Genome-wide analysis of lncRNAs, miRNAs, and mRNAs forming a prognostic scoring system in esophageal squamous cell carcinoma

Xiaobo Shi[1], You Li[2], Yuchen Sun[1], Xu Zhao[1], Xuanzi Sun[1], Tuotuo Gong[1], Zhinan Liang[1], Yuan Ma[1] and Xiaozhi Zhang[1]

[1] Department of Radiation Oncology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, China
[2] Department of Peripheral Vascular Diseases, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, China

ABSTRACT

Background. Esophageal squamous cell carcinoma (ESCC) is the main subtype of esophageal carcinoma. Protein coding genes and non-coding RNAs can be powerful prognostic factors in multiple cancers, including ESCC. However, there is currently no model that integrates multiple types of RNA expression signatures to predict clinical outcomes.

Methods. The sequencing data (RNA-sequencing and miRNA-sequencing) and clinical data of ESCC patients were obtained from The Cancer Genome Atlas (TCGA) database, and Differential gene expression analysis, Cox regression analysis and Spearman correlation analysis were used to construct prognosis-related lncRNA-mRNA co-expression network and scoring system with multiple types of RNA. The potential molecular mechanisms of prognostic mRNAs were explored by functional enrichment analysis.

Results. A total of 62 prognostic lncRNAs, eight prognostic miRNAs and 66 prognostic mRNAs were identified in ESCC (P-value < 0.05) and a prognosis-related lncRNA-mRNA co-expression network was created. Five prognosis-related hub RNAs (CDCA2, MTBP, CENPE, PBK, AL033384.1) were identified. Biological process analysis revealed that mRNAs in prognosis-related co-expression RNA network were mainly enriched in cell cycle, mitotic cell cycle and nuclear division. Additionally, we constructed a prognostic scoring system for ESCC using ten signature RNAs (MLIP, TNFSF10, SIK2, LINC01068, LINC00601, TTTY14, AC084262.1, LINC01415, miR-5699-3p, miR-552-5p). Using this system, patients in the low-risk group had better long-term survival than those in the high-risk group (log-rank, P-value < 0.0001). The area under the ROC curve (AUCs) revealed that the accuracy of the prediction model was higher than the accuracy of single type of RNA prediction model.

Conclusion. In brief, we constructed a prognostic scoring system based on multiple types of RNA for ESCC that showed high predicting prognosis performance, and deeply understood the regulatory mechanism of prognosis-related lncRNA-mRNA co-expression network.

Corresponding author
Xiaozhi Zhang,
zhangxiaozhi@xjtu.edu.cn

INTRODUCTION

Esophageal cancer is one of the most common malignancies, ranking 7th in global morbidity and 6th in cancer-related mortality (*Bray et al., 2018*). The pathological types of esophageal cancer are mainly squamous cell carcinoma and adenocarcinoma. Esophageal squamous cell carcinoma (ESCC) is prevalent in Asia, Africa, and South America, especially in China, where ESCC accounts for more than 90% (*Malhotra et al., 2017*). The main treatments for esophageal cancer include surgical resection, radiotherapy and chemotherapy. Although progress has been made in the diagnosis and treatment of esophageal cancer, the 5-year overall survival rate is only about 15–20% (*Chen et al., 2016*; *Fitzmaurice et al., 2015*; *Gavin et al., 2012*). At present, the gold standard of tumor treatment and survival prediction is still tumor node metastasis (TNM) staging system, but there are some limitations in clinical application (*Amin et al., 2017*). TNM staging can only include categorical variables such as tumor, lymph node or metastasis, while neglecting other important prognostic variables, such as genomics or transcriptome differences. TNM staging also is difficult to explain why patients of the same stage have different clinical outcomes after the same treatment, that is, it cannot distinguish individual differences in patients with the same stage. Therefore, it is necessary to establish a genome or transcriptome based prognostic score system to predict the clinical prognosis of individual patients more accurately.

According to some estimates, about 70% of the human genome is transcribed into RNA, the portion of the genome which codes for proteins is only about 2% (*Birney et al., 2007*; *Esteller, 2011*). In recent decades, protein coding genes and non-coding RNAs have been confirmed to play key roles in tumorigenesis and tumor progession. For ESCC, researchers have identified multiple driving genes, including TP53, NOTCH1, FAM135B, EP300, and TET2, and the mutation status of FAM135B, EP300 and TET2 are associated with the prognosis of patients (*Gao et al., 2014*; *Sawada et al., 2016*; *Song et al., 2014*). *Wen et al. (2019)* analyzed the expression profile of small non-coding RNAs in 145 ESCC samples, and established a prediction model composed of four-miRNAs, which was used to predict overall survival in LN-positive locoregional ESCC patients. *Sun et al. (2015)* analyzed the expression of GASC1-targeted gene in 149 tumor specimens from patients with ESCC, and identified a prediction model composed of three-gene (PPARG, MDM2, and NANOG), which may serve as a predictor for the poor prognosis of ESCC patients. *Li et al. (2016)* conducted whole-genome sequencing analysis of lncRNA expression in 12 ESCC tumor and normal tissues, and constructed a co-expression network composed of 119 differentially expressed lncRNA and 1350 correlated mRNAs to reveal the potential mechanism of ESCC. However, the individualized prognosis prediction model based on multiple types of RNA has not been reported in ESCC, and the prognosis-related lncRNA-mRNA co-expression network is lacking.

In this study, we comprehensively analyzed the expression and clinical data of lncRNAs, miRNAs and mRNAs of ESCC in the TCGA database. Using multivariate Cox regression analysis, we constructed a prognostic scoring system based on multiple types of RNA that divided ESCC patients into two groups (high-risk and low-risk) with a significant difference in overall survival (OS). The accuracy of the prognostic scoring system was higher than the accuracy of single type of RNA prediction model. Besides, we constructed a prognosis-related lncRNA-mRNA co-expression network in ESCC and the potential molecular mechanisms of prognostic mRNAs were explored by functional enrichment analyses. The presented analysis, we aim to provide novel clues for effective prediction of clinical outcomes.

MATERIAL AND METHODS

Data collection and pretreatment

The sequencing data (RNA-sequencing and miRNA-sequencing) and clinical information of ESCC patients were obtained from the TCGA database (https://portal.gdc.cancer.gov/). Based on the annotation file (Homo_sapiens.GRCh38.95.chr.gtf) downloaded from the Ensembl database (http://asia.ensembl.org/info/data/ftp/index.html), we identified 19876 lncRNAs and 19645 protein-coding genes. At the same time, we identified 2069 miRNA according to the annotation file (mature.fa) downloaded from miRBase database (http://mirbase.org/ftp.shtml). LncRNAs, mRNAs and miRNAs expressing raw count value >1 were screened for subsequent operation. This study was in line with the published guidelines provided by TCGA (https://cancergenome.nih.gov/publications/publicationguidelines). Since our data was obtained from the TCGA database, no ethics committee approval was required.

Differentially expressed analysis

The analysis and extraction of differentially expressed lncRNAs and mRNAs between 81 tumor tissues and 11 normal tissues were conducted by using the edgeR package of R language (*Robinson, McCarthy & Smyth, 2010*; *R Core Team, 2013*). Similarly, the differentially expressed miRNAs between 95 tumor tissues and 13 normal tissues were analyzed and extracted using edgeR package. $|log2FC| > 2$ and $FDR < 0.05$ (FC, fold change; FDR, false discovery rate) were considered to be significant. After edgeR normalization, log2 (normalized value $+1$) transformation was performed on the expression profiles of miRNAs, mRNAs and lncRNAs for subsequent manipulation.

Survival analysis

The clinical datasets of the ESCC cohort were downloaded from TCGA. Samples with a survival time of $t = 0$ days were removed to avoid introducing more mixed factors, and the remaining 80 samples were retained for the survival analysis. The clinical and pathological characteristics of the remaining 80 samples are summarized in Table 1. Univariate Cox regression analysis was used in R software to evaluate whether lncRNA, miRNA and mRNA were correlated with OS. RNAs with $P < 0.05$ were screened as prognostic biomarkers. RNAs with hazard ratio (HR) <1 were defined as protective signature, while RNAs with HR for death > 1 were defined as risky RNAs.

Table 1 Clinical characteristics of 80 patients with esophageal squamous cell carcinoma.

Characteristics	Number	Percent (%)
Gender		
Male	68	85
Female	12	15
Age (years)		
≤58	46	57.5
>58	34	42.5
Histologic grade		
G1	15	18.75
G2	38	47.5
G3	18	22.5
G_X	9	11.25
Tumor stage		
T1	7	8.75
T2	29	36.25
T3	40	50
T4	4	5
Node stage		
N0	42	52.5
N1 + N2	30	37.5
N_X	8	10
Metastasis stage		
M0	70	87.5
M1	5	6.25
M_X	5	6.25
Pathologic stage		
I	7	8.75
II	46	57.5
III	22	27.5
IV	4	5
–	1	1.25
Survival status		
No	24	30
Yes	56	70

LncRNA-mRNA co-expression network

The correlation between prognostic lncRNA and mRNA expression profiles was analyzed by Spearman method, and the lncRNA-mRNAs pairs that the absolute value of correlation coefficients >=0.4 and $p < 0.05$ were selected to construct the co-expression network. The co-expression network result was displayed using Cytoscape software version 3.6.0 (https://cytoscape.org/) (*Shannon, 2003*). CytoHubba, a plugin in the Cytoscape software, was adopted to calculate the degree of each node and select modules of hub genes from the network (*Chin et al., 2014*).

Functional enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional and pathway enrichment analysis were performed for mRNAs in prognosis-related co-expression RNA network using the Database for Annotation, Visualization and Integrated Discovery bioinformatics resources (DAVID) (https://david-d.ncifcrf.gov/), and $P < 0.05$ was considered as the cut-off criterion to screen the Enriched terms and pathways (*Altermann & Klaenhammer, 2005*; *Ashburner et al., 2000*; *Huang, Sherman & Lempicki, 2009a*; *Huang, Sherman & Lempicki, 2009b*).

Prognostic scoring system

RNAs with univariate Cox regression $P < 0.01$ were selected for the stepwise Cox regression procedures. Akaike information criterion (AIC) was used to evaluate the relative goodness of fitted model. Furthermore, Multivariate Cox regression coefficient was multiplied by the expression level of independent biomarkers ($P < 0.05$) to construct prognostic Cox models of lncRNA, miRNA and mRNA, respectively. Finally, a prognostic scoring system in ESCC was constructed, based on above-described multiple types of RNA. Receiver operating characteristics curves (ROC) and area under ROC curves (AUCs) were applied to evaluate the efficiency of each model. Statistical computing was performed using R software version 3.5.2. A flow diagram of the prognostic scoring system is presented in Fig. 1.

Statistical analysis

The statistical analyses in the present study were conducted by SPSS Statistics 18.0 and R software version 3.5.2. P value <0.05 was defined as statistically significance. Univariate Cox and multivariate Cox regression analyses were used to identify prognostic biomarkers. Survival curves were plotted by Kaplan–Meier (K-M) analysis, and differences in survival rates were assessed using a log-rank test.

RESULTS

Differentially expressed lncRNA, miRNA and mRNA

Analysis of expression profiles in ESCC compared with normal esophageal tissues identified a total of 1662 lncRNAs, 79 miRNAs and 2063 mRNAs (Table S1). Among them, 818 and 844 lncRNAs were respectively up-regulated and down-regulated (Fig. 2A); 52 miRNAs were up-regulated, and 27 were down-regulated (Fig. 2B); 869 up-regulated mRNAs and 1196 down-regulated mRNAs were obtained (Fig. 2C). Expression heatmaps were constructed by the top 50 up-regulation and the top 50 down-regulation to visualize the most significant lncRNAs, miRNAs and mRNAs (Fig. S1). The heatmap of the lncRNAs (Fig. S1A), miRNAs (Fig. S1B) and mRNAs (Fig. S1C) showed that the tumors clustered separately from the normal tissues.

Prognostic lncRNAs, miRNAs, mRNAs and co-expression network

Using univariate Cox regression analysis on the remaining 80 samples with survival times >0, 62 prognostic lncRNAs, eight prognostic miRNAs, and 66 prognostic mRNAs were identified in ESCC (P-value <0.05) (Table S2). The prognostic lncRNAs and mRNAs
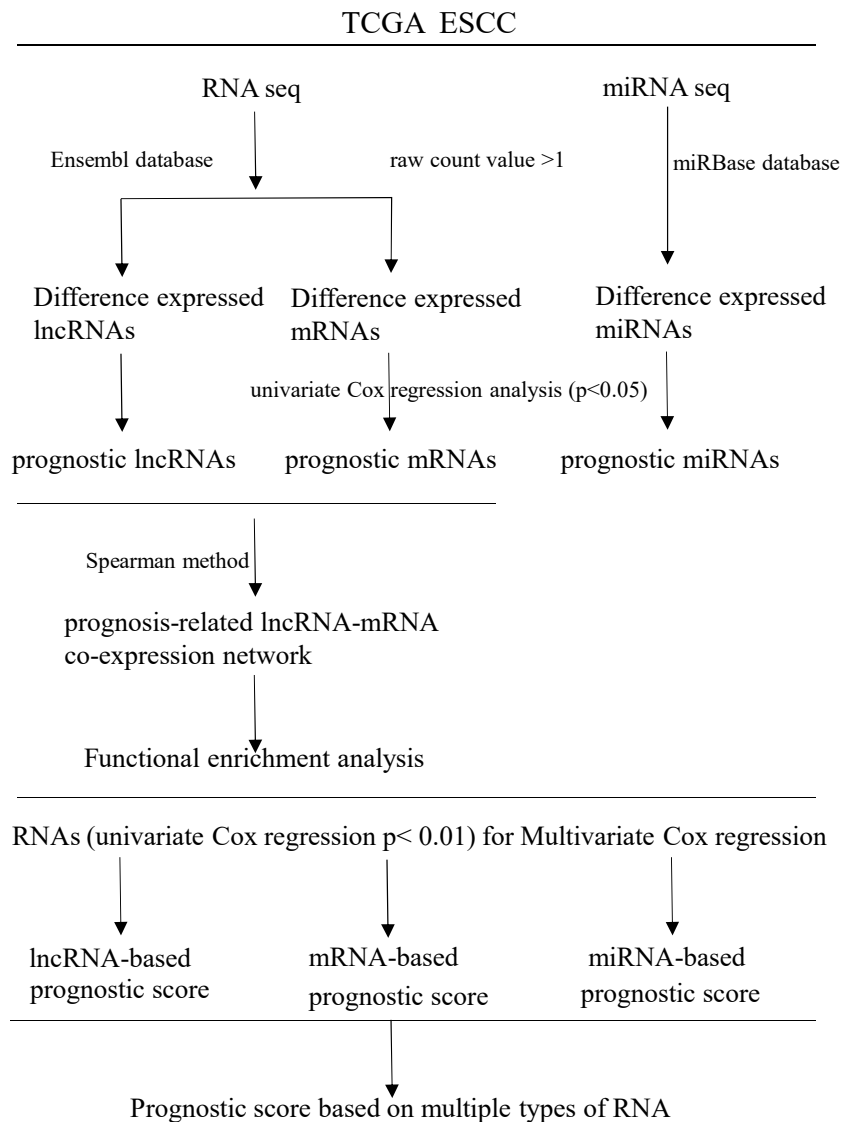
Figure 1 Flow diagram of the prognostic scoring system construction.

in ESCC were used to generate the co-expression network consisting of 22 lncRNAs, 40 mRNAs, and 77 interaction pairs (Fig. 3) (Table S3). Cytoscape analysis of the co-expression network revealed the top five prognostic RNAs (CDCA2, MTBP, CENPE, PBK, AL033384.1) (Table 2). Based on the median expression of each top 5 RNAs, 80 ESCC patients were divided into two groups (high expression vs low expression). The prognostic value of RNA was demonstrated by K-M plots (Fig. 4).

We then performed GO functional enrichment analysis of mRNAs in prognosis-related co-expression RNA network (Fig. 5). The results showed that the prognostic mRNAs mainly enriched in biological process (BP) including cell cycle, mitotic cell cycle and nuclear division. Cellular component (CC) analysis indicated enrichment in intracellular
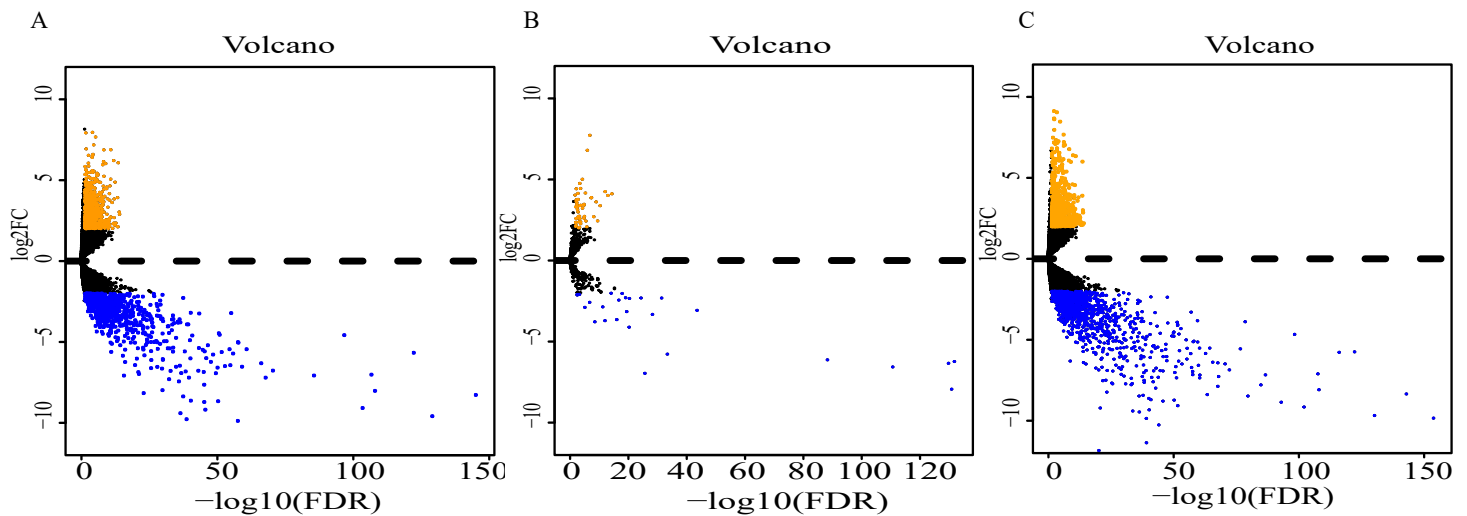
Figure 2 **Volcano plot of differentially expressed RNAs between ESCC and normal tissues.** (A) lncRNAs; (B)miRNAs; (C) mRNAs. Orange dots indicate upregulated RNAs, while blue dots indicate downregulated RNAs with statistical significance.

Table 2 **Top five in the prognosis-related co-expression RNA network ranked by degree method.**

Rank	Name	Score	HR
1	CDCA2	13	0.508863
2	MTBP	12	0.566911
3	CENPE	10	0.63653
4	PBK	6	0.627698
5	AL033384.1	5	1.460057

non-membrane-bounded organelle, non-membrane-bounded organelle and cytoskeletal part. Besides, in the molecular function (MF), the mRNAs were significantly clustered into purine nucleotide binding, ribonucleotide binding and ATP binding terms (Table S4). No pathways were significantly enriched in the KEGG enrichment analysis of prognostic mRNAs.

Prognostic scoring system

To create prognostic scoring system, RNAs with univariate Cox regression $P < 0.01$ were selected for the stepwise Cox regression procedures. Next, based on 5 lncRNAs, 2 miRNAs and 3 mRNAs respectively, we constructed three prediction models with single type of RNA to calculated the risk scores for predicted survival (Table S5). The formulas for the three prognostic models were as follows: lncRNA-based prognostic score = (0.447 × expression level of LINC01068) + (0.3677 × expression level of LINC00601) + (0.3075 × expression level of TTTY14) + (−0.8750 × expression level of AC084262.1) + (−0.4744 × expression level of LINC01415); miRNA-based prognostic score = (1.2932 × expression level of miR-5699-3p) + (0.7202 × expression level of miR-552-5p); mRNA-based prognostic score = (0.5139 × expression level of MLIP) + (0.5746 × expression level of TNFSF10) +
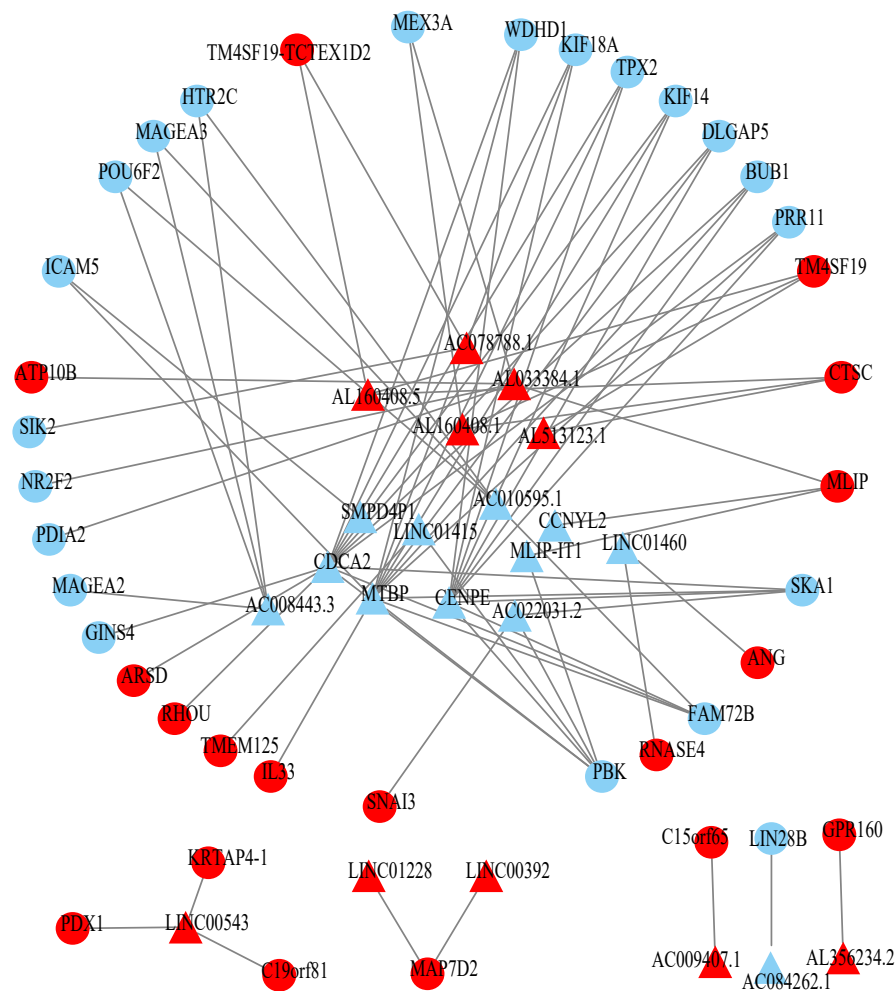
Figure 3 Prognosis-related co-expression RNA network in ESCC. The red represents the risky RNAs, and the blue represents the protective RNAs in ESCC. The triangle indicates lncRNAs and circle indicates mRNAs.

($-1.0069 \times$ expression level of SIK2). Of three prognostic models, seven RNAs were shown to be risky RNAs (LINC01068, LINC00601, TTTY14, miR-5699-3p, miR-552-5p, MLIP, TNFSF10, HR >1) and three RNAs were the protective RNAs (AC084262.1, LINC01415, SIK2, HR <1) (Figs. 6A–6C).

Using these three formulas, we calculated the prognostic score for each of the 80 patients separately and ranked them according to the increased prognostic scores. we divided the ESCC patients into two group (high-risk or low-risk) using the median prognostic score as a cutoff. As shown in Figs. 6D–6F, patients in the high-risk group had a worse prognosis than the low-risk group in all three models ($P < 0.0001$). We also used ROC curves to estimate the specificity and sensitivity of these prognostic models. All three prognostic models showed moderate prognostic evaluation ability, with AUC of 1 year values of 0.855, 0.859 and 0.785, separately, and AUC of 3 year values of 0.909, 0.709 and 0.762,
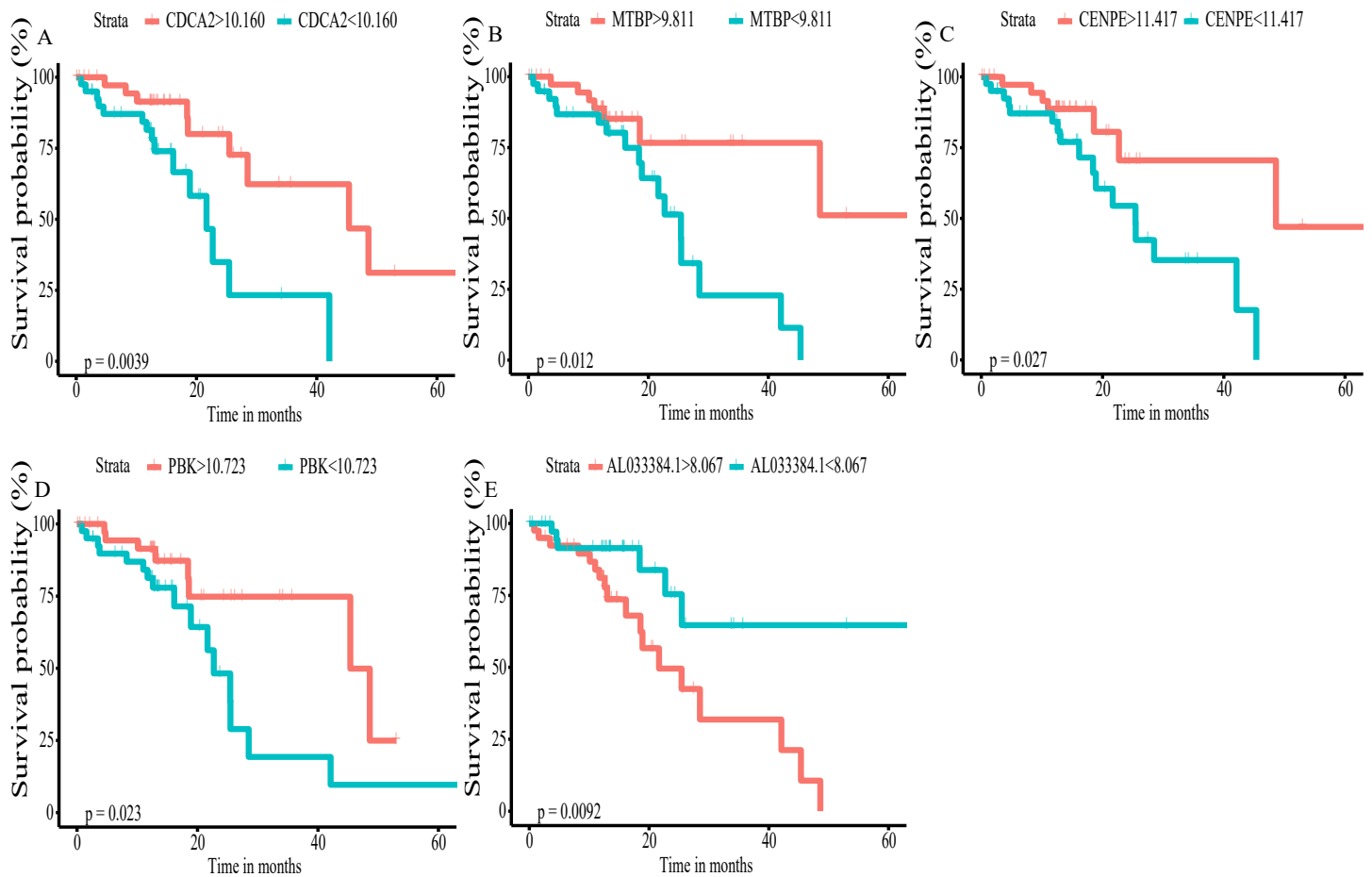
Figure 4 **Kaplan–Meier (K–M) survival curves for top5 RNAs in the prognosis-related co-expression RNA network.** (A) CDCA2; (B) MTBP; (C)CENPE; (D) PBK; (E) AL033384.1.

separately (Fig. 7). Figure 8 shows the distribution of patient prognostic scores, the survival status and tumor RNAs expression of all 80 ESCC patients. Patients in the high-risk group had more deaths than those in the low-risk group in all three models (Figs. 8D–8F). Moreover, patients in the low-risk group tend to express protective RNA, while patients in the high-risk group tend to express risky RNA (Figs. 8G–8I).

In order to improve the prediction accuracy and understand the potential molecular mechanism of prognostic markers, we constructed a prognostic Cox model with multiple types of RNA for ESCC using the ten RNAs provided above (Table S6). The formula was as follows: RNA-based prognostic score = (0.42895 × expression level of LINC01068) + (0.34829 × expression level of LINC00601) + (0.2185 × expression level of TTTY14) + (−1.393 × expression level of AC084262.1) + (−0.33364 × expression level of LINC01415) + (1.06024 × expression level of miR-5699-3p) + (0.34784 × expression level of miR-552-5p) + (0.3418 × expression level of MLIP) + (0.05437 × expression level of TNFSF10) + (−1.38365 × expression level of SIK2). The ten RNAs forest plots of RNA-based prognostic model was presented in Fig. 9. K-M analysis showed patients in the low-risk group had
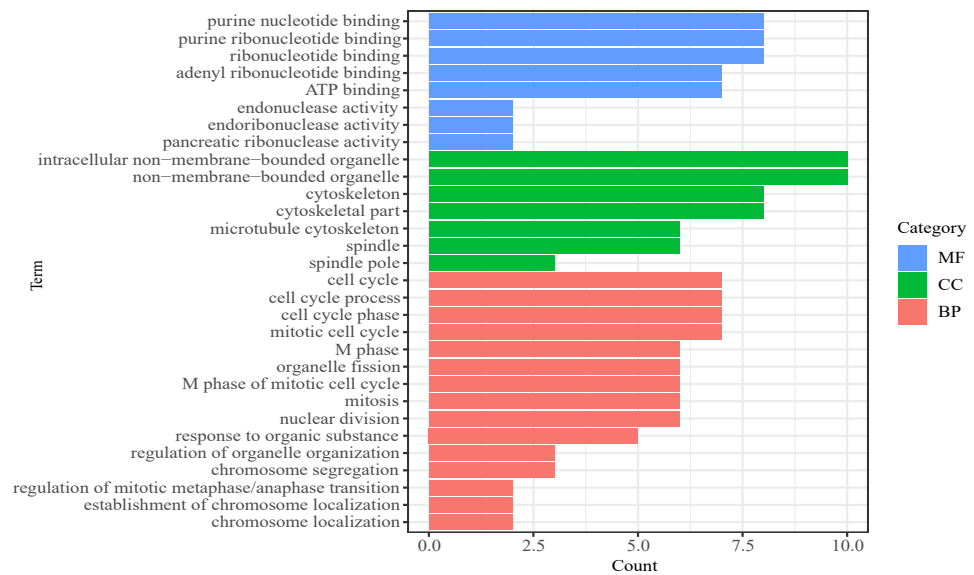
Figure 5 Gene Ontology (GO) analysis of mRNA in the prognosis-related co-expression RNA network.

better long-term survival than those in the high-risk group (*P*-value <0.0001; Fig. 10A). Furthermore, the AUC of 1 year value was 0.916 and 3 year value was 0.917 (Fig. 10B), indicating that the combination of different types of RNA patterns is a more accurate prognostic model than single type of RNA prediction model. For tumor staging, we also generated K-M plots and corresponding ROC curves. ESCC patients were also divided into two groups by tumor, node and metastasis (TNM) stage, and the prognosis of the two groups was different (*P*-value <0.05; Fig. 10C). The AUC of 1 year value and 3 year value based on TNM staging were 0.612 and 0.548, respectively (Fig. 10D). Although TNM staging is often used in clinical prognostic prediction, its prognostic AUC value is limited. Besides, combining multiple clinical parameters, we performed cox regression analysis of the prognostic score. As shown in Table 3, in both univariate Cox and multivariate Cox regression analysis, prognostic score was significantly correlated with survival (*P* < 0.001), that is, prognostic score was an independent prognostic factor in ESCC patients.

DISCUSSION

ESCC is one of the leading causes of cancer-associated mortality worldwide. Several studies have shown that lncRNAs, miRNAs, and mRNAs can be powerful prognostic factors in multiple cancers, including ESCC. MALAT1 has been identified as an important predictor of survival in ESCC (*Hu et al., 2015*). *Luo & Wu (2019)* verified that miR-375 may be a new prognostic marker of ESCC by meta-analysis. *Pan et al. (2014)* measured FOXCUT/FOXC1 in 82 ESCC tissues and adjacent noncancerous tissues by real-time quantitative PCR (qPCR), and found patients with upregulated FOXCUT or FOXC1 experienced a significantly worse prognosis than those with downregulated FOXCUT or
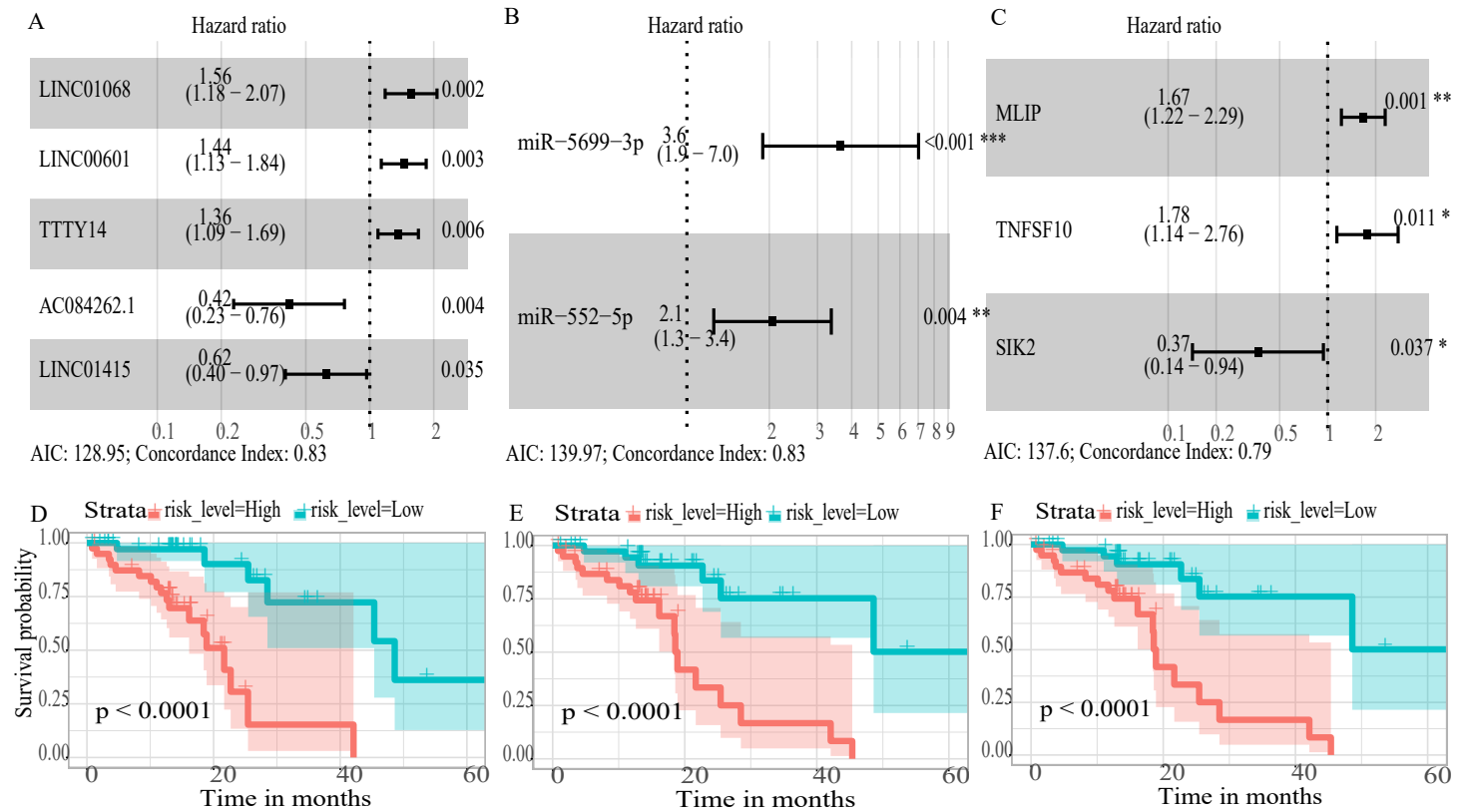
Figure 6 Forest plots of hazard ratios (HR) of the RNAs and Kaplan-Meier curves for overall survival (OS) of high-risk and low-risk patients based on prognostic scores in the TCGA ESCC cohort. (A) lncRNAs; (B)miRNAs; (C) mRNAs; (D) lncRNA-based prognostic score; (E) miRNA-based prognostic score; (F) mRNA-based prognostic score.
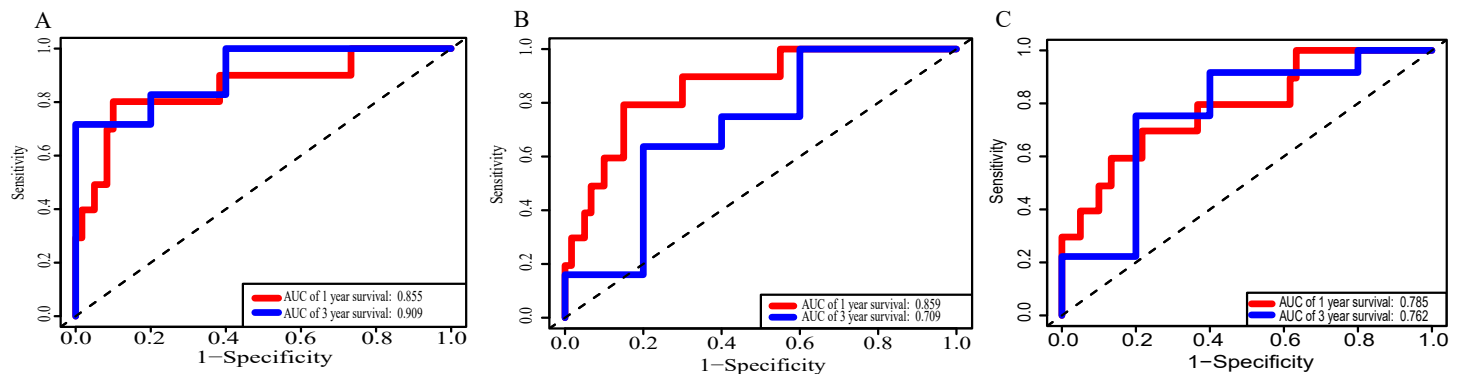
Figure 7 Receiver operating characteristic curves for survival prediction by prognostic score in TCGA ESCC cohort. (A) lncRNA-based prognostic score; (B) miRNA-based prognostic score; (C) mRNA-based prognostic score. The red indicates the AUC of 1 years survival and the blue indicates the AUC of 3 years survival.
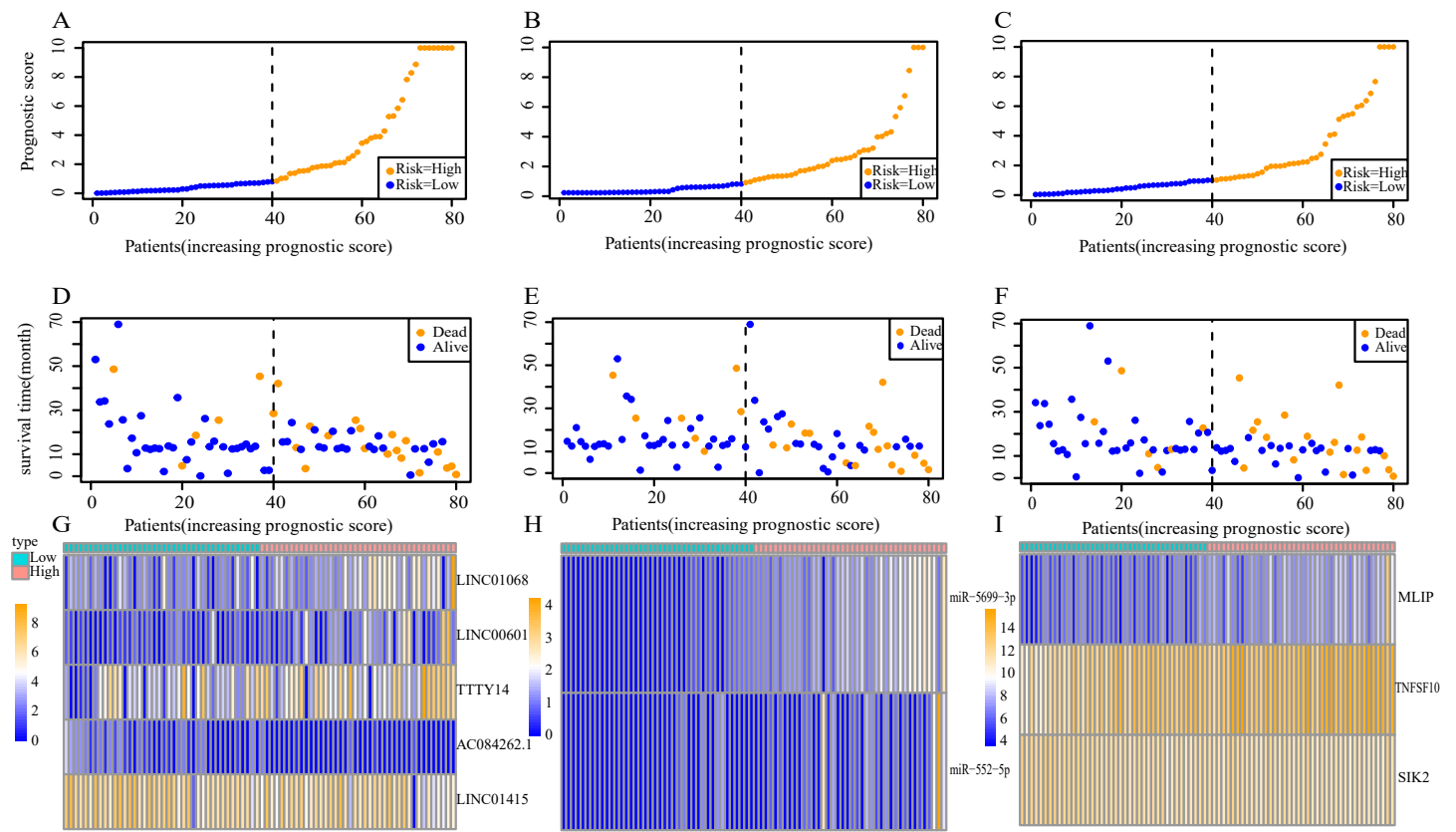
Figure 8 **Prognostic scores, survival and expression clustering heatmap of the signature RNAs of ESCC patients.** (A) Distribution of prognostic score based on lncRNA. (B) Distribution of prognostic score based on miRNA. (C) Distribution of prognostic score based on mRNA. (D) Distribution of patient survival status based on lncRNA prognostic scores. (E) Distribution of patient survival status based on miRNA prognostic scores. (F) Distribution of patient survival status based on mRNA prognostic scores. (G–I) Expression clustering heatmap of the signature RNAs.

FOXC1. However, the prediction models constructed in the previous studies mainly focus on one kind of RNA, which has limited prognostic efficacy.

In the present study, we comprehensively analyzed the expression data and clinical data of ESCC in the TCGA database, and identified 62 prognostic lncRNAs, 8 prognostic miRNAs, and 66 prognostic mRNAs. Using Cox regression analysis, We proposed three different prognostic models based on 5 lncRNAs, 2 miRNAs and 3 mRNAs respectively, which showed moderate prognostic assessment ability in predicting long-term survival of ESCC patients. Furthermore, a novel prognostic scoring system that included multiple types of RNA was proposed, which showed high predicting prognosis performance and was validated as an independent prognostic factor in ESCC patients. Of prognostic models, seven RNAs were shown to be risky RNAs (LINC01068, LINC00601, TTTY14, miR-5699-3p, miR-552-5p, MLIP, TNFSF10, HR >1) and three RNAs were the protective RNAs (AC084262.1, LINC01415, SIK2, HR <1).

A number of RNAs in the prognostic system used in the present study have been previously implicated in malignant tumors. TTTY14 (testis-specific transcript, Y-linked 14) was significantly correlated with overall survival for gastric cancer (GC) patients and
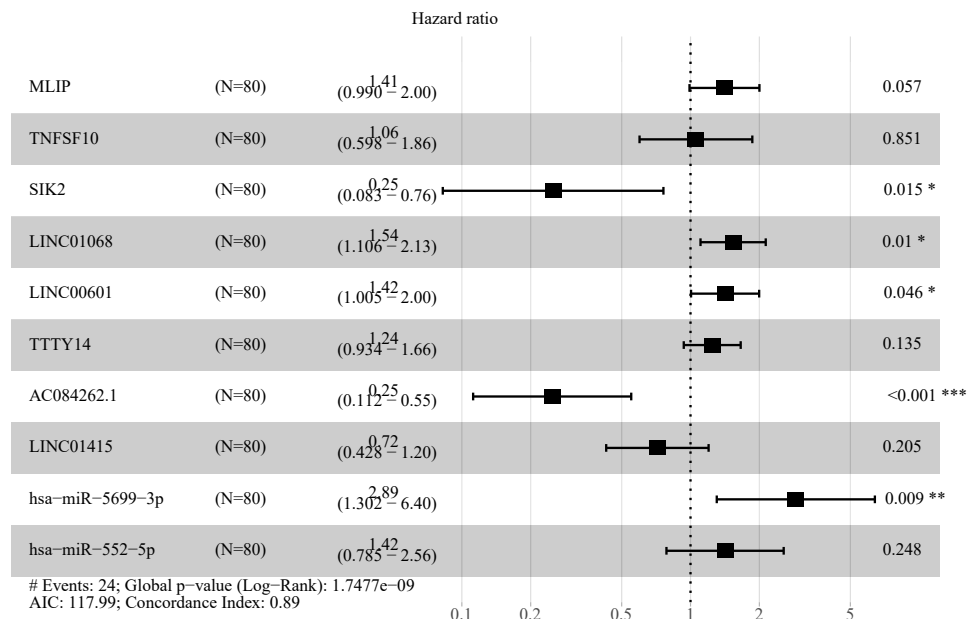
Figure 9 Forest plots of hazard ratios (HR) of the RNAs involved in prognostic scoring system based on multiple types of RNA.

oral squamous cell carcinoma (OSCC) patients and has been suggested to be involved in HPV (human papillomavirus)-Induced Oncogenesis (*Cheng et al., 2019*; *Goedert et al., 2016*; *Li et al., 2017*). miR-552-5p facilitates osteosarcoma cell proliferation and metastasis by targeting WIF1, which means miR-552-5p may become a new target for the treatment of osteosarcoma (*Cai et al., 2019*). TNFSF10 (TNF superfamily member 10), a cytokine that belongs to the tumor necrosis factor (TNF) ligand family,preferentially induces apoptosis in transformed and tumor cells, and TNFSF10 was significantly associated with overall survival in patients with liver cancer, breast cancer, non-small cell lung cancer and other tumors (*Koç Erbaşoğlu et al., 2019*; *McCarthy, 2005*; *Piras-Straub et al., 2015*). Frequent amplification of TNFSF10 was associated with the development and progression of esophageal cancer (*Chen et al., 2008*). SIK2 (salt inducible kinase 2) was a potential breast cancer suppressor, and compared with normal control, its expression level of breast cancer tissues and cell lines was reduced (*Maxfield et al., 2016*). However, functional studies of the other RNAs (LINC01068, LINC00601, AC084262.1, LINC01415, miR-5699-3p, MLIP) have not been reported in cancer research.

LncRNAs play an important role in a variety of biological processes (*Kornienko et al., 2013*). Accumulating evidence, suggesting that lncRNAs influence the expression of target gene by regulating the transcription and stability of target gene (*Batista & Chang, 2013*; *Tripathi et al., 2013*). LncRNA-mRNA co-expression network is an important way to analyze the function and regulation mechanism from a comprehensive perspective. We proposed a prognosis-related lncRNA-mRNA co-expression network in ESCC consisting of 22 lncRNAs, 40 mRNAs, and 77 interaction pairs. Five prognosis-related hub RNAs
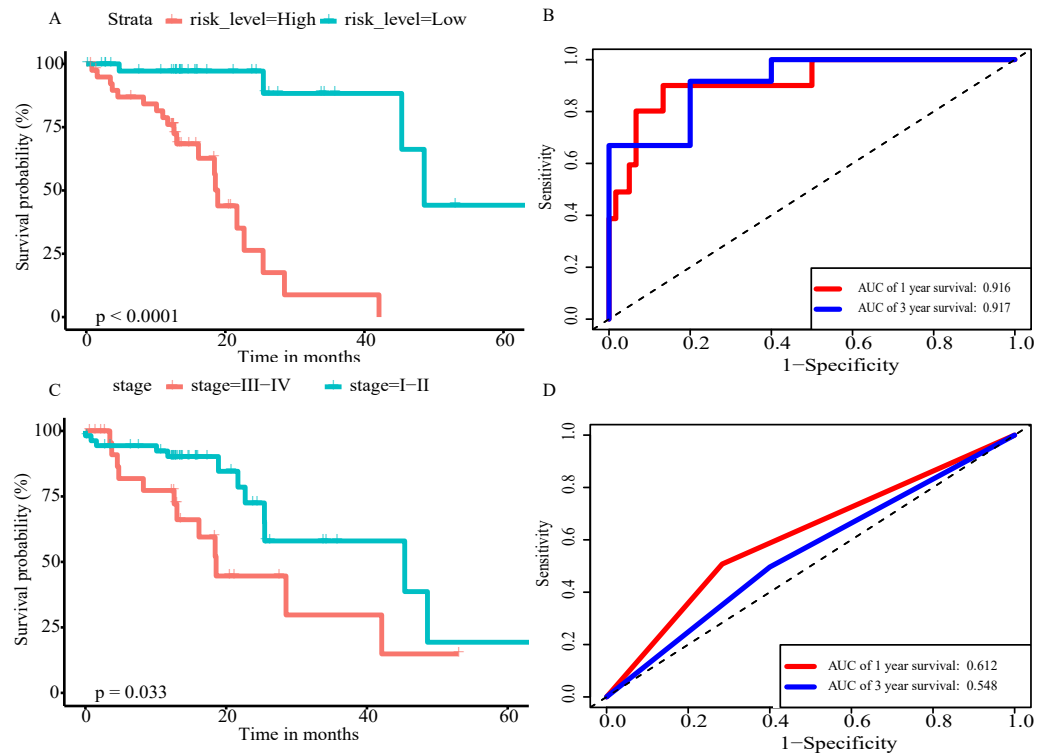
Figure 10 **The prognostic value of prognostic score and TNM stage in the TCGA datasets.** (A) Overall survival (OS) outcomes for high-risk and low-risk patients grouped by prognostic scoring system. (B) ROC curve with AUC of prognostic scoring system. (C) K–M plots of patients with different TNM staging. (D) ROC curve with AUC of TNM stage.

Table 3 **Univariate and multivariate Cox regression analysis of overall survival in ESCC.** Age, pathologic stage, tumor stage, histologic grade and prognostic score were continuous variable. Specifically, pathologic stage: I = 1, II = 2, III = 3, IV = 4; Tumor stage: T1 = 1,T2 = 2, T3 = 3, T4 = 4; Histologic grade: G1 = 1, G2 = 2, G3 = 3.

Variables	Univariate analysis		Multivariate analysis	
	Hazard radio (95% CI)	P	Hazard radio (95% CI)	P
Age	1.023 (0.965–1.086)	0.445	1.087 (0.969–1.221)	0.156
Gender (male/female)	0.033 (0.000–4.833)	0.180	0.000 (0.000–Inf)	0.973
Pathologic stage	1.868 (0.982–3.552)	0.057	7.133 (0.064–79.317)	0.110
Tumor stage	0.869 (0.427–1.768)	0.698	0.206 (0.035–1.218)	0.081
Node stage (N-/N+)	3.105 (1.077–8.953)	0.036	1.043 (0.184–5.908)	0.962
Metastasis stage (M-/M+)	3.431 (0.948–12.413)	0.060	0.004 (0.000–1.544)	0.069
Histologic grade	0.869 (0.418–1.807)	0.707	0.330 (0.088–1.234)	0.099
Prognostic score	1.090 (1.041–1.141)	<0.001	1.091 (1.034–1.15)	0.001

(CDCA2, MTBP, CENPE, PBK, AL033384.1) were identified and their prognostic value was verified by K-M plots.

Considering that mRNAs are the implementers of molecular function, GO enrichment analysis revealed that mRNAs in the prognosis-related co-expression RNA network were mainly enriched in cell cycle, mitotic cell cycle and nuclear division. Previous studies have shown that cell cycle pathway played an important role in the occurrence and development of esophageal squamous cell carcinoma (*Gao et al., 2014*; *Sanchez-Vega et al., 2018*), our observations were consistent with these results.

However, there were some limitations to this study, which should be considered when interpreting our results. First, in this study, only lncRNA, miRNA, and mRNA with both differential expression and prognostic value were included in the analysis. Therefore, the prognostic scoring system and co-expression network may not represent all molecular features that may be associated with ESCC overall survival. Second, several novel signature molecules with important prognostic significance in ESCC lack *in vivo* or *in vitro* experiments to determine their underlying molecular mechanisms. Finally, another limitation of the study was that the prognostic scoring system was not validated in another independent cohort.

CONCLUSIONS

In brief, we constructed a prognostic scoring system based on multiple types of RNA for ESCC that showed high predicting prognosis performance, and deeply understood the regulatory mechanism of prognosis-related lncRNA-mRNA co-expression network. These findings provide promising clues for effective prediction of clinical outcomes.

ACKNOWLEDGEMENTS

The authors would like to show thanks to TCGA project and its contributors for this valuable public data set.

Funding

This work was supported by the National Natural Science Foundation of China (No. 81773239). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
ational Natural Science Foundation of China: 81773239.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Xiaobo Shi and You Li conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Yuchen Sun and Xiaozhi Zhang conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Xu Zhao analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Xuanzi Sun and Zhinan Liang conceived and designed the experiments, prepared figures and/or tables, and approved the final draft.
- Tuotuo Gong and Yuan Ma analyzed the data, prepared figures and/or tables, and approved the final draft.

Data Availability

The clinical datasets are available from TCGA (TCGA_ESCA) and the expression profile of esophageal squamous cell carcinoma data are available in the Supplemental Tables.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.8368#supplemental-information.

REFERENCES

Altermann E, Klaenhammer TR. 2005. PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. *BMC Genomics* **6**:60 DOI 10.1186/1471-2164-6-60.

Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, Meyer L, Gress DM, Byrd DR, Winchester DP. 2017. The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *A Cancer Journal for Clinicians* **67**:93–99 DOI 10.3322/caac.21388.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: tool for the unification of biology. *Nature Genetics* **25**:25–29 DOI 10.1038/75556.

Batista PJ, Chang HY. 2013. Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152**:1298–1307 DOI 10.1016/j.cell.2013.02.012.

Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Stamatoyannopoulos JA, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland

GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SCJ, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Dutta A, Guigó R, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Flicek P, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermüller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung W-K, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Dermitzakis ET, Margulies EH, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei C-L, Ruan Y, Snyder M, Birney E, Struhl K, Gerstein M, Antonarakis SE, Gingeras TR, Brown JB, Flicek P, Fu Y, Keefe D, Birney E, Denoeud F, Gerstein M, Green ED, Kapranov P, Karaöz U, Myers RM, Noble WS, Reymond A, Rozowsky J, Struhl K, Siepel A, Stamatoyannopoulos JA, Taylor CM, Taylor J, Thurman RE, Tullius TD, Washietl S, Zheng D, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Collins FS, Margulies EH, Cooper GM, Asimenos G, Thomas DJ, Dewey CN, Siepel A, Birney E, Keefe D, Hou M, Taylor J, Nikolaev S, Montoya-Burgos JI, Löytynoja A. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**:799–816 DOI 10.1038/nature05874.

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *A Cancer Journal for Clinicians* **68**:394–424 DOI 10.3322/caac.21492.

Cai W, Xu Y, Yin J, Zuo W, Su Z. 2019. miR-552-5p facilitates osteosarcoma cell proliferation and metastasis by targeting WIF1. *Experimental and Therapeutic Medicine* **17**:3781–3788 DOI 10.3892/etm.2019.7361.

Chen J, Guo L, Peiffer DA, Zhou L, Chan OTM, Bibikova M, Wickham-Garcia E, Lu S-H, Zhan Q, Wang-Rodriguez J, Jiang W, Fan J-B. 2008. Genomic profiling of 766 cancer-related genes in archived esophageal normal and carcinoma tissues. *Indian Journal of Cancer* **122**:2249–2254 DOI 10.1002/ijc.23397.

Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. 2016. Cancer statistics in China, 2015. *CA: A Cancer Journal for Clinicians* **66**:115–132 DOI 10.3322/caac.21338.

Cheng C, Wang Q, Zhu M, Liu K, Zhang Z. 2019. Integrated analysis reveals potential long non-coding RNA biomarkers and their potential biological functions for disease

free survival in gastric cancer patients. *Cancer Cell International* **19**:Article 123 DOI 10.1186/s12935-019-0846-6.

Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. 2014. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology* **8**(**Suppl 4**):S11 DOI 10.1186/1752-0509-8-s4-s11.

Esteller M. 2011. Non-coding RNAs in human disease. *Nature Reviews Genetics* **12**:861–874 DOI 10.1038/nrg3074.

Fitzmaurice C, Dicker D, Pain A, Hamavid H, Moradi-Lakeh M, Macintyre MF, Allen C, Hansen G, Woodbrook R, Wolfe C, Hamadeh RR, Moore A, Werdecker A, Gessner BD, Te Ao B, McMahon B, Karimkhani C, Yu C, Cooke GS, Schwebel DC, Carpenter DO, Pereira DM, Nash D, Kazi DS, De Leo D, Plass D, Ukwaja KN, Thurston GD, Yun Jin K, Simard EP, Mills E, Park E-K, Catalá-López F, Deveber G, Gotay C, Khan G, Hosgood HD, Santos IS, Leasher JL, Singh J, Leigh J, Jonas JB, Sanabria J, Beardsley J, Jacobsen KH, Takahashi K, Franklin RC, Ronfani L, Montico M, Naldi L, Tonelli M, Geleijnse J, Petzold M, Shrime MG, Younis M, Yonemoto N, Breitborde N, Yip P, Pourmalek F, Lotufo PA, Esteghamati A, Hankey GJ, Ali R, Lunevicius R, Malekzadeh R, Dellavalle R, Weintraub R, Lucas R, Hay R, Rojas-Rueda D, Westerman R, Sepanlou SG, Nolte S, Patten S, Weichenthal S, Abera SF, Fereshtehnejad S-M, Shiue I, Driscoll T, Vasankari T, Alsharif U, Rahimi-Movaghar V, Vlassov VV, Marcenes WS, Mekonnen W, Melaku YA, Yano Y, Artaman A, Campos I, Maclachlan J, Mueller U, Kim D, Trillini M, Eshrati B, Williams HC, Shibuya K, Dandona R, Murthy K, Cowie B, Amare AT, Antonio CA, Castañeda Orjuela C, Van Gool CH, Violante F, Oh I-H, Deribe K, Soreide K, Knibbs L, Kereselidze M, Green M, Cardenas R, Roy N, Tillmann T, Li Y, Krueger H, Monasta L, Dey S, Sheikhbahaei S, Hafezi-Nejad N, Kumar GA, Sreeramareddy CT, Dandona L, Wang H, Vollset SE, Mokdad A, Salomon JA, Lozano R, Vos T, Forouzanfar M, Lopez A, Murray C, Naghavi M. 2015. The global burden of cancer 2013. *JAMA Oncology* **1**:505–527 DOI 10.1001/jamaoncol.2015.0735.

Gao Y-B, Chen Z-L, Li J-G, Hu X-D, Shi X-J, Sun Z-M, Zhang F, Zhao Z-R, Li Z-T, Liu Z-Y, Zhao Y-D, Sun J, Zhou C-C, Yao R, Wang S-Y, Wang P, Sun N, Zhang B-H, Dong J-S, Yu Y, Luo M, Feng X-L, Shi S-S, Zhou F, Tan F-W, Qiu B, Li N, Shao K, Zhang L-J, Zhang L-J, Xue Q, Gao S-G, He J. 2014. Genetic landscape of esophageal squamous cell carcinoma. *Nature Genetics* **46**:1097–1102 DOI 10.1038/ng.3076.

Gavin AT, Francisci S, Foschi R, Donnelly DW, Lemmens V, Brenner H, Anderson LA. 2012. Oesophageal cancer survival in Europe: a EUROCARE-4 study. *Cancer Epidemiology* **36**(**6**):505–512 DOI 10.1016/j.canep.2012.07.009.

Goedert L, Plaça JR, Nunes EM, Debom GN, Espreafico EM. 2016. Long noncoding RNAs in HPV-induced oncogenesis. *Advances in Tumor Virology* **6**:1–9 DOI 10.4137/ATV.S29816.

Hu L, Wu Y, Tan D, Meng H, Wang K, Bai Y, Yang K. 2015. Up-regulation of long noncoding RNA MALAT1 contributes to proliferation and metastasis in esophageal squamous cell carcinoma. *Journal of Experimental & Clinical Cancer Research* **34**:Article 7 DOI 10.1186/s13046-015-0123-z.

Huang DW, Sherman BT, Lempicki RA. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**:1–13 DOI 10.1093/nar/gkn923.

Huang DW, Sherman BT, Lempicki RA. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**:44–57 DOI 10.1038/nprot.2008.211.

Koç Erbaşoğlu Ö, Horozoğlu C, Ercan Ş, Kara HV, Turna A, Farooqi AA, Yaylım İ. 2019. Effect of trail C1595T variant and gene expression on the pathogenesis of non-small cell lung cancer. *Libyan Journal of Medicine* **14**:Article 1535746 DOI 10.1080/19932820.2018.1535746.

Kornienko AE, Guenzl PM, Barlow DP, Pauler FM. 2013. Gene regulation by the act of long non-coding RNA transcription. *BMC Biology* **11**:Article 59 DOI 10.1186/1741-7007-11-59.

Li S, Chen X, Liu X, Yu Y, Pan H, Haak R, Schmidt J, Ziebolz D, Schmalz G. 2017. Complex integrated analysis of lncRNAs-miRNAs-mRNAs in oral squamous cell carcinoma. *Oral Oncology* **73**:1–9 DOI 10.1016/j.oraloncology.2017.07.026.

Li Y, Shi X, Yang W, Lu Z, Wang P, Chen Z, He J. 2016. Transcriptome profiling of lncRNA and co-expression networks in esophageal squamous cell carcinoma by RNA sequencing. *Tumour Biology* **37**:13091–13100 DOI 10.1007/s13277-016-5227-3.

Luo HS, Wu DH. 2019. Identification of miR-375 as a potential prognostic biomarker for esophageal squamous cell cancer: a bioinformatics analysis based on TCGA and meta-analysis. *Pathology, Research and Practice* **215**:512–518 DOI 10.1016/j.prp.2019.01.009.

Malhotra GK, Yanala U, Ravipati A, Follet M, Vijayakumar M, Are C. 2017. Global trends in esophageal cancer. *Journal of Surgical Oncology* **115**:564–579 DOI 10.1002/jso.24592.

Maxfield KE, Macion J, Vankayalapati H, Whitehurst AW. 2016. SIK2 restricts autophagic flux to support triple-negative breast cancer survival. *Molecular and Cellular Biology* **36**:3048–3057 DOI 10.1128/mcb.00380-16.

McCarthy MM. 2005. Evaluating the expression and prognostic value of TRAIL-R1 and TRAIL-R2 in breast cancer. *Clinical Cancer Research* **11**:5188–5194 DOI 10.1158/1078-0432.ccr-05-0158.

Pan F, Yao J, Chen Y, Zhou C, Geng P, Mao H, Fang X. 2014. A novel long non-coding RNA FOXCUT and mRNA FOXC1 pair promote progression and predict poor prognosis in esophageal squamous cell carcinoma. *International Journal of Clinical and Experimental Pathology* **7**:2838–2849.

Piras-Straub K, Khairzada K, Trippler M, Baba HA, Kaiser GM, Paul A, Canbay A, Weber F, Gerken G, Herzer K. 2015. TRAIL expression levels in human hepatocellular carcinoma have implications for tumor growth, recurrence and survival. *Indian Journal of Cancer* **136(4)**:E154–E160 DOI 10.1002/ijc.29139.

R Core Team. 2013. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. *Available at https://www.R-project.org/*.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**:139–140 DOI 10.1093/bioinformatics/btp616.

Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, Dimitriadoy S, Liu DL, Kantheti HS, Saghafinia S, Chakravarty D, Daian F, Gao Q, Bailey MH, Liang W-W, Foltz SM, Shmulevich I, Ding L, Heins Z, Ochoa A, Gross B, Gao J, Zhang H, Kundra R, Kandoth C, Bahceci I, Dervishi L, Dogrusoz U, Zhou W, Shen H, Laird PW, Way GP, Greene CS, Liang H, Xiao Y, Wang C, Iavarone A, Berger AH, Bivona TG, Lazar AJ, Hammer GD, Giordano T, Kwong LN, McArthur G, Huang C, Tward AD, Frederick MJ, McCormick F, Meyerson M, Van Allen EM, Cherniack AD, Ciriello G, Sander C, Schultz N, Caesar-Johnson SJ, Demchok JA, Felau I, Kasapi M, Ferguson ML, Hutter CM, Sofia HJ, Tarnuzzer R, Wang Z, Yang L, Zenklusen JC, Zhang J, Chudamani S, Liu J, Lolla L, Naresh R, Pihl T, Sun Q, Wan Y, Wu Y, Cho J, Defreitas T, Frazer S, Gehlenborg N, Getz G, Heiman DI, Kim J, Lawrence MS, Lin P, Meier S, Noble MS, Saksena G, Voet D, Zhang H, Bernard B, Chambwe N, Dhankani V, Knijnenburg T, Kramer R, Leinonen K, Liu Y, Miller M, Reynolds S, Shmulevich I, Thorsson V, Zhang W, Akbani R, Broom BM, Hegde AM, Ju Z, Kanchi RS, Korkut A, Li J, Liang H, Ling S, Liu W, Lu Y, Mills GB, Ng K-S, Rao A, Ryan M, Wang J, Weinstein JN, Zhang J, Abeshouse A, Armenia J, Chakravarty D, Chatila WK, De Bruijn I, Gao J, Gross BE, Heins ZJ, Kundra R, La K, Ladanyi M, Luna A, Nissan MG, Ochoa A, Phillips SM, Reznik E, Sanchez-Vega F, Sander C, Schultz N, Sheridan R, Sumer SO, Sun Y, Taylor BS, Wang J, Zhang H, Anur P, Peto M, Spellman P, Benz C, Stuart JM, Wong CK, Yau C, Hayes DN, Parker JS, Wilkerson MD, Ally A, Balasundaram M, Bowlby R, Brooks D, Carlsen R, Chuah E, Dhalla N, Holt R, Jones SJM, Kasaian K, Lee D, Ma Y, Marra MA, Mayo M, Moore RA, Mungall AJ, Mungall K, Robertson AG, Sadeghi S, Schein JE, Sipahimalani P, Tam A, Thiessen N, Tse K, Wong T, Berger AC, Beroukhim R, Cherniack AD, Cibulskis C, Gabriel SB, Gao GF, Ha G, Meyerson M, Schumacher SE, Shih J, Kucherlapati MH, Kucherlapati RS, Baylin S, Cope L, Danilova L, Bootwalla MS, Lai PH, Maglinte DT, Van Den Berg DJ, Weisenberger DJ, Auman JT, Balu S. 2018. Oncogenic signaling pathways in the cancer genome atlas. *Cell* **173**:321–337 DOI 10.1016/j.cell.2018.03.035.

Sawada G, Niida A, Uchi R, Hirata H, Shimamura T, Suzuki Y, Shiraishi Y, Chiba K, Imoto S, Takahashi Y, Iwaya T, Sudo T, Hayashi T, Takai H, Kawasaki Y, Matsukawa T, Eguchi H, Sugimachi K, Tanaka F, Suzuki H, Yamamoto K, Ishii H, Shimizu M, Yamazaki H, Yamazaki M, Tachimori Y, Kajiyama Y, Natsugoe S, Fujita H, Mafune K, Tanaka Y, Kelsell DP, Scott CA, Tsuji S, Yachida S, Shibata T, Sugano S, Doki Y, Akiyama T, Aburatani H, Ogawa S, Miyano S, Mori M, Mimori K. 2016. Genomic landscape of esophageal squamous cell carcinoma in a Japanese population. *Gastroenterology* **150**:1171–1182 DOI 10.1053/j.gastro.2016.01.035.

Shannon P. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**:2498–2504 DOI 10.1101/gr.1239303.

Song Y, Li L, Ou Y, Gao Z, Li E, Li X, Zhang W, Wang J, Xu L, Zhou Y, Ma X, Liu L, Zhao Z, Huang X, Fan J, Dong L, Chen G, Ma L, Yang J, Chen L, He M, Li M, Zhuang X, Huang K, Qiu K, Yin G, Guo G, Feng Q, Chen P, Wu Z, Wu J, Ma L, Zhao J, Luo L, Fu M, Xu B, Chen B, Li Y, Tong T, Wang M, Liu Z, Lin D, Zhang X, Yang H, Wang J, Zhan Q. 2014. Identification of genomic alterations in oesophageal squamous cell cancer. *Nature* **509**:91–95 DOI 10.1038/nature13176.

Sun L-L, Wu J-Y, Wu Z-Y, Shen J-H, Xu X-E, Chen B, Wang S-H, Li E-M, Xu L-Y. 2015. A three-gene signature and clinical outcome in esophageal squamous cell carcinoma. *Indian Journal of Cancer* **136(6)**:E569–E577 DOI 10.1002/ijc.29211.

Tripathi V, Shen Z, Chakraborty A, Giri S, Freier SM, Wu X, Zhang Y, Gorospe M, Prasanth SG, Lal A, Prasanth KV. 2013. Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLOS Genetics* **9**:e1003368 DOI 10.1371/journal.pgen.1003368.

Wen J, Wang G, Xie X, Lin G, Yang H, Luo K, Liu Q, Ling Y, Xie X, Lin P, Chen Y, Zhang H, Rong T, Fu J. 2019. Prognostic value of a four-miRNA signature in patients with lymph node positive locoregional esophageal squamous cell carcinoma undergoing complete surgical resection. *Annals of Surgery* Epub ahead of print Apr 30 2019 DOI 10.1097/sla.0000000000003369.