



Prediction of Recurrence in Cervical Cancer Using a Nine-lncRNA Signature

Yu Mao*, Lixin Dong, Yue Zheng, Jing Dong and Xin Li

Department of Oncology, First Hospital of Qinhuangdao, Qinhuangdao, China

OPEN ACCESS

Edited by:

Madhuchanda Bhattacharjee,
University of Hyderabad, India

Reviewed by:

Ao Li,
University of Science and Technology
of China, China
Izabela Makalowska,
Adam Mickiewicz University
in Poznań, Poland

*Correspondence:

Yu Mao
ayabluewin@163.com

Specialty section:

This article was submitted to
Bioinformatics and Computational
Biology,
a section of the journal
Frontiers in Genetics

Received: 11 October 2018

Accepted: 15 March 2019

Published: 03 April 2019

Citation:

Mao Y, Dong L, Zheng Y, Dong J
and Li X (2019) Prediction
of Recurrence in Cervical Cancer
Using a Nine-lncRNA Signature.
Front. Genet. 10:284.
doi: 10.3389/fgene.2019.00284

Background and Objective: As a common cancer type in women, cervical cancer remains one of the leading causes of cancer-associated mortalities world wide. Recent evidence has demonstrated the regulatory role of a large number of long non-coding RNAs (lncRNAs) in cervical cancer. Here, we aimed to identify new biomarkers that related with the recurrence through comprehensive bioinformatics analysis.

Methods: Firstly, we collected online lncRNA expression data of cervical cancer patients which were divided into training, validation, and test set. Then we developed a nine-lncRNA signature from training set by conducting LASSO Cox regression model along with 10-fold cross validation. The prognostic value of this risk score was validated in all the three sets using Kaplan–Meier analysis, C-index, time-dependent ROC curves and dynamic AUC. Biological function of these lncRNAs in cervical cancer cells were evaluated by performing gene ontology biological process enrichment and Kyoto Encyclopedia of Genes and Genomes signaling pathways analysis.

Results: According to the results, a higher predict accuracy was observed in the nine-lncRNA signature than that of FIGO stage in all the three sets. Stratified analysis also demonstrated that the nine-lncRNA signature can predict the recurrence of cervical cancer within FIGO stage. The potential mechanisms underlying the nine-lncRNAs from the signature were also identified according to the gene enrichment analysis.

Conclusion: In the present article, we provided a reliable prognostic tool to facilitate the individual management of patients with cervical cancer after treatment.

Keywords: cervical cancer, lncRNA, prognosis, GEO, TCGA

INTRODUCTION

As the second most common cause of female cancer-associated mortalities worldwide, cervical cancer ranks the fourth most frequently diagnosed cancer (Jemal et al., 2011; Torre et al., 2015). Previous studies demonstrated that cervical cancer was closely associated with human papillomavirus (HPV) infection (Walboomers et al., 1999; Castellsagué et al., 2006). Evidence suggested that HPV mediated the genomic instability and somatic mutations which play important roles in the pathogenesis of cervical cancer (Cancer Genome Atlas Research Network et al., 2017). Despite great advancement achieved in the standard treatment such as surgery, radiotherapy, and chemotherapy, the prognosis for patients with cervical cancer remains poor (Fuller et al., 2007).

Hence, identification of the new suitable prognosis biomarkers are critical for cervical cancer. Recent studies have showed that genomic factors could be the indicators for the prognosis of cervical cancer (Mao et al., 2018a).

Previous studies based on the gene expression data have identified a series of gene signature which was used as recurrence predictive model. Huang et al. (2012) has identified a 7-gene signature which associated with the relapse and survival in patients with early stage cervical carcinoma. Lee et al. (2013) selected 12-genes and constructed a prognostic score model for recurrence prediction. However, limited number of studies investigated whether the long non-coding RNA (lncRNA) signature can predict the recurrence and disease free survival time of cervical cancer.

As the RNA transcripts longer than 200 nucleotides, lncRNAs lack the ability of generating protein (Gupta et al., 2010). However, they can still contribute to the modulation of tumor progression associated biological processes via chromatin remodeling, transcription and post-transcriptional processing (Tano and Akimitsu, 2012; Wang et al., 2014). In addition, recent studies have showed the association between lncRNAs and survivals of human cancers, such as prostate cancer, breast cancer and gastric cancer (Hu et al., 2014; Zhu et al., 2016; Li et al., 2018).

The Gene Expression Omnibus is an international public repository that archives and freely distributes microarray, next-generation sequencing, and other forms of high-throughput functional genomic data sets (Kodahl et al., 2014; Clough and Barrett, 2016). The Cancer Genome Atlas (TCGA) is a large-scale cancer genome project which provides researchers with multi-dimensional maps of the key genomic changes and clinic-pathological information in 33 types of cancer (Collins and TCGA Project Team, 2007). Hence, we downloaded the lncRNA data from the GEO and performed lncRNA profiling on cervical cancer patients. Finally, a prognostic, nine-lncRNA signature for cervical cancer was constructed from the training set of GEO and its predictive accuracy was further validated two independent validation sets.

MATERIALS AND METHODS

Data Source

As we have previously depicted, we firstly downloaded the “MINiML formatted family file(s)” of GSE44001 datasets from GEO which has the microarray data of 300 patients with cervical cancer and then processed these data sets using R to generate the lncRNA expression matrix. The clinical information (recurrence status, disease free survival time, and FIGO stage) of 300 patients was also extracted and patients were randomly assigned to a training set ($n = 150$) and an internal validation set ($n = 150$). RNA sequencing (RNA-seq) and corresponding clinical data (recurrence status and disease free survival time) were downloaded from the publicly available TCGA database. After excluded those without complete clinical and survival information, a number of 49 patients with cervical cancer were enrolled into the external test set. Each lncRNA expression level

was determined by the value of Reads Per Kilobase of exon model per Million mapped reads (RPKM) (Mao et al., 2018b, 2019b).

Microarray Data Analysis and lncRNA Signature Construction

Data preprocessing was performed according to our previous published study (Mao et al., 2018b,c, 2019a). After quantile normalization and log₂-scale transformation, we delineated the box plot (Supplementary Figure S1). Then we performed cox regression analysis with Least Absolute Shrinkage and Selection Operator (LASSO), which is a parameter selection method that manage high-dimensional regression variables with no prior feature selection step by shrinking all regression coefficients and forcing many variables to be exactly zero (Tibshirani, 1997; Mao et al., 2019b). To achieve variable selection and shrinkage, we put the normalized lncRNA expression data into LASSO Cox regression. During this process, the penalty regularization parameter λ was chosen via the cross-validation routine with an n -fold = 10 by using R package “glmnet” (Friedman et al., 2010). The value of lambda.min, the lambda value giving minimum mean cross-validated error was calculated using R. By applying the lambda.min, a nine-lncRNA signature was identified based on the expression of lncRNAs weighted by the coefficients from LASSO penalized regression. Then the score of each sample was calculated according to expression levels of the RNAs (Exp_i) and LASSO coefficients (L_i).

$$\text{Risk score} = \sum_{i=1}^n Exp_i \times L_i$$

The resulting score allowed the division of patients into two classes, namely the high-risk group and low risk group based on the median risk score value.

Prognostic Signature Validation

Firstly, Kaplan–Meier and Chi-square analysis was performed using Graphpad prism. C-index was calculated using R with Package “survival” (Therneau and Grambsch, 2000). After that, a time-dependent ROC (receiver operating characteristic) curves along with the dynamic area under the time specific ROC curves (dynamic -AUC) was obtained by using R with Package “risksetROC” (Heagerty et al., 2000; Heagerty and Zheng, 2005).

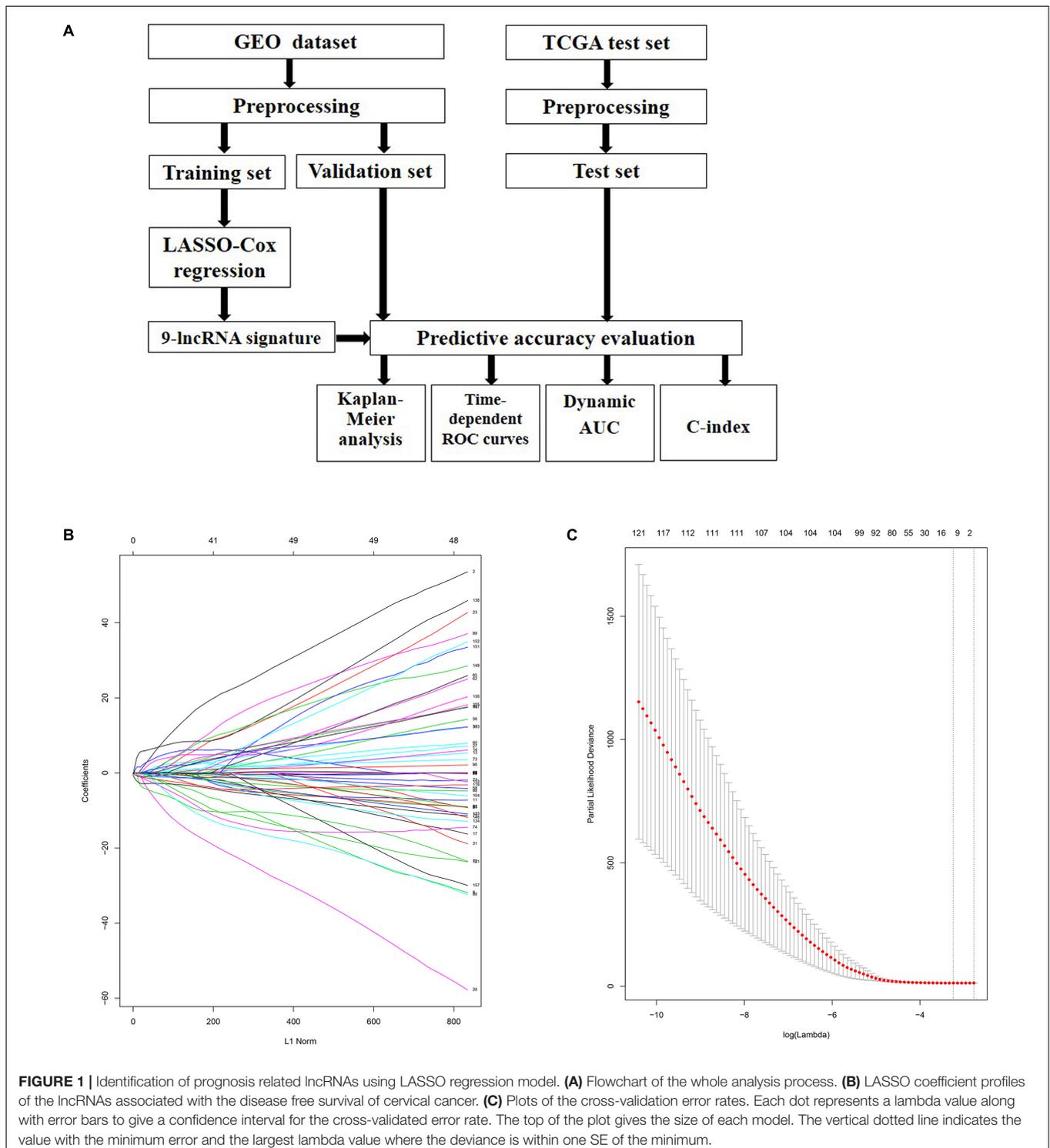
Functional Annotation of lncRNA Target Genes

Identifying target genes of lncRNAs is an important step in studying the function of lncRNA in cervical cancer. In this study, we first predict the target genes of lncRNAs in the signature by using the starBase v2.0 (Yang et al., 2011; Li et al., 2014). Then these genes were put into gene ontology (GO) biological process enrichment, Kyoto Encyclopedia of Genes and Genomes (KEGG), and REACTOM¹ signaling pathways analysis. The enrichment analysis results, including enrichment score, the count of genes enriched in the terms and false discovery rate were

¹<https://reactome.org/>

shown as pictures which plotted using R with Package “ggplot2” (Wickham, 2015). The potential relationship among these target genes was analyzed using Search Tool for the Retrieval of Interacting Genes (STRING) which is an online tool designed to evaluate the protein–protein interaction (PPI) information (Damian et al., 2015). The PPI network were then plotted.

Moreover, we identified the related genes of the nine-lncRNAs in the signature by calculating pearson correlation coefficient between lncRNAs and mRNAs using TCGA datasets as previously depicted (Mao et al., 2018b,c, 2019a). Genes with pearson correlation coefficient >0.60 or <-0.40 was considered as associated with genes in the signature and was enrolled into



the analysis. After that, these genes were also put into functional enrichment and pathway analysis which was then visualized using Cytoscape software with ClueGO and CluePedia (Shannon et al., 2003; Bindea et al., 2009).

RESULTS

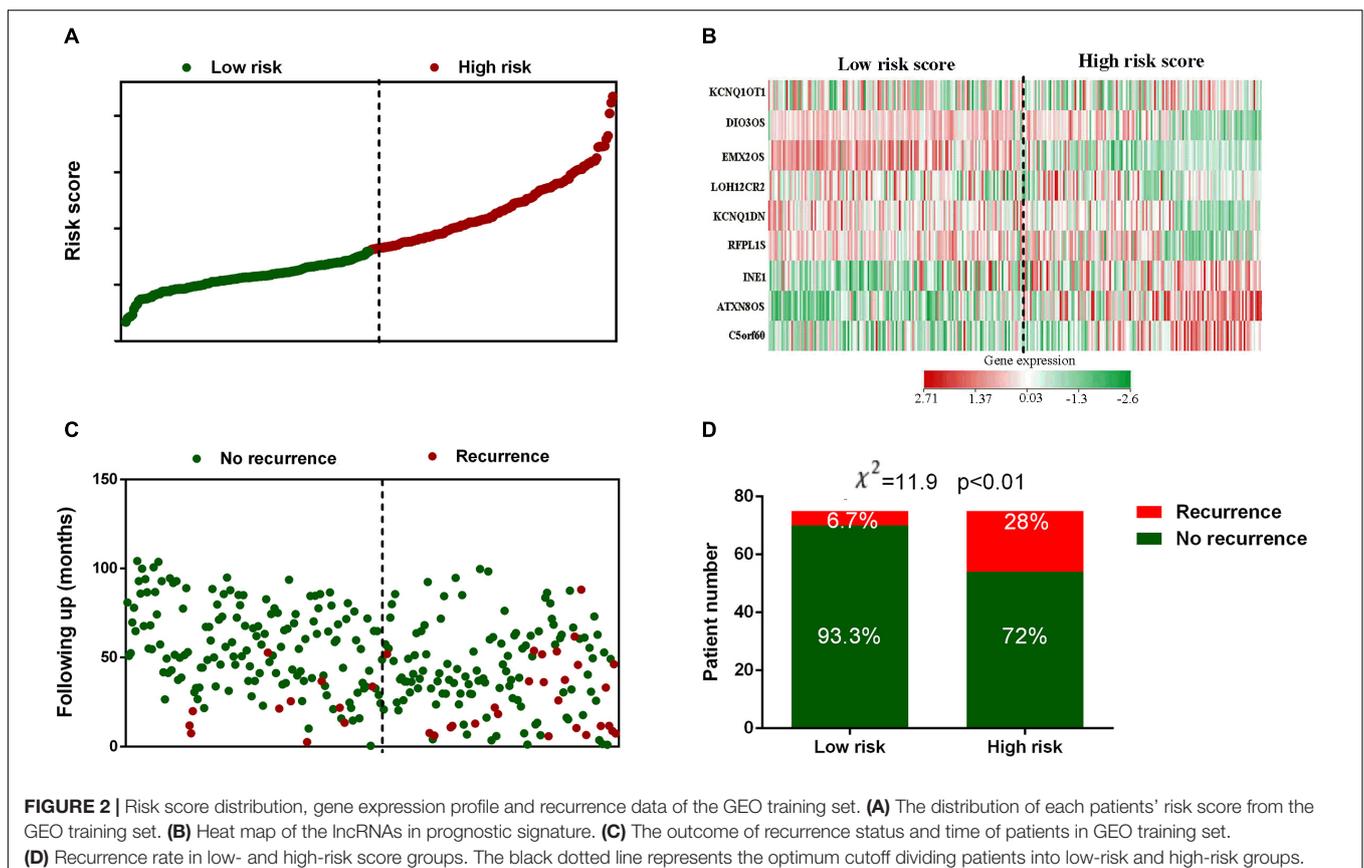
Prognosis Related lncRNA Identification and Signature Generation

A flowchart which depicted the whole process of our analysis was firstly plotted in **Figure 1A**. Samples in GEO datasets were firstly quantile normalized and the distributions for the dataset of lncRNA profiles in each patient was shown as box plot using the R software package (**Supplementary Figure S1**). 300 patients from the GSE44001 cohort were randomly divided into a training cohort ($n = 150$) and internal validation cohort ($n = 150$). LASSO Cox regression model along with 10-fold cross validation was performed to analyze the expression data in the training cohort (**Figure 1B**). According to the results, a set of nine-lncRNAs along with their coefficients were identified and included into a risk score formula. Hence, we developed a nine-lncRNA signature based on their expression level and coefficients. Risk score = $(3.31562585 * ATXN8OS) + (0.13987057 * C5orf60) + (-0.43216636 * DIO3OS) + (-0.92247218 * EMX2OS) + (1.13309789 * INE1) + (-4.48055889 * KCNQ1DN) + (-0.08067727 * KCNQ1OT1) + (-0.09737496 * LOH12CR2) + (-0.66622831 * RFPL1S)$ (**Figure 1C**).

The risk score for each patient in the training set was calculated and plotted in **Figure 2A**. Besides, the corresponding heatmap of lncRNA expression level in the signature was also presented. **Figure 2C** showed the distribution of disease free survival time and recurrence status of each patient which ranked according to the score in **Figure 2A**. According to **Figures 2B,C**, three lncRNAs in the signature had positive coefficients including ATXN8OS, C5orf60, and INE1 which revealed that higher expression level of these lncRNAs was associated with shorter disease free survival while the other six lncRNAs including DIO3OS, EMX2OS, KCNQ1DN, KCNQ1OT1, LOH12CR2, and RFPL1S had negative coefficients which indicated that their expression level was negatively related with the possibility of cervical cancer recurrence. Moreover, Chi-square analysis showed that the recurrence rate in high risk group was significant higher than low risk group (**Figure 2D**).

Validation of the lncRNA Signature's Survival Predict Accuracy

The robustness of the lncRNA signature was tested by evaluating their ability to classify the high-risk group and low risk group in all the three datasets. The gene signature based risk



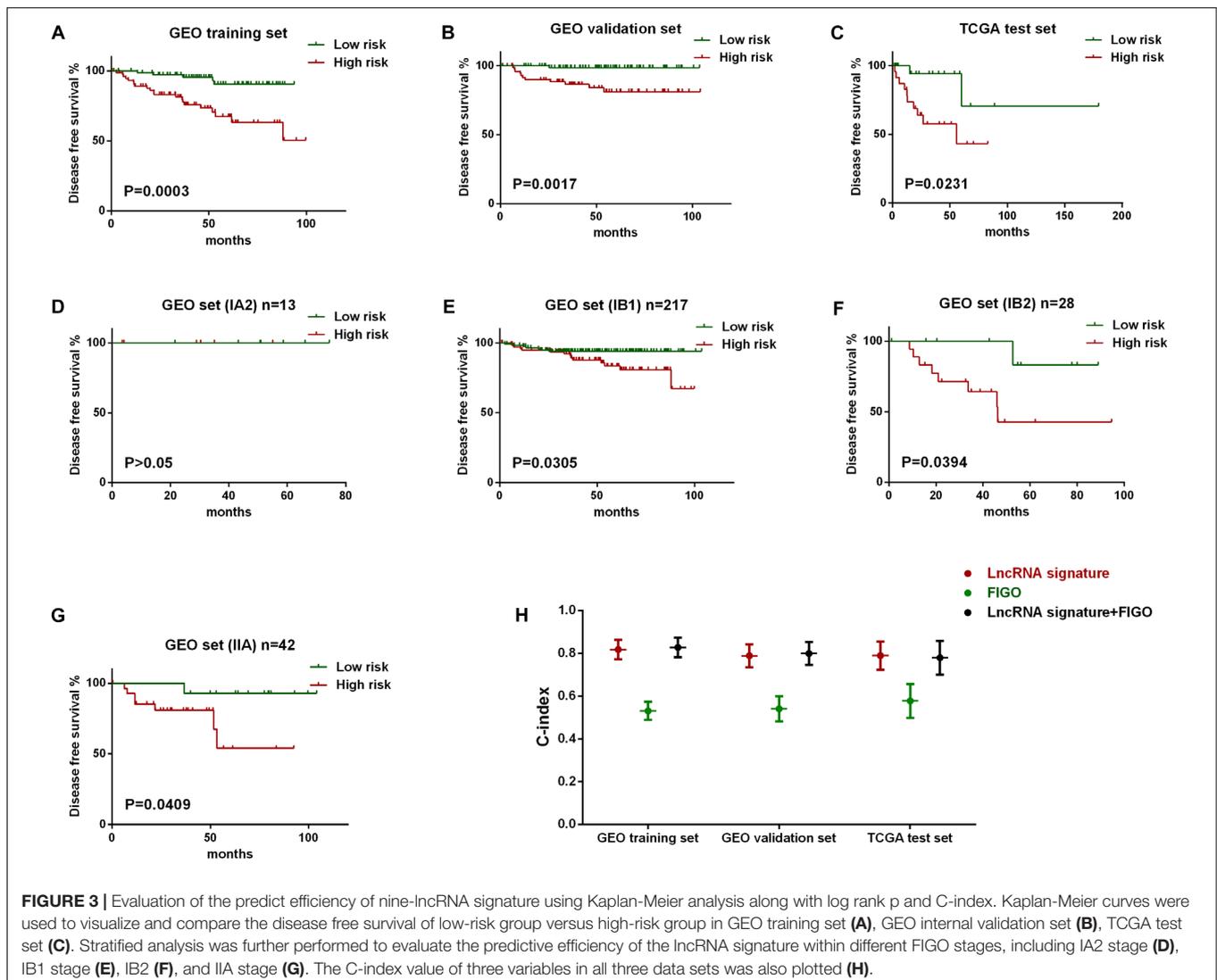
score for each patient was firstly calculated. Then patients were divided into high- or low-risk group according to the median value. Kaplan–Meier curves were plotted, along with log rank p -test, to compare the disease free survival of the two groups. According to the results, significant differences in Kaplan–Meier survival analysis were observed in high- and low- risk group separated by the lncRNA signature in GEO training (Figure 3A) and validation set (Figure 3B). Similar outcomes were also found in external validation set (TCGA test set) (Figure 3C).

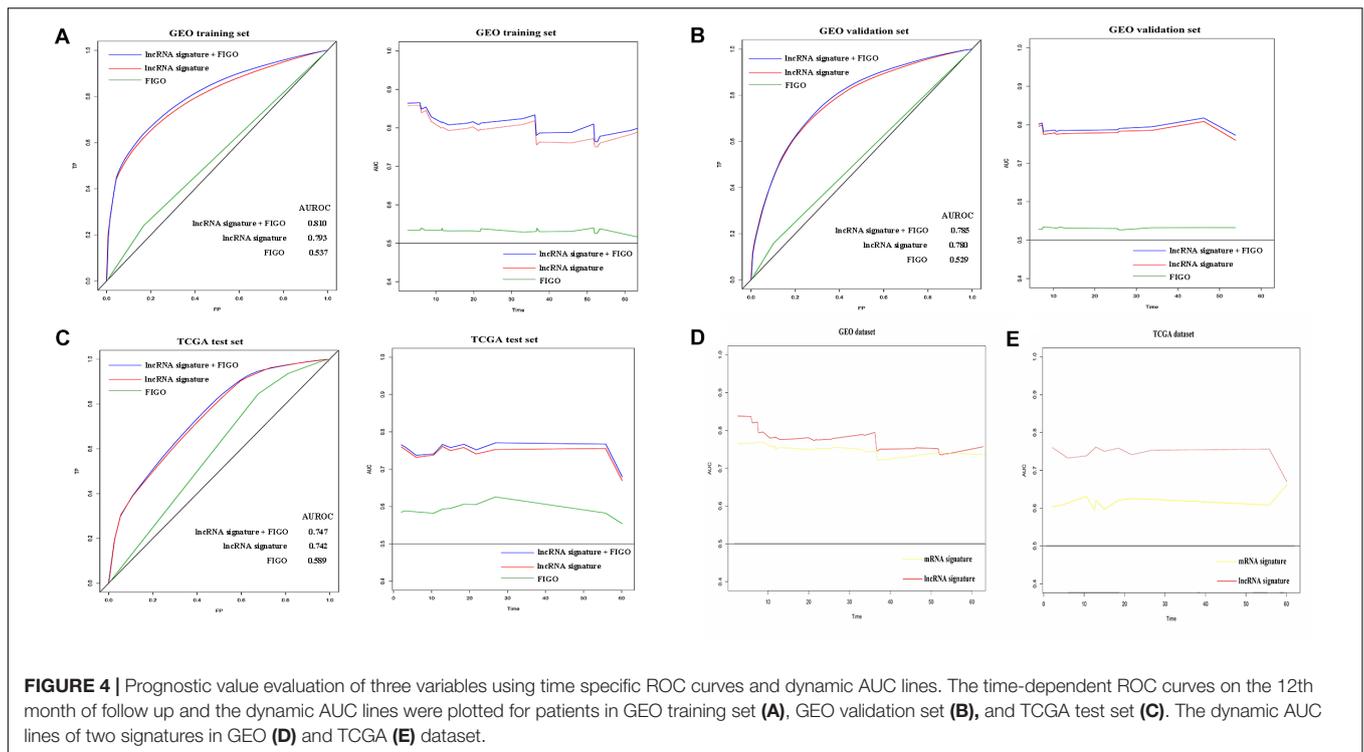
Besides, stratified analysis was further performed to evaluate the predictive efficiency of the lncRNA signature within FIGO stages. Figures 3E–G showed that the lncRNA signature can predict the tumor recurrence in patients of different FIGO stage, except for those in IA2 stage as a results of the limited number (Figure 3D). We also evaluate the predict accuracy of the lncRNA signature by using C-index which showed that the C-index value of lncRNA signature was higher than FIGO and a new

variable combining both has the highest value than either alone (Figure 3H).

Further Evaluation of the Nine-lncRNA Signature Predictive Efficiency Using Time-Dependent ROC Curves and Dynamic AUC

Next, we tried to determine the sensitivity and specificity of the predictive model and the changes of predictive accuracy over time. Hence, we introduced time-dependent ROC (receiver operating characteristic) curves and dynamic AUC to assess the predictive accuracy of lncRNA signature, FIGO stage and a new variable combined both. Firstly, we plotted the time-dependent ROC curves and calculated the corresponding AUC on the 12th month of follow up. As shown in Figure 4, predict accuracy of the new variable which combined both the signature and FIGO was better than either alone on the 12th month of follow up in all three sets. In addition, the lncRNA signature





has a better predictive accuracy than FIGO stage in all the three subsets (Figure 4).

Subsequently, the dynamic area under curves (dynamic - AUC) at continuous time was calculated and depicted as the line chart. Results showed that the lncRNA signature combined with FIGO stage has a higher dynamic AUC level than either alone. The lncRNA alone also showed high accuracy with the dynamic AUC estimates exceeding 0.75 which was much better than FIGO stage alone. Similarly, the nine-lncRNA signature showed better predict accuracy than FIGO stage in the GEO validation set and TCGA test set as shown in Figures 4B,C. Hence, the lncRNA signature could predict the recurrence of cervical cancer patients with high efficiency.

Furthermore, we compared the recurrence predict performance of the nine-lncRNA signature with a 12-mRNA signature which was established in a previously published article (Lee et al., 2013). Results showed that similar predict accuracy was observed between the two signatures in GEO datasets (Figure 4D). However, the nine-lncRNA signature showed a better predict ability than the 12-gene signature in TCGA dataset (Figure 4E).

Functional Annotation of lncRNA Associated Genes

In order to further describe the function of lncRNAs in the signature, we performed the function annotation and enrichment analysis. Studies on lncRNA have demonstrated that lncRNAs usually act as competing endogenous RNAs (ceRNAs) which modulate the gene expression and maintain the functional balance of various gene networks (Kartha and Subramanian,

2014). Hence, we first identified the target genes of the lncRNAs in the signature and then put these genes into GO, KEGG, and REACTOM analysis. Results showed that target genes participate in biological processes, such as RNA stability and RNA splicing, and pathways such as RNA transport and spliceosome (Figures 5A,B). The PPI network of the target genes were shown in Figure 5C.

Moreover, genes, which was considered as related with the lncRNAs in the signature, were put into GO biological process and KEGG signaling pathways analysis. As shown in Supplementary Figure S2, lncRNA-related genes participate in biological processes such as cell proliferation and cell cycle. These data indicated the potential regulatory mechanism of lncRNAs in the signature.

DISCUSSION

Cervical cancer is a common gynecological malignancy with high mortality and its incidence of the younger trend in recent years. Due to the advancements in systemic chemotherapy and radiotherapy, the incidence of cervical cancer and mortality have dropped significantly. However, its recurrence rate and mortality in developing countries still remains high (Saslow et al., 2012). Considering the poor prognosis of cervical cancer, novel biomarkers, especially genetic biomarkers characterizing the inner molecular complexity, began to be used as predictors for the disease free survival and overall survival. With the development of high-throughput technology, more and more attention have been paid to the gene expression profiling which can be used to identify the biomarker that related to the heterogeneities and

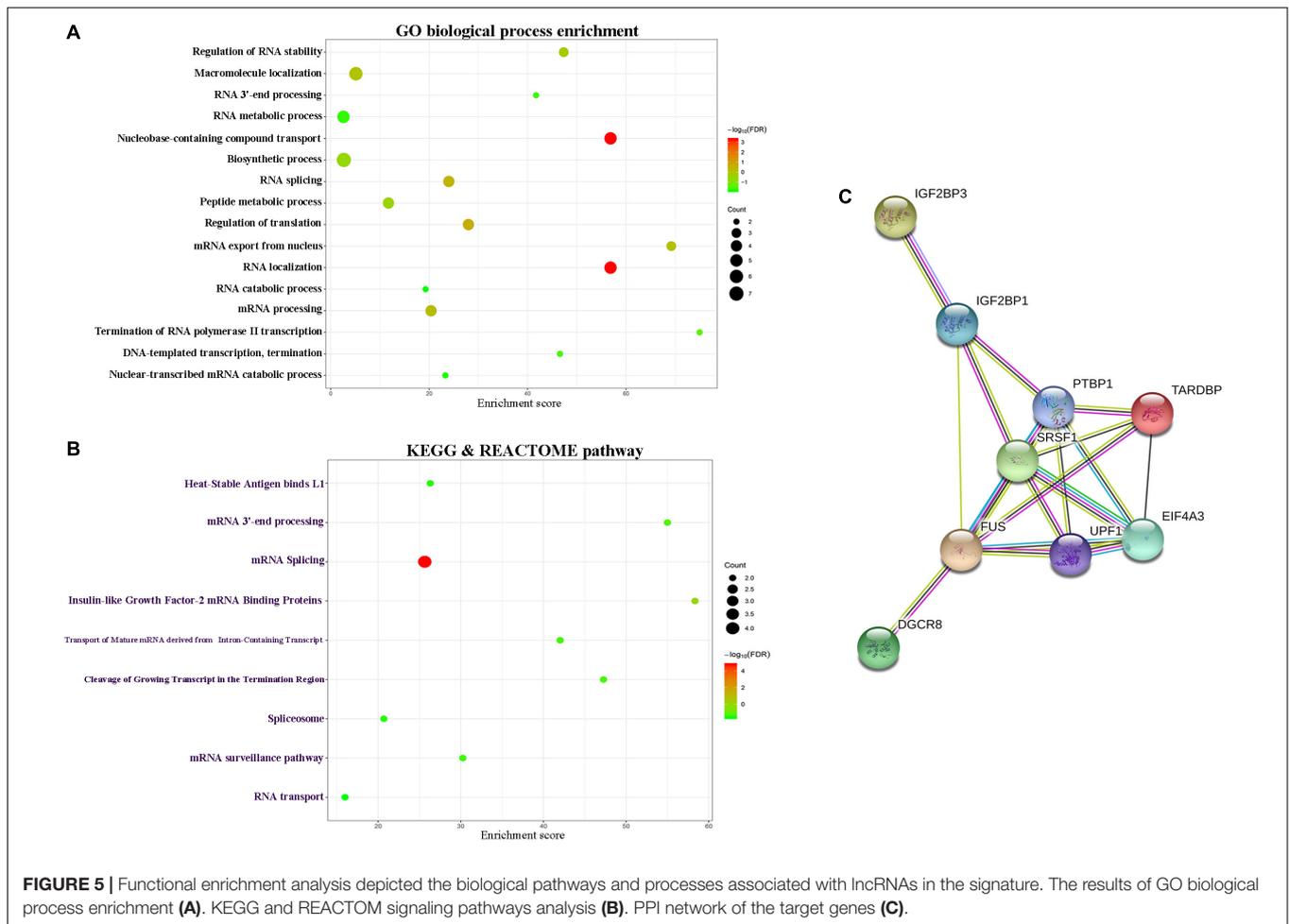


FIGURE 5 | Functional enrichment analysis depicted the biological pathways and processes associated with lncRNAs in the signature. The results of GO biological process enrichment (A), KEGG and REACTOM signaling pathways analysis (B), PPI network of the target genes (C).

molecular of cervical cancer (Zhan et al., 2016; Dai et al., 2017). Previous studies have analyzed open-access datasets and establish an integrated model which could predict glioblastoma prognosis with high accuracy (Zhang et al., 2016). By applying the public breast cancer dataset, a series of recurrent copy number aberrations in complex patterns was also discovered via non-negative sparse singular value decomposition (Xi and Li, 2016). Here we downloaded and mined the open-access lncRNAs expression datasets of cervical cancer from GEO and performed a series analysis.

The microarray data we downloaded from GEO belongs to high-throughput biological data. In order to solve the common problem “curse-of-dimensionality” (small sample size combined with a very large number of genes) in high-throughput biological data, LASSO Cox regression model along with 10-fold cross validation was applied to analyze the expression data in the training cohort. As we have previously depicted, LASSO regression is good at handling high dimensional regression variables with no prior feature selection step by shrinking all regression coefficients toward zero, and thus forcing many regression variables to be exactly zero (Tibshirani, 1997; Mao et al., 2019a,b). It has been demonstrated that LASSO Cox regression model can achieve

high stability and accurate predictions in dealing with the “curse-of-dimensionality” data (Algamal and Lee, 2015). By performing LASSO Cox regression model along with 10-fold cross validation, we identified a series of lncRNAs with the most powerful prognostic ability. At the same time, the regression coefficients were presented by LASSO Cox regression, based on which a nine-lncRNA signature for the prediction of disease free survival of cervical cancer was constructed.

In evaluating the predict accuracy of the nine-lncRNA signature, we first plotted the Kaplan–Meier curves and calculated the C-index of the signature. According to the results, significant difference was observed between the survive time along with status of the two groups separated by the same criteria in all three sets. Besides, further study showed that the nine-lncRNA signature has predict efficiency within different FIGO stage. Similarly, the C-index of lncRNA signature was higher than FIGO and a new variable combining both has the highest value than either alone. Moreover, we calculated the AUC for time specific ROC curves at continuous time point and dynamic AUC line was plotted to depict the temporal changes in accuracy. Results showed that the nine-lncRNA signature has a higher accuracy than FIGO stage alone. The new variable combined

both nine-lncRNA signature and FIGO stage has better predict capacity for recurrence than either alone.

To gain more insights into the modulatory roles of the lncRNAs in the signature, we performed functional enrichment and annotation analysis for nine-lncRNAs in the signature. Results showed that lncRNAs in the signature might regulate biological processes, such as RNA splicing, cell proliferation and cell cycle, and pathways such as RNA transport and cell adhesion molecules. These data indicated the potential regulatory mechanism of lncRNAs in the signature.

Moreover, we also explored the regulatory mechanism of nine-lncRNAs in the signature by searching the published article. Among the nine-lncRNAs, lncRNA ATXN8OS participate in spinocerebellar ataxia by affecting the localization and activity of splicing factors and mutations in the ATXN8OS are associated with the amyotrophic lateral sclerosis (Moseley et al., 2006; Hirano et al., 2018). Dysregulation of lncRNA DIO3OS was closely related with inflammatory bowel disease (Wang et al., 2018). The downregulation of lncRNA EMX2OS might independently predict shorter recurrence-free survival of classical papillary thyroid cancer (Gu et al., 2018). LncRNA EMX2OS was also identified as associated with myalgic encephalomyelitis/chronic fatigue syndrome (Yang et al., 2018). LncRNA INE1 was considered as the potential hotspot for neurogenetic disorders (Thiselton et al., 2002). In Wilms' tumors, reduced expression of lncRNA KCNQ1DN existing far from the H19/IGF2 region and may play regulatory role in tumor progression (Xin et al., 2000). LncRNA KCNQ1OT1 facilitates the progression of non-small-cell lung carcinoma via modulating miRNA-27b-3p/HSP90AA1 axis (Dong et al., 2018). On the other side, lncRNA KCNQ1OT1 controls maternal p57 expression in muscle cells by promoting H3K27me3 accumulation to an intragenic MyoD-binding region (Andresini et al., 2019). Studies on RFPL1S suggests that RFPL1S may function as a post-transcriptional regulation of the sense RFPL genes (Seroussi et al., 1999).

CONCLUSION

In summary, we conducted comprehensive comparative analysis of lncRNA expression pattern and constructed

a nine-lncRNA signature that can be applied to predict disease free survival in cervical cancer. Gene annotation and functional enrichment analysis further revealed the underlying mechanisms whereby lncRNAs in the signature exerts their biological roles in tumor progression. Although further study are still needed to confirm the established signature, our study here still provide valuable indication for both the basic research and clinical treatment of cervical cancer.

DATA AVAILABILITY

The datasets generated for this study can be found in GEO (GSE44001) and TCGA.

AUTHOR CONTRIBUTIONS

YM contributed to the study design, data profiling, and manuscript draft. YZ and LD prepared the figures and tables. JD and XL performed the language editing. All authors reviewed and approved the final manuscript.

FUNDING

This work was supported by Scientific Research Fund of Hebei Health and Family Planning Commission (20181192).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00284/full#supplementary-material>

FIGURE S1 | Box plots showed the distributions of genes for the 300 samples were nearly the same after normalization.

FIGURE S2 | Functional enrichment analysis depicted the biological pathways and processes associated with correlated genes. The results of GO biological process enrichment **(A)** and KEGG signaling pathways analysis **(B)**.

REFERENCES

- Algamil, Z. Y., and Lee, M. H. (2015). *Penalized Logistic Regression with the Adaptive Lasso for Gene Selection in High-Dimensional Cancer Classification*. Oxford: Pergamon Press, Inc.
- Andresini, O., Rossi, M. N., Matteini, F., Petrai, S., Santini, T., and Maione, R. (2019). The long non-coding RNA Kcnq1ot1 controls maternal p57 expression in muscle cells by promoting H3K27me3 accumulation to an intragenic MyoD-binding region. *Epigenetics Chromatin* 12:8. doi: 10.1186/s13072-019-0253-1
- Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25, 1091–1093. doi: 10.1093/bioinformatics/btp101
- Cancer Genome Atlas Research Network, Barretos Cancer Hospital, Baylor College of Medicine, Harvard Medical School, and Albert Einstein College of Medicine, et al. (2017). Integrated genomic and molecular characterization of cervical cancer. *Nature*, 543, 378–384. doi: 10.1038/nature21386
- Castellsagué, X., Díaz, M., Sanjosé, S. D., Muñoz, N., Herrero, R., Franceschi, S., et al. (2006). Worldwide human papillomavirus etiology of cervical adenocarcinoma and its cofactors: implications for Screening and Prevention. *J. Natl. Cancer Inst.* 98, 303–315. doi: 10.1093/jnci/djj067
- Clough, E., and Barrett, T. (2016). The gene expression omnibus database. *Methods Mol. Biol.* 1418, 93–110. doi: 10.1007/978-1-4939-3578-9_5
- Collins, A., and TCGA Project Team (2007). The cancer genome atlas (TCGA) pilot project. *Cancer Res.* 67, LB-247–LB-247.
- Dai, F., Mei, L., Meng, S., Ma, Z., Guo, W., Zhou, J., et al. (2017). The global expression profiling in esophageal squamous cell carcinoma. *Genomics* 109, 241–250. doi: 10.1016/j.ygeno.2017.04.005
- Damian, S., Andrea, F., Stefan, W., Kristoffer, F., Davide, H., Jaime, H. C., et al. (2015). STRING v10: protein-protein interaction networks, integrated

- over the tree of life. *Nucleic Acids Res.* 43, D447–D452. doi: 10.1093/nar/gku1003
- Dong, Z., Yang, P., Qiu, X., Liang, S., Guan, B., Yang, H., et al. (2018). KCNQ1OT1 facilitates progression of non-small-cell lung carcinoma via modulating miRNA-27b-3p/HSP90AA1 axis. *J. Cell Physiol.* doi: 10.1002/jcp.27788 [Epub ahead of print].
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22. doi: 10.18637/jss.v033.i01
- Fuller, C. D., Wang, S. J., Thomas, C. R., Hoffman, H. T., Weber, R. S., and Rosenthal, D. I. (2007). Conditional survival in head and neck squamous cell carcinoma. *Cancer* 109, 1331–1343. doi: 10.1002/cncr.22563
- Gu, Y., Feng, C., Liu, T., Zhang, B., and Yang, L. (2018). The downregulation of lncRNA EMX2OS might independently predict shorter recurrence-free survival of classical papillary thyroid cancer. *PLoS One* 13:e0209338. doi: 10.1371/journal.pone.0209338
- Gupta, R. A., Shah, N., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076. doi: 10.1038/nature08975
- Heagerty, P. J., Lumley, T., and Pepe, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 56, 337–344. doi: 10.1111/j.0006-341X.2000.00337.x
- Heagerty, P. J., and Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics* 61, 92–105. doi: 10.1111/j.0006-341X.2005.030814.x
- Hirano, M., Samukawa, M., Isono, C., Saigoh, K., Nakamura, Y., and Kusunoki, S. (2018). Noncoding repeat expansions for ALS in Japan are associated with the ATXN8OS gene. *Neurol. Genet.* 4, e252. doi: 10.1212/NXG.0000000000000252
- Hu, Y., Chen, H. Y., Yu, C. Y., Xu, J., Wang, J. L., Qian, J., et al. (2014). A long non-coding RNA signature to improve prognosis prediction of colorectal cancer. *Oncotarget* 5, 2230–2240. doi: 10.18632/oncotarget.1895
- Huang, L., Zheng, M., Zhou, Q. M., Zhang, M. Y., Yu, Y. H., Yun, J. P., et al. (2012). Identification of a 7-gene signature that predicts relapse and survival for early stage patients with cervical carcinoma. *Med. Oncol.* 29, 2911–2918. doi: 10.1007/s12032-012-0166-3
- Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., and Forman, D. (2011). Global cancer statistics. *CA Cancer J. Clin.* 61, 69–90. doi: 10.3322/caac.20107
- Kartha, R. V., and Subramanian, S. (2014). Competing endogenous RNAs (ceRNAs): new entrants to the intricacies of gene regulation. *Front. Genet.* 5:8. doi: 10.3389/fgene.2014.00008
- Kodahl, A. R., Lyng, M. B., Binder, H., Cold, S., Gravgaard, K., Knoop, A. S., et al. (2014). Novel circulating microRNA signature as a potential non-invasive multi-marker test in ER-positive early-stage breast cancer: a case control study. *Mol. Oncol.* 8, 874–883. doi: 10.1016/j.molonc.2014.03.002
- Lee, Y. Y., Kim, T. J., Kim, J. Y., Choi, C. H., Do, I. G., Song, S. Y., et al. (2013). Genetic profiling to predict recurrence of early cervical cancer. *Gynecol. Oncol.* 131, 650–654. doi: 10.1016/j.ygyno.2013.10.003
- Li, J., Wang, W., Xia, P., Wan, L., Zhang, L., Yu, L., et al. (2018). Identification of a five-lncRNA signature for predicting the risk of tumor recurrence in breast cancer patients. *Int. J. Cancer* 143, 2150–2160. doi: 10.1002/ijc.31573
- Li, J. H., Liu, S., Zhou, H., Qu, L. H., and Yang, J. H. (2014). starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 42, D92–D97. doi: 10.1093/nar/gkt1248
- Mao, X., Qin, X., Li, L., Zhou, J., Zhou, M., Li, X., et al. (2018a). A 15-long non-coding RNA signature to improve prognosis prediction of cervical squamous cell carcinoma. *Gynecol. Oncol.* 149, 181–187. doi: 10.1016/j.ygyno.2017.12.011
- Mao, Y., Fu, Z., Zhang, Y., Dong, L., Zhang, Y., Zhang, Q., et al. (2018b). A seven-lncRNA signature predicts overall survival in esophageal squamous cell carcinoma. *Sci. Rep.* 8, 8823. doi: 10.1038/s41598-018-27307-2
- Mao, Y., Fu, Z., Zhang, Y., Dong, L., Zhang, Y., Zhang, Q., et al. (2018c). A six-microRNA risk score model predicts prognosis in esophageal squamous cell carcinoma. *J. Cell Physiol.* 234, 6810–6819. doi: 10.1002/jcp.27429
- Mao, Y., Fu, Z., Dong, L., Zheng, Y., Dong, J., and Li, X. (2019a). Identification of a 26-lncRNAs risk model for predicting overall survival of cervical squamous cell carcinoma based on integrated bioinformatics analysis. *DNA Cell Biol.* doi: 10.1089/dna.2018.4533 [Epub ahead of print].
- Mao, Y., Fu, Z., Zhang, Y., Dong, L., Zhang, Y., Zhang, Q., et al. (2019b). A six-microRNA risk score model predicts prognosis in esophageal squamous cell carcinoma. *J. Cell Physiol.* 234, 6810–6819. doi: 10.1002/jcp.27429
- Moseley, M. L., Zu, T., Ikeda, Y., Gao, W., Mosemiller, A. K., Daughters, R. S., et al. (2006). Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nat. Genet.* 38, 758–769. doi: 10.1038/ng1827
- Saslow, D., Solomon, D., Lawson, H. W., Killackey, M., Kulasingam, S., Cain, J., et al. (2012). American cancer society, american society for colposcopy and cervical pathology, and american society for clinical pathology screening guidelines for the prevention and early detection of cervical cancer. *Am. J. Clin. Pathol.* 62, 516–542. doi: 10.1309/AJCPGTD94EVRJCG
- Seroussi, E., Kedra, D., Pan, H. Q., Peyrard, M., Schwartz, C., Scambler, P., et al. (1999). Duplications on human chromosome 22 reveal a novel ret finger protein-like gene family with sense and endogenous antisense transcripts. *Genome Res.* 9, 803–814. doi: 10.1101/gr.9.9.803
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Tano, K., and Akimitsu, N. (2012). Long non-coding RNAs in cancer progression. *Front. Genet.* 3:219. doi: 10.3389/fgene.2012.00219
- Therneau, T. M., and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. New York, NY: Springer. doi: 10.1007/978-1-4757-3294-8
- Thiselton, D. L., Mcdowall, J., Brandau, O., Ramser, J., D'Esposito, F., Bhattacharya, S. S., et al. (2002). An integrated, functionally annotated gene map of the DXS8026-ELK1 interval on human Xp11.3-Xp11.23: potential hotspot for neurogenetic disorders. *Genomics* 79, 560–572. doi: 10.1006/geno.2002.6733
- Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* 16, 385–395. doi: 10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3
- Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global cancer statistics, 2012. *CA Cancer J. Clin.* 65, 87–108. doi: 10.3322/caac.21262
- Walboomers, J. M., Jacobs, M. V., Manos, M. M., Bosch, F. X., Kummer, J. A., Shah, K. V., et al. (1999). Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J. Pathol.* 189, 12–19. doi: 10.1002/(SICI)1096-9896(199909)189:1<12::AID-PATH431>3.0.CO;2-F
- Wang, G. Y., Zhu, Y. Y., and Zhang, Y. Q. (2014). The functional role of long non-coding RNA in digestive system carcinomas. *Bull. Cancer* 101, 27–31. doi: 10.1684/bdc.2014.2023
- Wang, S., Hou, Y., Chen, W., Wang, J., Xie, W., Zhang, X., et al. (2018). KIF9AS1, LINC01272 and DIO3OS lncRNAs as novel biomarkers for inflammatory bowel disease. *Mol. Med. Rep.* 17, 2195–2202. doi: 10.3892/mmr.2017.8118
- Wickham, H. (2015). ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* 3, 180–185. doi: 10.1002/wics.147
- Xi, J., and Li, A. (2016). “Discovering recurrent copy number aberrations in complex patterns via non-negative sparse singular value decomposition,” in *Proceedings of the IEEE/ACM Transactions on Computational Biology and Bioinformatics*, (Piscataway, NJ: IEEE). doi: 10.1109/TCBB.2015.2474404
- Xin, Z., Soejima, H., Higashimoto, K., Yatsuki, H., Zhu, X., Satoh, Y., et al. (2000). A novel imprinted gene, KCNQ1DN, within the WT2 critical region of human chromosome 11p15.5 and its reduced expression in Wilms' tumors. *J. Biochem.* 128, 847–853. doi: 10.1093/oxfordjournals.jbchem.a022823
- Yang, C. A., Bauer, S., Ho, Y. C., Sotzny, F., Chang, J. G., and Scheibenbogen, C. (2018). The expression signature of very long non-coding RNA in myalgic encephalomyelitis/chronic fatigue syndrome. *J. Transl. Med.* 16, 231. doi: 10.1186/s12967-018-1600-x
- Yang, J. H., Li, J. H., Shao, P., Zhou, H., Chen, Y. Q., and Qu, L. H. (2011). starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Res.* 39, D202–D209. doi: 10.1093/nar/gkq1056

- Zhan, C., Yan, L., Wang, L., Jiang, W., Zhang, Y., Xi, J., et al. (2016). Landscape of expression profiles in esophageal carcinoma by the cancer genome atlas data. *Dis. Esophagus* 29, 920–928. doi: 10.1111/dote.12416
- Zhang, Y., Li, A., Peng, C., and Wang, M. (2016). “Improve glioblastoma multiforme prognosis prediction by using feature selection and multiple kernel learning,” in *Proceedings of the IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 825–835. doi: 10.1109/TCBB.2016.2551745
- Zhu, X., Tian, X., Yu, C., Shen, C., Yan, T., Hong, J., et al. (2016). A long non-coding RNA signature to improve prognosis prediction of gastric cancer. *Mol. Cancer* 15, 60. doi: 10.1186/s12943-016-0544-0

Conflict of Interest Statement: 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.

Copyright © 2019 Mao, Dong, Zheng, Dong and Li. 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.