Identification of a Prognostic Colorectal Cancer Model Including LncRNA FOXP4-AS1 and LncRNA BBOX1-AS1 Based on Bioinformatics Analysis

Zhi-Liang Shi,[1,2] Guo-Qiang Zhou,[2] Jian Guo,[2] Xiao-Ling Yang,[2] Cheng Yu,[2]
Cheng-Long Shen,[2] and Xin-Guo Zhu[1]

Abstract

Background: Knowledge about the prognostic role of long noncoding RNA (lncRNA) in colorectal cancer (CRC) is limited. Therefore, we constructed a lncRNA-related prognostic model based on data from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA).

Materials and Methods: CRC transcriptome and clinical data were downloaded from the GSE20916 dataset and the TCGA database, respectively. R software was used for data processing and analysis. The differential lncRNA expression within the two datasets was first screened, and then intersections were measured. Cox regression and the Kaplan–Meier method were used to evaluate the effects of various factors on prognosis. The area under the curve (AUC) of the receiver operating characteristic curve and a nomogram based on multivariate Cox analysis were used to estimate the prognostic value of the lncRNA-related model. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were applied to elucidate the significantly involved biological functions and pathways.

Results: A total of 11 lncRNAs were crossed. The univariate Cox analysis screened out two lncRNAs, which were analyzed in the multivariate Cox analysis. A nomogram based on the two lncRNAs and other clinicopathological risk factors was constructed. The AUC of the nomogram was 0.56 at 3 years and 0.71 at 5 years. The 3-year nomogram model was compared with the ideal model, which showed that some indices of the 3-year model were consistent with the ideal model, suggesting that our model was highly accurate. The GO and KEGG enrichment analyses showed that positive regulation of secretion by cells, positive regulation of secretion, positive regulation of exocytosis, endocytosis, and the calcium signaling pathway were differentially enriched in the two-lncRNA-associated phenotype.

Conclusions: A two-lncRNA prognostic model of CRC was constructed by bioinformatics analysis. The model had moderate prediction accuracy. LncRNA BBOX1-AS1 and lncRNA FOXP4-AS1 were identified as prognostic biomarkers.

Keywords: bioinformatics analysis, colorectal cancer, lncRNA, prognostic model

Introduction

In the United States, the incidence and mortality of colorectal cancer (CRC) both currently rank third among all cancers in both men and women.[1] In 2020 in the United Sates, 147,950 new cases of CRC will likely be diagnosed, and an estimated 53,200 CRC-related deaths are expected to occur.[1] In addition, the prevalence of CRC in young and middle-aged people (<50 years) is increasing.[2] Despite advances in treatment methods, many patients still face

[1]Department of General Surgery, The First Affiliated Hospital of Soochow University, Suzhou, China.
[2]Department of Gastrointestinal Surgery, Changshu No. 2 Hospital, Suzhou, China.

Address correspondence to: Xin-Guo Zhu; Department of General Surgery, The First Affiliated Hospital of Soochow University; 188 Shizi Street, Gusu District, Suzhou City, Suzhou 215006, Jiangsu Province, China
E-mail: **xgzhu45@163.com**

poor prognoses because of early metastasis, the absence of a typical clinical presentation, and the lack of sensitive screening methods for early-stage CRC.

In recent years, researchers have demonstrated the involvement of genome stability and aberrant gene expression in CRC prognosis.[3–6] Still, the mechanism of CRC survival remains unclear, which hampers efforts to improve CRC prognosis. Individualized systemic treatment may prolong survival and enhance quality of life. Therefore, an effective prediction model is critical for the accurate assessment of CRC prognosis.

Long noncoding RNAs (lncRNAs), whose transcripts are longer than 200 nucleotides, do not code for proteins.[7] They have recently been recognized as important regulators in tumor development and progression.[8,9] Research has shown that through interactions with RNAs, proteins, and lipids, lncRNAs play a pivotal role in mediating the signal transduction pathways of cancer,[10] suggesting their clinical potential as prognostic biomarkers and therapeutic targets.[11] Some studies have also revealed the roles of lncRNAs in cancer prognosis and progression.[12–16]

The Cancer Genome Atlas (TCGA) is a landmark cancer genomics program. It contains >20,000 primary cancer samples with gene information for 33 cancer types, including a valuable collection of multi-omics data regarding transcriptomes, DNA methylation, copy number variation, and other variables. In this study, we performed a global analysis of prognostic lncRNAs from the TCGA database and constructed a two-lncRNA prognostic model. Functional analyses of the two lncRNAs were performed. Furthermore, we developed and validated a predictive nomogram that integrated our newly discovered two-lncRNA signature with the traditional clinicopathological risk factors of CRC patients in the TCGA cohort. This two-lncRNA prognostic model and nomogram might help to more accurately predict CRC prognosis, and they may also help to guide postoperative treatment and follow-up.

Materials and Methods

Data acquisition

We used the search term ''colorectal cancer'' as the keyword in the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/), limiting the search range to ''Expression profiling by array'' and ''Homo sapiens.'' After screening, we selected the GSE20916 chip dataset from the GPL570 chip platform (Affymetrix Human Genome U133Plus 2.0 Array), which consisted of 46 cancer patients and 44 normal controls. We downloaded GSE20916 data for use with the GEOquery R package.

A dataset from the TCGA database, namely the COAD and READ projects, along with clinicopathological data, including information about 41 normal samples and 471 tumor samples, were downloaded from the Xena Functional Genomics Explorer of the University of California Santa Cruz. Since the information was retrieved from the TCGA database (a public database), ethical approval was not needed for this research. Data collection and processing were consistent with TCGA data policies for protecting human subjects (http://cancergenome.nih.gov/publications/publications guidelines).

Data processing and identification of differentially expressed lncRNAs

We re-annotated the GEO data into the lncRNA dataset using the SeqMap tool, removing the lncRNAs that could
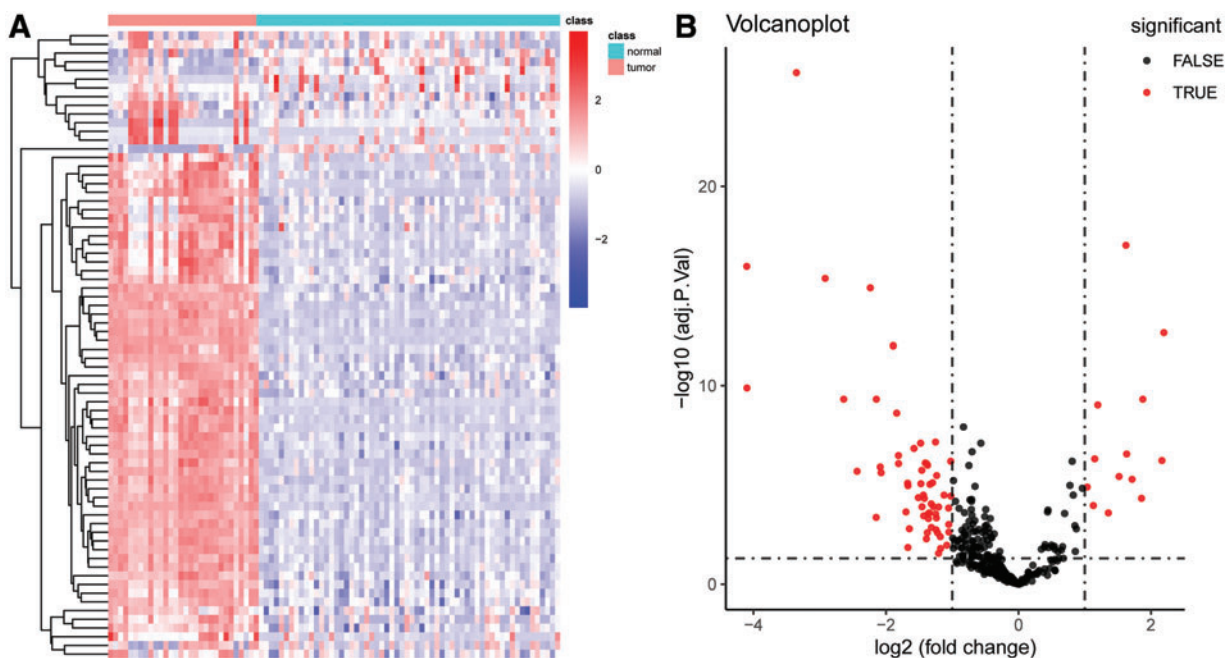

FIG. 1. **(A)** Heatmap of the DElncRNAs in GSE20916. **(B)** Volcano diagram of the DElncRNAs in GSE20916. DElncRNAs, differentially expressed long noncoding RNAs. Color images are available online.
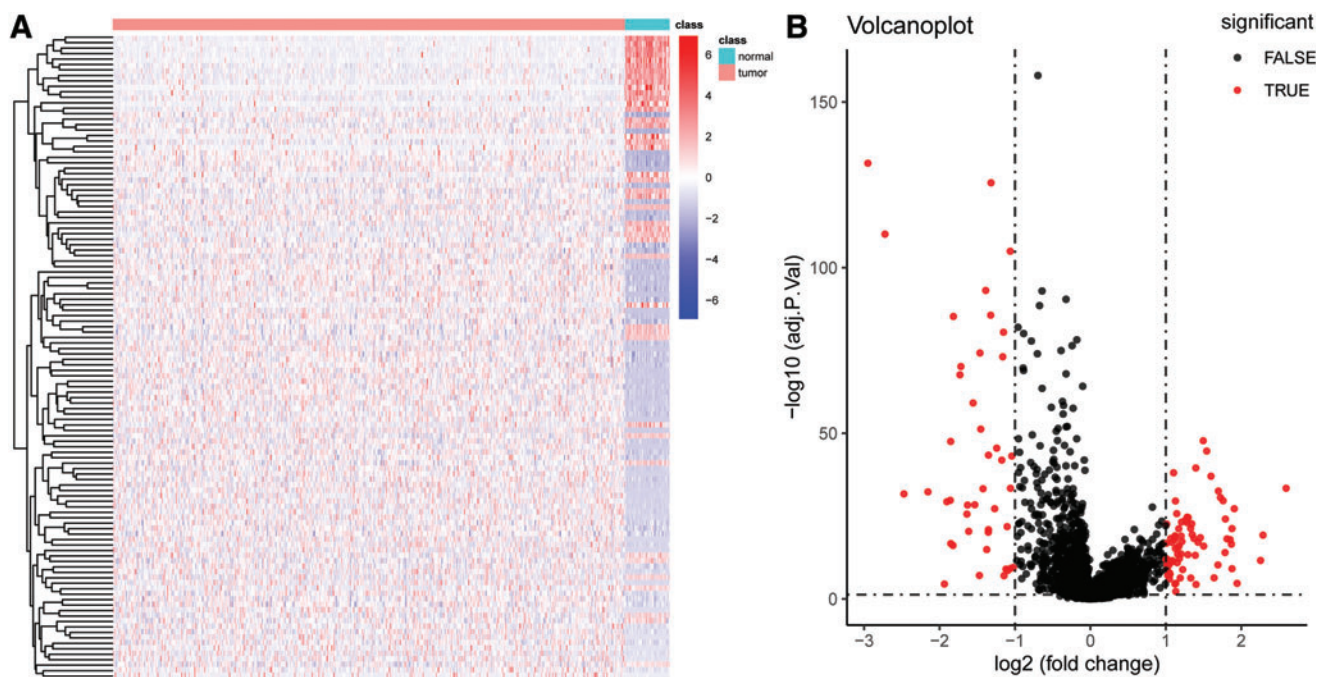
FIG. 2. **(A)** Heatmap of the DElncRNAs in the TCGA dataset. **(B)** Volcano diagram of the DElncRNAs in TCGA dataset. TCGA, The Cancer Genome Atlas. Color images are available online.

not locate chromosome position and a probe that corresponded to multiple lncRNAs. We noted the average expression of multiple probes corresponding to specific genes, and we changed the probe names to standard gene symbols. Finally, an expression matrix of 500 lncRNAs was obtained from the GEO database.

We then examined the clinical information of 471 tumor samples from the TCGA dataset and excluded the samples with missing clinical information, such as age, sex, survival time, survival status, pM, pT, and pN. Finally, 448 tumor samples were used for subsequent analysis. The RNA sequencing data in the fragments per kilobase per million (FPKM) format were converted to the transcripts per million (TPM) format for further analysis. The lncRNAs included 3prime_overlapping_ncRNA, antisense_RNA, bidirectional_promoter_lncRNA, lincRNA, macro_lncRNA, non_coding, processed_transcript, sense_intronic, and sense_overlapping. We downloaded the hg38 genome annotation file (Homo_
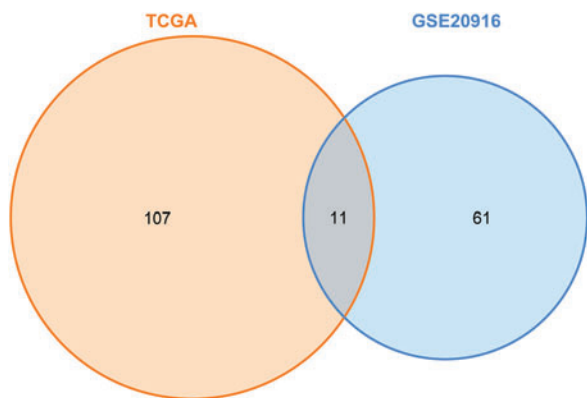

FIG. 3. Venn diagram of the GSE20916 and TCGA datasets. Color images are available online.

TABLE 1. THE CANCER GENOME ATLAS CLINICAL INFORMATION TRAINING AND VALIDATION DATASETS

Clinical characteristics	Test n = 224	Train n = 224	p
Age	66.8 (12.4)	67.1 (13.2)	0.510
pM			0.246
M0	170 (76.9%)	160 (72.7%)	
M1	25 (11.3%)	37 (16.8%)	
MX	26 (11.8%)	23 (10.5%)	
pN			0.405
N0	139 (62.1%)	125 (55.8%)	
N1	48 (21.4%)	56 (25.0%)	
N2	37 (16.5%)	43 (19.2%)	
pT			0.197
T1/T2	51 (22.9%)	38 (17.0%)	
T3	150 (67.3%)	156 (69.6%)	
T4	22 (9.87%)	30 (13.4%)	
Gender			0.57
Female	108 (48.2%)	101 (45.1%)	
Male	116 (51.8%)	123 (54.9%)	
Race			0.716
Black	27 (12.1%)	32 (14.3%)	
Others	83 (37.1%)	85 (37.9%)	
White	114 (50.9%)	107 (47.8%)	
OS			0.3
Alive	181 (80.8%)	171 (76.3%)	
Death	43 (19.2%)	53 (23.7%)	
OS.time	809 (688)	926 (792)	0.095

OS, overall survival.

Names	Hazard ratio	95% CI	p
MUC2	0.927	0.834–1.03	0.16
AC016027.1	0.408	0.136–1.227	0.111
SATB2-AS1	0.942	0.734–1.209	0.639
FOXP4-AS1	1.496	1.021–2.192	0.039*
DPP10-AS1	0.826	0.578–1.181	0.296
PVT1	1.089	0.696–1.703	0.709
BBOX1-AS1	1.403	1.029–1.913	0.032*
AC092718.4	0.97	0.637–1.477	0.887
DLGAP1-AS2	1.064	0.736–1.539	0.74
LINC01082	1.189	0.795–1.777	0.399
SNHG4	1.027	0.692–1.526	0.893

*$p < 0.05$.
CI, confidence interval.

sapiens.GRCh38.90.chr.gtf) based on the above set, and then determined the gene biotype. Next, we screened the lncRNA data based on the gene biotype and obtained an expression matrix of 14,376 lncRNAs from the TCGA dataset.

We used the limma package to analyze the differentially expressed lncRNAs (DELs) of the GEO and TCGA data. For the DEL threshold values, we used a |logFC| of >1 and an adjusted *p*-value of <0.05.

Definition and evaluation of the lncRNA-related prognostic model

We divided the TCGA dataset according to a 1:1 sample and obtained a training group (224 samples) and a validation group (224 samples). Then, we performed univariate Cox analysis to determine the association between lncRNA expression and overall survival (OS). The lncRNAs with *p*-values <0.05 were analyzed by multivariate Cox analysis.

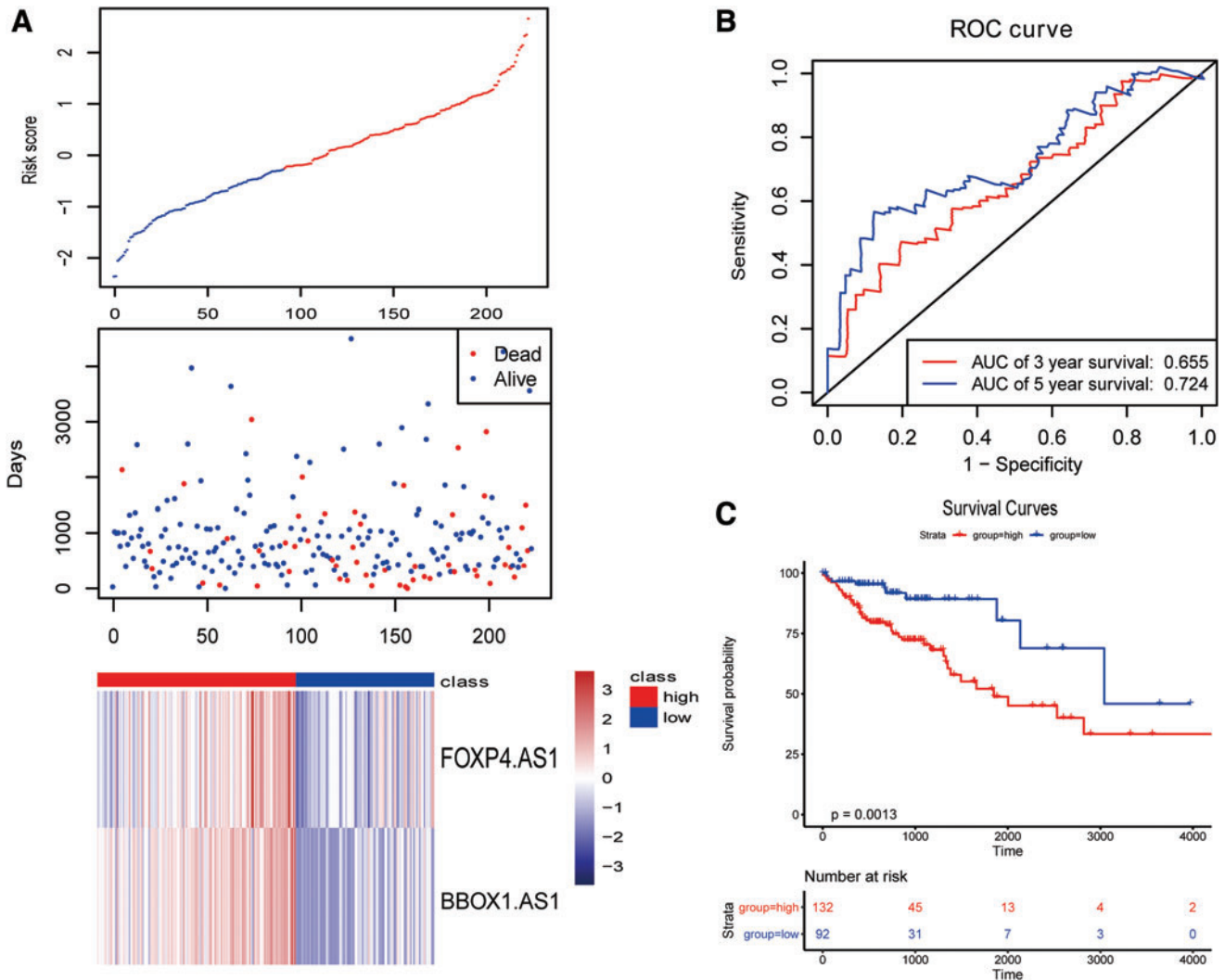The model was constructed in the training set using the following formula:

FIG. 4. **(A)** The risk score diagram in the training dataset and heatmap of the screened lncRNAs expression. **(B)** The ROC curve of the model for 3 and 5 years in the training dataset. **(C)** Survival curve of the training dataset. ROC, receiver operating characteristic. Color images are available online.

$$\text{Risk sore} = \sum_{i=1}^{N} (\exp * \text{coef.})$$

where N is the number of lncRNAs, exp is the lncRNA expression value, and coef is the messenger RNA (mRNA) coefficient in the multivariate Cox regression analysis. The testing set and whole set were used to validate the prediction accuracy of the model. The time-dependent receiver operating characteristic (ROC) curve was used to estimate the prediction accuracy of the model in terms of survival by calculating the area under the curve (AUC). We used the surv_cutpoint function in the survminer package to determine the best cutoff value according to the risk score and divided the patients into high- and low-risk groups. The Kaplan–Meier method was used to evaluate the survival difference between the high- and low-risk groups.

Univariate and multivariate Cox proportional hazard regression analyses were performed to assess whether the prognostic model was independent of other clinicopathological features (including age, sex, race, and risk score).

Construction and evaluation of the nomogram

To identify the predicted 3- and 5-year survival probabilities, a nomogram was constructed based on multivariate Cox analysis. Nomograms can predict the prognosis of a patient with cancer by simplifying the complicated prediction model into a profiled chart. A point scale was created to determine the points for the variables, and the sum of the points assigned to each variable was rescaled to a range from 0 to 100. Worse prognoses were represented by higher point totals. The calibration curves were graphically estimated by mapping the predicted probabilities of the nomogram against the actual observed rates. A concordance index (C-index) was used to assess the discrimination of the nomogram. The prediction accuracy was compared between the nomogram and separate prognostic factors using the C-index and ROC analyses.

Co-expression regulatory network and functional analysis

We used gene coexpression to predict the lncRNA-associated target genes in the TCGA training set. The filter
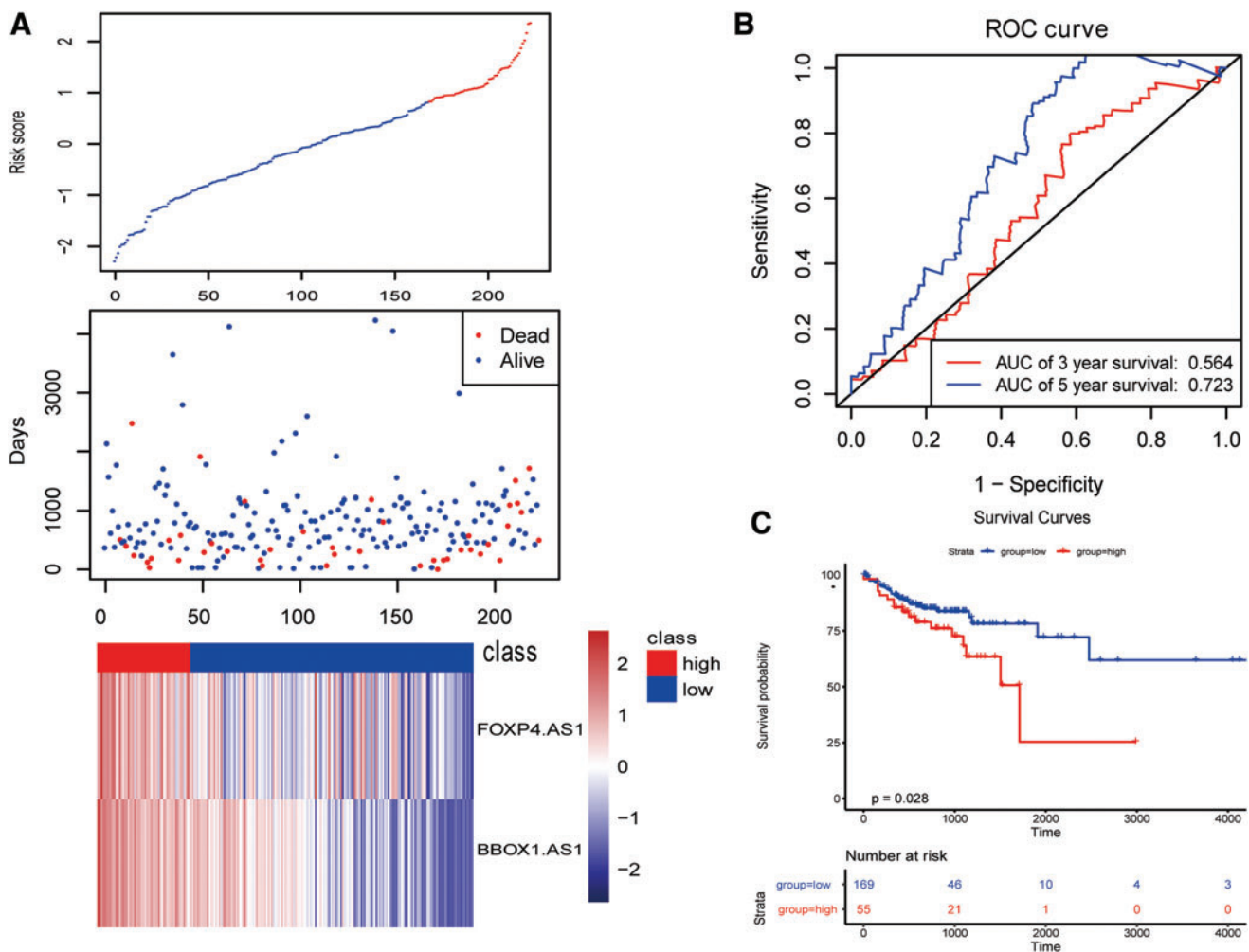

FIG. 5. (A) The risk score diagram in the validation set and heatmap of the screened lncRNAs expression. (B) The model in the test dataset for 3 and 5 years of ROC curves. (C) Survival curve of the validation set. Color images are available online.

threshold was |COR| >0.3 and $p < 0.05$. The lncRNA-mRNA co-expression network was visualized by using Metascape. The clusterProfiler R package was used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of the lncRNA-related mRNAs. All analyses were performed with default parameters. We also analyzed the correlations of the three most significant mRNAs of each lncRNA.

Results

Identification of differentially expressed lncRNAs

We used the limma package to analyze differences within the GSE20916 dataset. We used a |logFC| of >1 and an adjusted *p*-value of <0.05 as the differential analysis condition. The volcano diagram and heatmap are given in Figure 1A, B. We screened out 72 DELs, including 13 up-regulated and 59 downregulated lncRNAs (Accessory 1). We analyzed the differential lncRNA expression in the CRC patients from the TCGA dataset and set the same threshold as that used with the GSE20916 dataset. The volcano diagram and heatmap are given in Figure 2A, B. We obtained 118 DELs (Accessory 2). A total of 11 lncRNAs were crossed (Fig. 3).

Derivation of the lncRNA prognostic model

The clinicopathological features of the CRC patients from the TCGA database (age, sex, tumor–node–metastasis [TNM] stage, race, and OS) are given in Table 1.

First, we performed univariate Cox analysis to study the correlations between the crossed DELs and OS of the CRC patients in the training set. With $p < 0.05$ as an identification standard, two lncRNAs were screened out (Table 2).

We then conducted multivariate Cox analysis of the two lncRNAs in the training dataset. The coefficients for each lncRNA were the coefficients in the multivariate Cox analysis. The following model was derived:

$$\text{Risk score} = (1.4123 \times \text{expression value of FOXP4} - \text{AS1}) + (1.3378 \times \text{expression value of BBOX1} - \text{AS1}).$$

We used the survminer package to determine the best cutoff value based on the risk score and divided the patients into high- and low-risk groups. Then, we performed survival analysis based on the risk score and found that patients at high and low risk could be significantly separated. At the
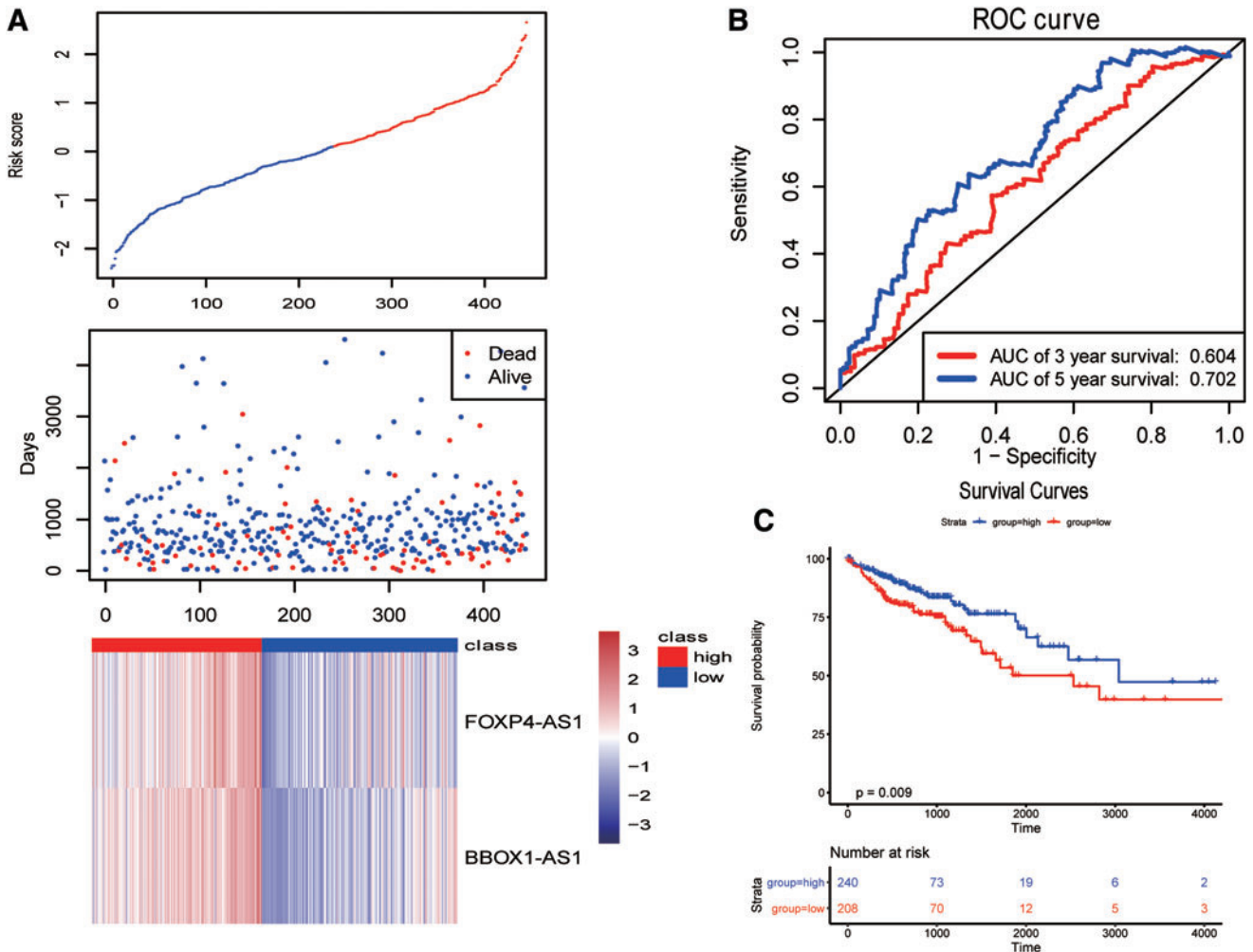

FIG. 6. **(A)** The risk score diagram in the total dataset and heatmap of the screened lncRNAs expression. **(B)** The 3- and 5-year ROC curve of the model in the total dataset. **(C)** Survival curve of the total dataset. Color images are available online.

same time, we calculated the ROC curve and found that the AUC was 0.724, which proved that the predictive ability of the model was strong. The survival and ROC curves are given in Figure 4.

We then validated the two lncRNAs in the validation set (Fig. 5).

Finally, we validated the model using all the TCGA samples (Fig. 6). As given in Figures 4–6, the 5-year AUCs were all >0.7, suggesting that the model had good predictive ability.

Risk model and clinical characteristics analysis

To explore the relationships between the prediction accuracy of the risk model and clinical characteristics, we analyzed the predictive relationships between the clinical features and risk score by using univariate and multivariate Cox analyses. The *p*-value of the predictive significance of the risk score was minimal (Tables 3 and 4). The survival curves for the meaningful variables in the univariate Cox analysis are given in Figure 7.

Line diagram and decline curve analysis of the model

We use TNM stages together with the risk score to construct the column line graph model (Fig. 8A). We calculated the prediction accuracy of the nomogram by using ROC analysis, which showed that the 3- and 5-year AUCs were 0.601 and 0.638, respectively (Fig. 8B). Figure 8C shows the 3-year nomogram model compared with the ideal model. Some indices of the 3-year model were basically consistent with those of the ideal model, suggesting that we con-

TABLE 4. MULTIVARIATE COX ANALYSIS

Characteristics	Hazard ratio	95% CI	p
Age			
<Median	Ref.		
≥Median	2.858	1.037–3.516	0.077
M			
M0	Ref.		
M1	5.058	2.302–11.113	<0.001
MX	2.728	1.046–7.115	0.04
N			
N0	Ref.		
N1	1.413	0.595–3.359	0.433
N2	1.827	0.815–4.096	0.144
T			
T1/T2	Ref.		
T3	1.226	0.399–3.77	0.018
T4	2.219	0.611–8.059	0.024
Gender			
Female	Ref.		
Male	0.66	0.355–1.227	0.189
Race			
Black	Ref.		
Others	1.319	0.44–3.95	0.621
White	1.126	0.405–3.131	0.82
Risk score			
Low	Ref.		
High	1.495	1.106–2.023	0.009

structed a model with high accuracy. Decline curve analysis showed that our model had a good net benefit in 3 and 5 years, indicating the substantial practical clinical value of the model.

LncRNA coexpression network and functional assessment of lncRNA-related mRNAs

A total of 1977 target genes (the absolute value of the screening threshold was >0.3, *p* < 0.05) were included in the analysis (Fig. 9, Accessory 3). Then, we conducted enrichment analysis of the selected target genes, for which the enrichment pathway selection threshold was *p* < 0.05. The significantly enriched pathways are given in Figure 10.

For biological processes, associated differentially expressed genes were mainly enriched in positive regulation of secretion by cells, positive regulation of secretion, and positive regulation of exocytosis. The molecular functions of these genes were enriched in DNA-directed 5′-3′ RNA polymerase activity, low-density lipoprotein particle receptor binding, 5′-3′ RNA polymerase activity, RNA polymerase activity, and lipoprotein particle receptor binding. The cellular components for these genes were mainly involved in the Flemming body. The KEGG enrichment analysis showed that the differentially expressed genes were mainly enriched in endocytosis and the calcium signaling pathway.

We also analyzed the correlations of the three most positively significant mRNAs of each lncRNA (Fig. 11).

Discussion

LncRNAs are noncoding RNAs with transcript lengths >200 nucleotides, and they have no significant protein-coding potential.[17] The proportion of coding RNA in the
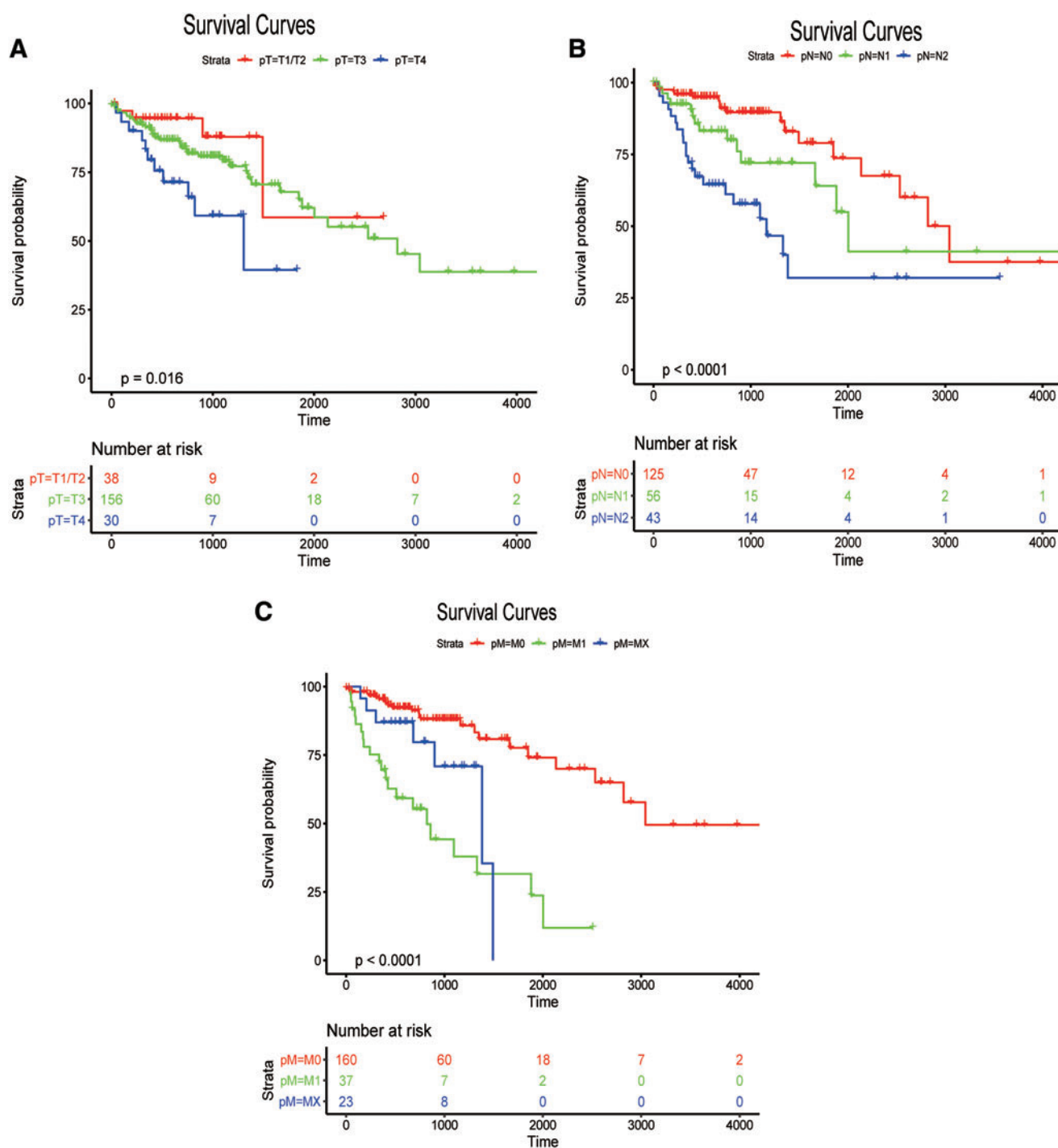
TABLE 3. UNIVARIATE COX ANALYSIS

| Characteristics	Hazard ratio	95% CI	p
Age			
<Median	Ref.	Ref.	Ref.
>Median	2.154	1.798–3.875	0.158
M			
M0	Ref.	Ref.	Ref.
M1	5.94	3.229–10.924	<0.001
MX	2.907	1.221–6.925	0.016
N			
N0	Ref.		
N1	1.891	0.94–3.805	0.074
N2	3.96	2.107–7.442	<0.001
T			
T1/T2	Ref.		
T3	1.756	0.622–4.954	0.288
T4	3.999	1.271–12.587	0.018
Gender			
Female	Ref.		
Male	1.093	0.634–1.886	0.749
Race			
Black	Ref.		
Others	1.131	0.411–3.115	0.812
White	1.414	0.546–3.661	0.476
Risk score			
Low	Ref.		
High	2.451	1.114–4.89	<0.001

FIG. 7. **(A)** The survival curves of the pT stage. **(B)** The survival curves of the pN stage. **(C)** The survival curves of the pM stage. Color images are available online.

human transcriptome is estimated to be <2%, whereas the proportion of noncoding RNA is ~98%, of which lncRNA, which regulates at least 70% of all gene expression, accounts for >80%.[17,18] LncRNA is widely involved in many important biological functions, such as cell proliferation, cell survival, and apoptosis, mainly by regulating gene expression through epigenetic, transcriptional, and posttranscriptional control.[19] The functions of lncRNAs are mainly to: (1) act as molecular signals to respond to intracellular or extracellular signals and as regulators of specific signaling

FIG. 8. **(A)** Line diagram. **(B)** ROC curve of line diagram. **(C)** The calibration plots for predicting 3- and 5-year OS nomogram-predicted probability of survival is plotted on the x-axis; actual survival is plotted on the y-axis. **(D–E)** DCA curves of 3 and 5 years. OS, overall survival. Color images are available online.
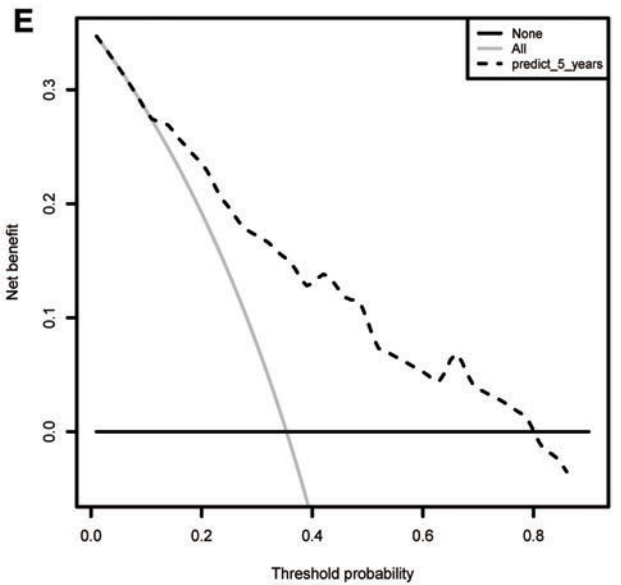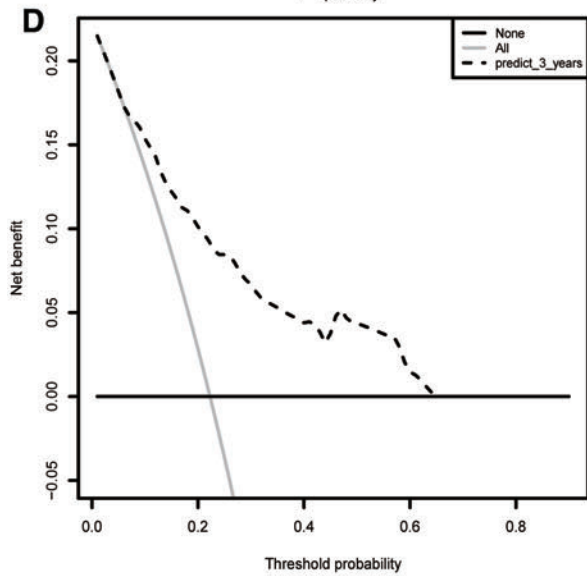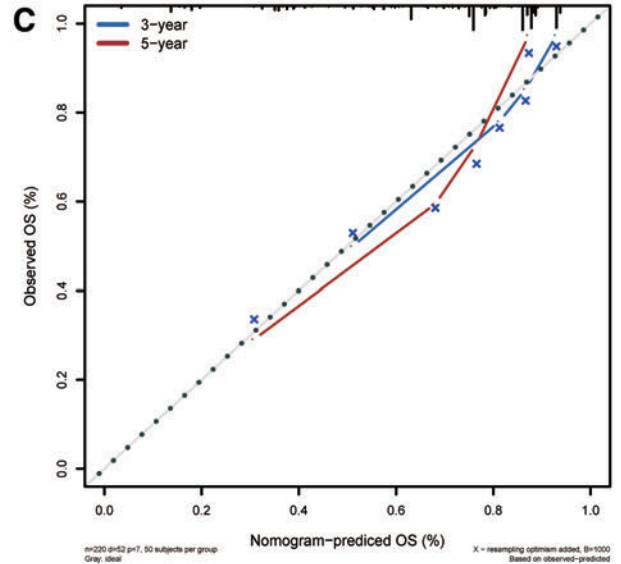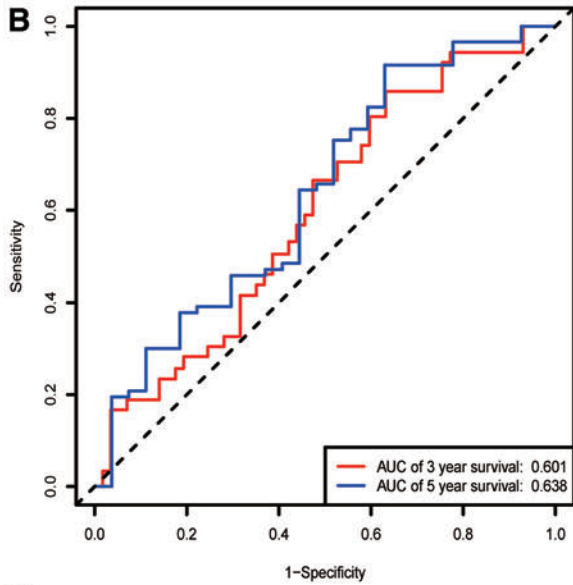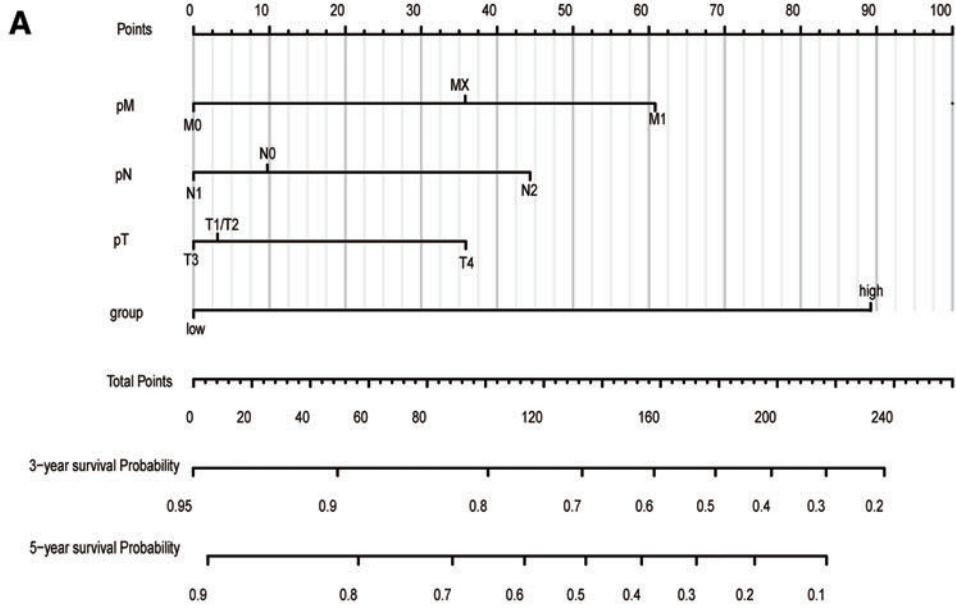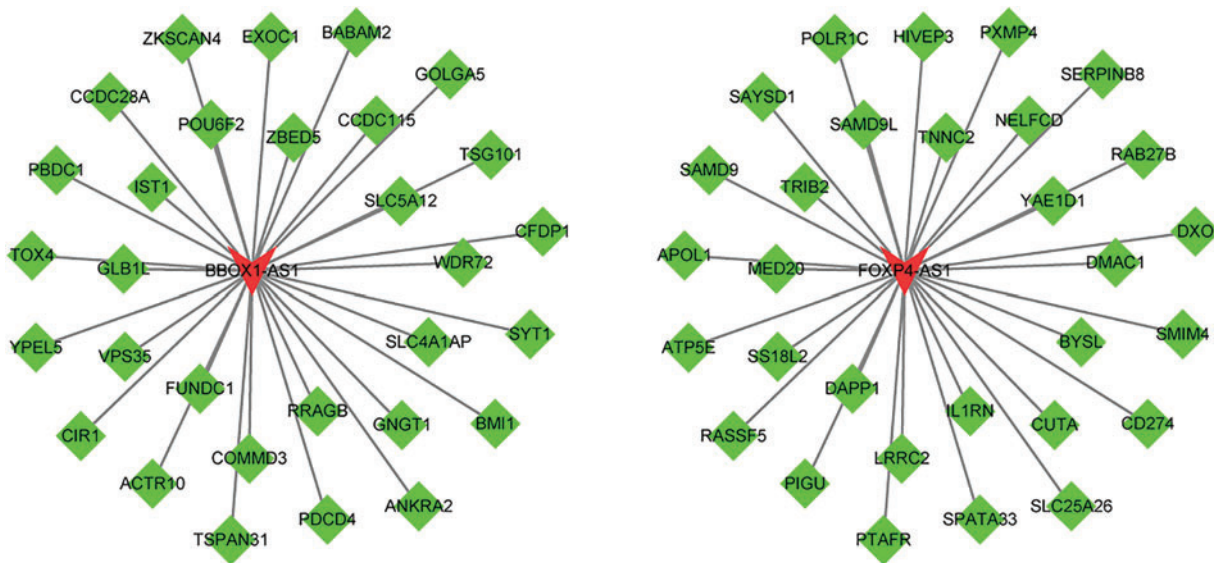
FIG. 9. LncRNA coexpressed regulatory network with mRNA Green represents target genes, and red represents lncRNA. mRNA, messenger RNA. Color images are available online.

pathways; (2) act as molecular decoys by binding some RNAs or proteins so that they leave their specific regions and lose their normal functions; (3) act as molecular guides to guide some proteins to specific mRNA or chromosomal sites, thereby affecting gene transcription, mRNA stability, or translation; and (4) act as molecular scaffolds to bind multiple molecules, allowing them to perform their specific functions. Subcellular localization of lncRNA is also closely related to its function. LncRNA in the nucleus may be involved in chromatin regulation, gene transcription, and alternative splicing of transcripts, among other activities, whereas lncRNA in the cytoplasm may be closely related to competitive endogenous RNA (ceRNA) and mRNA stability or mRNA translation.[20–22] Research has shown that lncRNAs are involved in cancer phenotypes of proliferation, growth suppression, motility, immortality, angiogenesis, and viability.[23] LncRNAs can be used to identify cancers, provide prognostic value for cancers, and inform therapeutic options for cancer patients.[23–33] Some abnormally expressed lncRNAs in CRC are closely related to cancer progression. LncRNA FTX promotes CRC progression by regulating the miR-192-5p/EIF5A2 axis.[34] LncRNA NBR2 inhibits CRC invasion and migration by downregulating miRNA-21,[35] whereas lncRNA ST8SIA6-AS1 promotes CRC cell proliferation, migration, and invasion.[36] Some lncRNAs, such as lncRNA HOTAIR, lncRNA PVT1, circulating lncRNA DANCER, and others, can act as prognostic factors in CRC.[25,37,38] Although there have been many studies on the associations between lncRNAs and CRC, the roles and mechanisms of lncRNAs in CRC and their clinical applications still need extensive exploration.

Recently, studies have reported different genetic models for predicting human cancer.[39–43] A weighted prognosis signature of six lncRNAs, including LINC01583, LINC00276, LUNAR1, DKFZp434J0226, SFTA1P, and OGFOD3, was constructed for predicting CRC prognosis. However, the samples were only acquired from GSE datasets, and the number of samples was <100, which may not provide enough accuracy. In our research, we screened out 11 crossed DELs from the

GSE20916 and TCGA datasets. To explore the relationships between the expression of individual DELs and clinical prognosis, we first included one factor in the regression model to fit the univariate Cox analysis, and two key lncRNAs related to CRC prognosis were further identified to incorporate into the multivariate Cox analysis. Through the multivariate Cox analysis of the two lncRNAs in the training dataset, we constructed a risk score model that could help to separate patients into high- and low-risk groups. We found that the AUC of the ROC curve was 0.724 in the training set, which was higher than that in a previous study (AUC = 0.683).[43] In the test validation set and total TCGA dataset, the AUCs of the ROC curves were 0.723 and 0.702, respectively. They were all >0.70, demonstrating that our model had better predictive power than that of the previous study.

Moreover, we analyzed the predictive relationships between the clinical features and risk score and found that the *p*-value of the predictive significance of the risk score was minimal, further demonstrating the high predictive power of the prognostic model. The TNM staging system is used to predict CRC prognosis. Nomograms have been found to be more accurate than the TNM staging system.[44] In 2000, Massacesi et al.[45] reported the first nomogram for advanced CRC. Since that study, various models have been constructed to predict prognoses in cancer patients.[46–48] Nevertheless, there have been few prediction models combining lncRNA information with CRC clinical features. In our study, we identified a prognostic model with two CRC lncRNAs, and we constructed a nomogram and risk classification system. The 3-year nomogram model was comparable with the ideal model, suggesting that we constructed a model with high accuracy. Survival analysis indicated that the risk score model could significantly discriminate prognostic differences between high- and low-risk score groups. Our nomogram could be a clinically valuable prognostic model for CRC patients.

Among the two lncRNAs in the model, lncRNA BBOX1-AS1 promotes the progression of gastric, colorectal, and cervical cancer through the ceRNA pathway.[49–51]
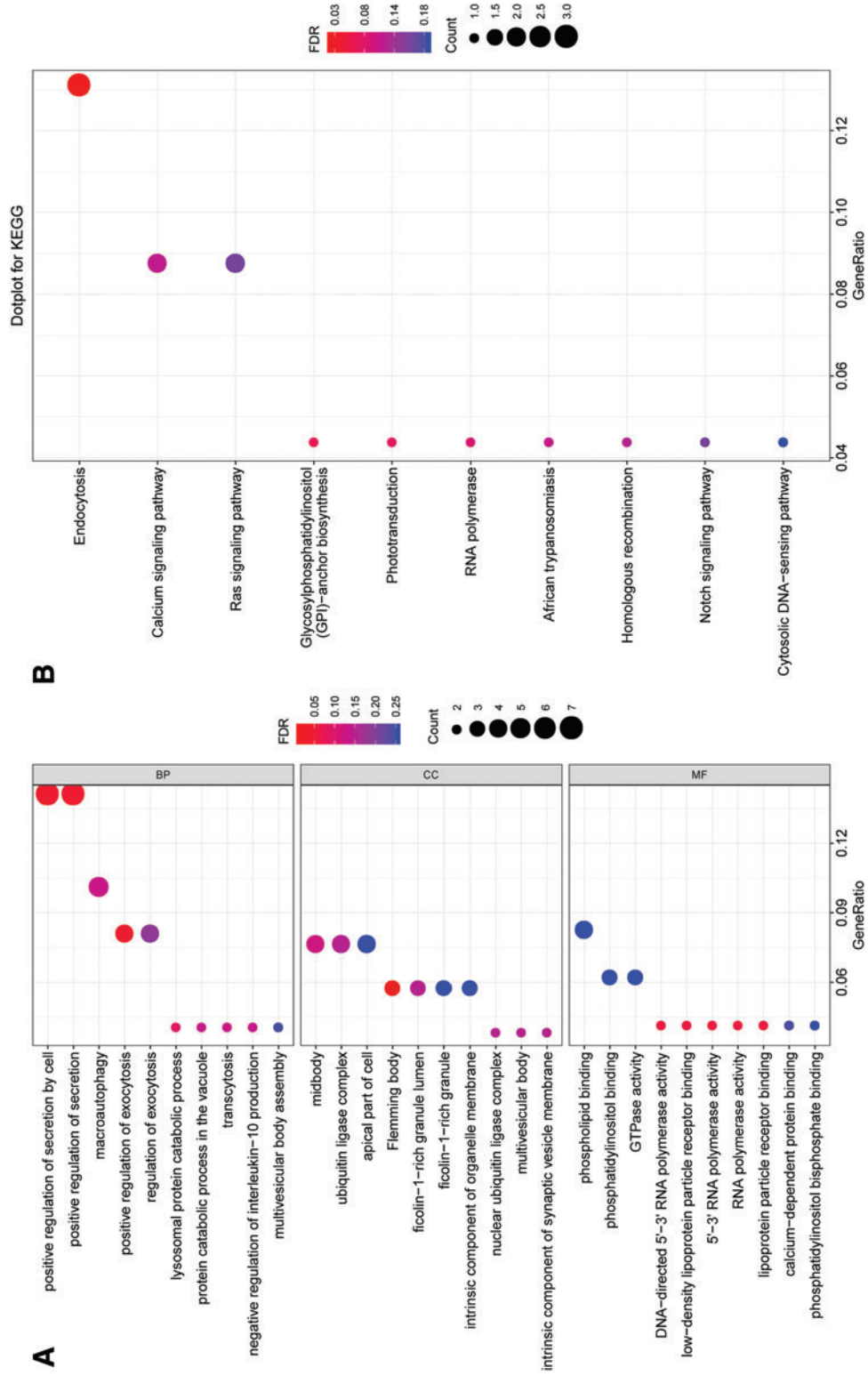
FIG. 10. Functional analysis of coexpressed mRNA (**A**). GO functional enrichment analysis (**B**). KEGG pathway enrichment analysis. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes. Color images are available online.
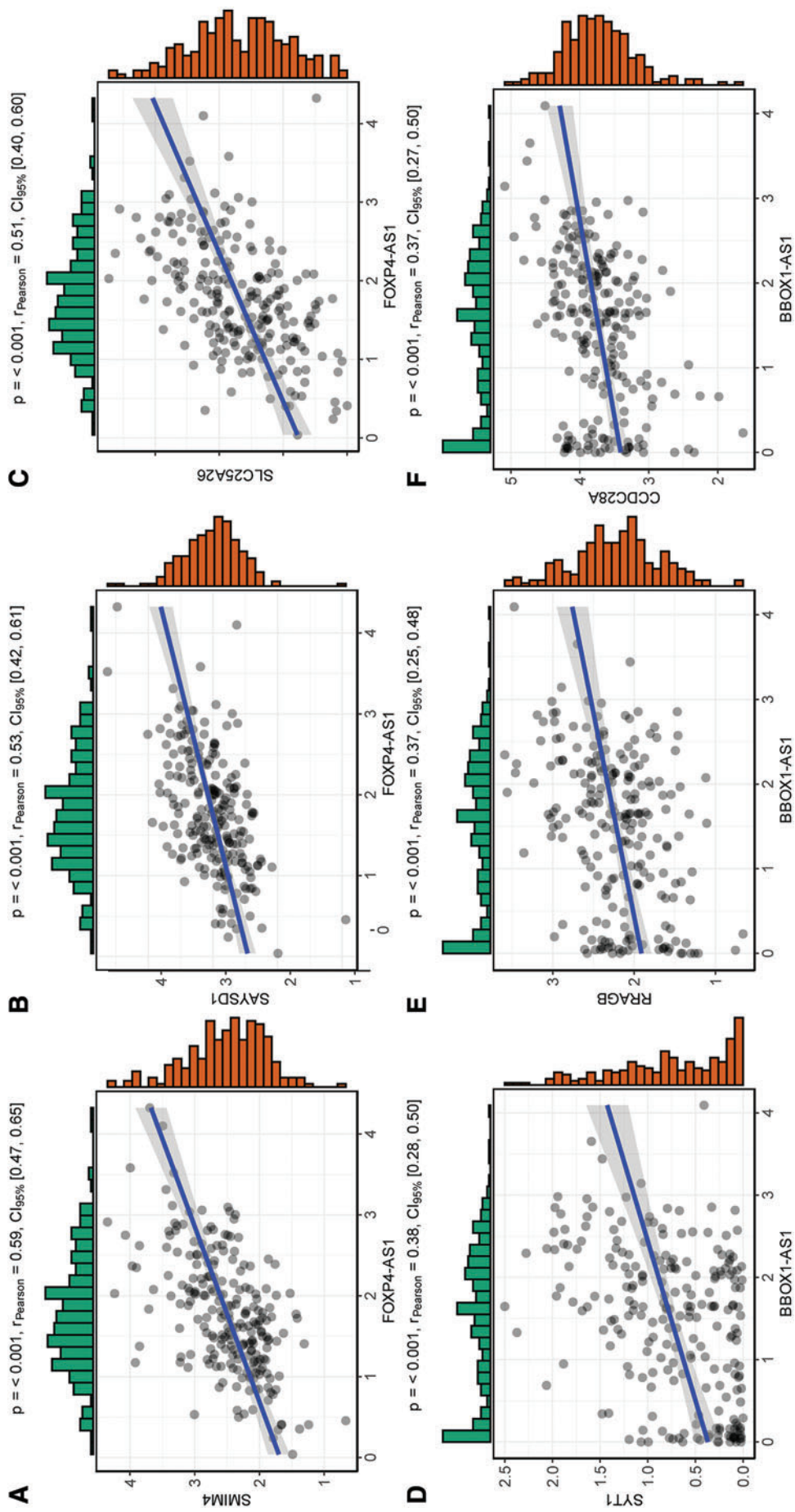
FIG. 11. **(A)** Correlation scatterplot of lncRNA FOXP4-AS1 and SMIM4. **(B)** Correlation scatterplot of lncRNA FOXP4-AS1 and SAYSD1. **(C)** Correlation scatterplot of lncRNA FOXP4-AS1 and SLC25A26. **(D)** Correlation scatterplot of lncRNA BBOX1-AS1 and SYT1. **(E)** Correlation scatterplot of lncRNA BBOX1-AS1 and RRAGB. **(F)** Correlation scatterplot of lncRNA BBOX1-AS1 and CCDC28A. Color images are available online.

Synaptotagmin-1 (SYT1), which is significantly associated with BBOX1-AS1, can promote colon cancer cell proliferation, migration, and invasion.[52] LncRNA FOXP4-AS1 plays an important role in the progression of some cancers, such as osteosarcoma, prostate cancer, and gastric cancer.[53–57] Similarly, lncRNA FOXP4-AS1 can be predictive of poor prognoses in CRC patients.[58] One of the three most positively significant mRNAs associated with FOXP4-AS1, that is, SLC25A26, which is downregulated owing to gene promoter hypermethylation, can promote cancer cell survival and proliferation.[59] However, the potential mechanisms of these two lncRNAs in CRC require further exploration. An advantage of our study is the use of many methods and tools along with systematic bioinformatics methods to process a large amount of data. Nevertheless, there were some shortcomings. First, there were differences in the sample sizes of the GEO and TCGA datasets. In addition, the prognostic model only incorporated lncRNA expression and did not consider the prognostic impact of other gene changes, such as micro RNA (miRNA) and mRNA expression. In the future, we expect to collect new samples and follow-up data. Moreover, we expect to further explore the mechanism at the molecular biological level while considering more factors that may affect prognosis to construct a more robust and reliable prognostic prediction model, which should be compared with traditional classical prognostic models for clinical application.

In conclusion, we constructed a two-lncRNA model, including lncRNA FOXP4-AS1 and lncRNA BBOX1-AS1, for predicting CRC prognosis. The two-lncRNA model could accurately predict CRC prognosis. In addition, these two lncRNAs could be involved in several pathways associated with CRC progression.

Data Availability Statement

The datasets analyzed during this study are available from the corresponding author upon reasonable request.

Disclosure Statement

No competing financial interests exist.

Funding Information

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020;70:7.
2. Siegel RL, Miller KD, Goding Sauer A, et al. Colorectal cancer statistics, 2020. CA Cancer J Clin 2020;70:145.
3. Zhu Z, Dong W. Overexpression of HHLA2, a member of the B7 family, is associated with worse survival in human colorectal carcinoma. Onco Targets Ther 2018;11:1563.
4. Lee SJ, Lee J, Park SH, et al. c-MET overexpression in colorectal cancer: A poor prognostic factor for survival. Clin Colorectal Cancer 2018;17:165.
5. Kang S, Na Y, Joung SY, et al. The significance of microsatellite instability in colorectal cancer after controlling for clinicopathological factors. Medicine (Baltimore) 2018;97:e0019.
6. Copija A, Waniczek D, Witkoś A, et al. Clinical significance and prognostic relevance of microsatellite instability in sporadic colorectal cancer patients. Int J Mol Sci 2017;18:107.
7. Morris KV, Mattick JS. The rise of regulatory RNA. Nat Rev Genet 2014;15:423.
8. Huarte M. The emerging role of lncRNAs in cancer. Nat Med 2015;21:1253.
9. Fatica A, Bozzoni I. Long non-coding RNAs: New players in cell differentiation and development. Nat Rev Genet 2014;15:7.
10. Lin C, Yang L. Long noncoding RNA in cancer: Wiring signaling circuitry. Trends Cell Biol 2018;28:287.
11. Fatima R, Akhade VS, Pal D, Rao SM. Long noncoding RNAs in development and cancer: Potential biomarkers and therapeutic targets. Mol Cell Ther 2015;3:5.
12. Zhang G, Chi N, Lu Q, et al. LncRNA PTCSC3 is a biomarker for the treatment and prognosis of gastric cancer. Cancer Biother Radiopharm 2020;35:77.
13. Zhang XT, Pan SX, Wang AH, et al. Long non-coding RNA (lncRNA) X-inactive specific transcript (XIST) plays a critical role in predicting clinical prognosis and progression of colorectal cancer. Med Sci Monit 2019;25:6429.
14. Zhou J, Lin J, Zhang H, et al. LncRNA HAND2-AS1 sponging miR-1275 suppresses colorectal cancer progression by upregulating KLF14. Biochem Biophys Res Commun 2018;503:1848.
15. Yang HY, Huang CP, Cao MM, et al. Long non-coding RNA CRNDE may be associated with poor prognosis by promoting proliferation and inhibiting apoptosis of cervical cancer cells through targeting PI3K/AKT. Neoplasma 2018;65:872.
16. Li X, Zhao X, Yang B, et al. Long non-coding RNA HOXD-AS1 promotes tumor progression and predicts poor prognosis in colorectal cancer. Int J Oncol 2018;53:21.
17. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet 2016;17:47.
18. Djebali S, Davis CA, Merkel A, et al. Landscape of transcription in human cells. Nature 2012;489:101.
19. Lee JT. Epigenetic regulation by long noncoding RNAs. Science 2012;338:1435.
20. Fu M, Zou C, Pan L, et al. Long noncoding RNAs in digestive system cancers: Functional roles, molecular mechanisms, and clinical implications (Review). Oncol Rep 2016;36:1207.
21. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell 2011;43:904.
22. Zhang Y, Liu X, Li Q, Zhang Y. lncRNA LINC00460 promoted colorectal cancer cells metastasis via miR-939-5p sponging. Cancer Manag Res 2019;11:1779.
23. Schmitt AM, Chang HY. Long noncoding RNAs in cancer pathways. Cancer Cell 2016;29:452.
24. Liu CN, Zhang HY. Serum lncRNA LOXL1-AS1 is a diagnostic and prognostic marker for epithelial ovarian cancer. J Gene Med 2020;22:e3233.
25. Pan X, Cheng R, Zhu X, et al. Prognostic significance and diagnostic value of overexpressed lncRNA PVT1 in colorectal cancer. Clin Lab 2019;65:12.
26. Yoruker EE, Keskin M, Kulle CB, et al. Diagnostic and prognostic value of circulating lncRNA H19 in gastric cancer. Biomed Rep 2018;9:181.
27. Lu H, Liu H, Yang X, et al. LncRNA BLACAT1 may serve as a prognostic predictor in cancer: Evidence from a meta-analysis. Biomed Res Int 2019;2019:1275491.

28. Luzón-Toro B, Fernández RM, Martos-Martínez JM, et al. LncRNA LUCAT1 as a novel prognostic biomarker for patients with papillary thyroid cancer. Sci Rep 2019;9: 14374.

29. Ma PJ, Guan QK, Xu DW, et al. LncRNA PANDAR as a prognostic marker in Chinese cancer. Clin Chim Acta 2017; 475:172.

30. Xu H, Zhou J, Tang J, et al. Identification of serum exosomal lncRNA MIAT as a novel diagnostic and prognostic biomarker for gastric cancer. J Clin Lab Anal 2020;34: e23323.

31. Jiang MC, Ni JJ, Cui WY, et al. Emerging roles of lncRNA in cancer and therapeutic opportunities. Am J Cancer Res 2019;9:1354.

32. Liu L, Zhan Y, Huang Y, Huang L. LncRNA FGD5-AS1 can be predicted as therapeutic target in oral cancer. J Oral Pathol Med 2020;49:243.

33. Tamang S, Acharya V, Roy D, et al. SNHG12: An lncRNA as a potential therapeutic target and biomarker for human cancer. Front Oncol 2019;9:901.

34. Zhao K, Ye Z, Li Y, et al. LncRNA FTX contributes to the progression of colorectal cancer through regulating miR-192-5p/EIF5A2 axis. Onco Targets Ther 2020;13: 2677.

35. Bai J, Xu J, Zhao J, Zhang R. LncRNA NBR2 suppresses migration and invasion of colorectal cancer cells by downregulating miRNA-21. Hum Cell 2020;33:98.

36. Huang CM, Cao GY, Yang CX, et al. LncRNA ST8SIA6-AS1 promotes colorectal cancer cell proliferation, migration and invasion by regulating the miR-5195/PCBP2 axis. Eur Rev Med Pharmacol Sci 2020;24:4203.

37. Shen X, Xue Y, Cong H, et al. Circulating lncRNA DANCR as a potential auxillary biomarker for the diagnosis and prognostic prediction of colorectal cancer. Biosci Rep 2020;40:BSR20191481.

38. Chen S, Zhang C, Feng M. Prognostic value of lncRNA HOTAIR in colorectal cancer: A meta-analysis. Open Med (Wars) 2020;15:76.

39. Kam AE, Pendurti G, Shah UH, et al. Survival outcome and prognostic model of patients with colorectal cancer on phase 1 trials. Invest New Drugs 2019;37:490.

40. Wang D, Yang Y, Jin L, et al. Prognostic models based on postoperative circulating tumor cells can predict poor tumor recurrence-free survival in patients with stage II-III colorectal cancer. J Cancer 2019;10:4552.

41. Wang L, Shi J, Huang Y, et al. A six-gene prognostic model predicts overall survival in bladder cancer patients. Cancer Cell Int 2019;19:229.

42. Zhao X, Liu J, Liu S, et al. Construction and validation of an immune-related prognostic model based on TP53 status in colorectal cancer. Cancers (Basel) 2019;11:1722.

43. Zhao J, Xu J, Shang AQ, Zhang R. A six-lncRNA expression signature associated with prognosis of colorectal cancer patients. Cell Physiol Biochem 2018;50:1882.

44. Touijer K, Scardino PT. Nomograms for staging, prognosis, and predicting treatment outcomes. Cancer 2009; 115(13 Suppl):3107.

45. Massacesi C, Norman A, Price T, et al. A clinical nomogram for predicting long-term survival in advanced colorectal cancer. Eur J Cancer 2000;36:2044.

46. Li H, Gao C, Liu L, et al. 7-lncRNA assessment model for monitoring and prognosis of breast cancer patients: Based on Cox regression and co-expression analysis. Front Oncol 2019;9:1348.

47. Liu Z, Li M, Hua Q, et al. Identification of an eight-lncRNA prognostic model for breast cancer using WGCNA network analysis and a Coxproportional hazards model based on L1-penalized estimation. Int J Mol Med 2019;44: 1333.

48. Xing L, Zhang X, Chen A. Prognostic 4-lncRNA-based risk model predicts survival time of patients with head and neck squamous cell carcinoma. Oncol Lett 2019;18:3304.

49. Liu J, Zhu J, Xiao Z, et al. BBOX1-AS1 contributes to colorectal cancer progression by sponging hsa-miR-361-3p and targeting SH2B1. FEBS Open Bio 2020 [Epub ahead of print]; DOI: 10.1002/2211-5463.12802.

50. Xu J, Yang B, Wang L, et al. LncRNA BBOX1-AS1 upregulates HOXC6 expression through miR-361-3p and HuR to drive cervical cancer progression. Cell Prolif 2020;53: e12823.

51. Yang Y, Yu Q, Li B, et al. BBOX1-AS1 accelerates gastric cancer proliferation by sponging miR-3940-3p to upregulate BIRC5 expression. Dig Dis Sci 2020 [Epub ahead of print]; DOI: 10.1007/s10620-020-06308-0.

52. Lu H, Hao L, Yang H, et al. miRNA-34a suppresses colon carcinoma proliferation and induces cell apoptosis by targeting SYT1. Int J Clin Exp Pathol 2019;12:2887.

53. Chen RY, Ju Q, Feng LM, et al. The carcinogenic complex lncRNA FOXP4-AS1/EZH2/LSD1 accelerates proliferation, migration and invasion of gastric cancer. Eur Rev Med Pharmacol Sci 2019;23:8371.

54. Li Y, Li T, Yang Y, et al. YY1-induced upregulation of FOXP4-AS1 and FOXP4 promote the proliferation of esophageal squamous cell carcinoma cells. Cell Biol Int 2020;44:1447.

55. Wu X, Xiao Y, Zhou Y, et al. LncRNA FOXP4-AS1 is activated by PAX5 and promotes the growth of prostate cancer by sequestering miR-3184-5p to upregulate FOXP4. Cell Death Dis 2019;10:472.

56. Yang L, Ge D, Chen X, et al. FOXP4-AS1 participates in the development and progression of osteosarcoma by downregulating LATS1 via binding to LSD1 and EZH2. Biochem Biophys Res Commun 2018;502:493.

57. Zhao J, Yang T, Li L. LncRNA FOXP4-AS1 is involved in cervical cancer progression via regulating miR-136-5p/CBX4 axis. Onco Targets Ther 2020;13:2347.

58. Li J, Lian Y, Yan C, et al. Long non-coding RNA FOXP4-AS1 is an unfavourable prognostic factor and regulates proliferation and apoptosis in colorectal cancer. Cell Prolif 2017;50:e12312.

59. Menga A, Palmieri EM, Cianciulli A, et al. SLC25A26 overexpression impairs cell function via mtDNA hypermethylation and rewiring of methyl metabolism. FEBS J 2017;284:967.