic Biomarkers in Ovarian Cancer. *Cancer Lett.* 473, 139–147. doi:10.1016/j.canlet.2019.12.043
- Sherafatian, M., and Arjmand, F. (2019). Decision Tree-Based Classifiers for Lung Cancer Diagnosis and Subtyping Using TCGA miRNA Expression Data. *Oncol. Lett.* 18, 2125–2131. doi:10.3892/ol.2019.10462
- Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer Statistics, 2019. *CA A. Cancer J. Clin.* 69, 7–34. doi:10.3322/caac.21551
- Sun, J., Wei, Q., Zhou, Y., Wang, J., Liu, Q., and Xu, H. (2017). A Systematic Analysis of FDA-Approved Anticancer Drugs. *BMC Syst. Biol.* 11, 87–43. doi:10.1186/s12918-017-0464-7
- The Cancer Genome Atlas Research Network (2014). Comprehensive Molecular Profiling of Lung Adenocarcinoma. *Nature* 511, 543–550. doi:10.1038/nature13385
- Therneau, T., and Lumley, T. (2010). Survival Analysis, Including Penalised Likelihood. *R. Package Version 2*, 36–14.
- Volinia, S., Calin, G. A., Liu, C.-G., Ambs, S., Cimmino, A., Petrocca, F., et al. (2006). A microRNA Expression Signature of Human Solid Tumors Defines Cancer Gene Targets. *Pnas* 103, 2257–2261. doi:10.1073/pnas.0510565103
- Wang, C., Ren, T., Wang, K., Zhang, S., Liu, S., Chen, H., et al. (2017). Identification of Long Non-coding RNA P34822 as a Potential Plasma Biomarker for the Diagnosis of Hepatocellular Carcinoma. *Sci. China Life Sci.* 60, 1047–1050. doi:10.1007/s11427-017-9054-y
- Wang, J., Han, X., and Sun, Y. (2017). DNA Methylation Signatures in Circulating Cell-free DNA as Biomarkers for the Early Detection of Cancer. *Sci. China Life Sci.* 60, 356–362. doi:10.1007/s11427-016-0253-7
- Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., and Li, T. (2009). miRecords: an Integrated Resource for microRNA-Target Interactions. *Nucleic Acids Res.* 37, D105–D110. doi:10.1093/nar/gkn851
- Yang, Y., Yu, Y., and Lu, S. (2020). Effectiveness of PD-1/pd-L1 Inhibitors in the Treatment of Lung Cancer: Brightness and challenge. *Sci. China Life Sci.* 63, 1499–1514. doi:10.1007/s11427-019-1622-5
- Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118
- Yu, S.-L., Chen, H.-Y., Chang, G.-C., Chen, C.-Y., Chen, H.-W., Singh, S., et al. (2008). MicroRNA Signature Predicts Survival and Relapse in Lung Cancer. *Cancer Cell* 13, 48–57. doi:10.1016/j.ccr.2007.12.008

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Wu, Lou, Ma, Xu and Shi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.