GENETIC TESTING AND MOLECULAR BIOMARKERS
Volume 25, Number 8, 2021
Mary Ann Liebert, Inc.
Pp. 517–527
DOI: 10.1089/gtmb.2021.0066

ORIGINAL ARTICLES

A Five-mRNA Expression Signature to Predict Survival in Oral Squamous Cell Carcinoma by Integrated Bioinformatic Analyses

Hejia Guo,[1–4] Cuiping Li,[2–4] Xiaoping Su,[2–4] and Xuanping Huang[1–4,i]

Objectives: This study was designed to identify a messenger RNA (mRNA) expression signature to predict survival in patients with oral squamous cell carcinoma (OSCC).
Methods: mRNA expression profiles were integrated with clinical data from 280 samples, including 19 normal tissues and 261 OSCC tissues in The Cancer Genome Atlas. We identified differentially expressed mRNAs (DEmRNAs) between the OSCC and normal tissue samples and developed a novel mRNA-focused expression signature using a Cox regression analysis and other bioinformatic methods. The prognostic value of this signature was evaluated by Kaplan–Meier analysis, multivariable COX regression, and receiver operating characteristic (ROC) curve analysis. Protein–protein interaction (PPI) network, gene ontology, and Kyoto Encyclopedia of Genes and Genomes enrichment analysis were performed to predict the function of the DEmRNAs. Signature-related mRNAs were analyzed by gene set enrichment analyses (GSEA) and validated by quantitative real-time polymerase chain reaction (qRT-PCR) in 20 paired OSCC and adjacent healthy tissues.
Results: We identified a novel 5-mRNA expression signature (*HOXA1*, *CELSR3*, *HIST1H3J*, *ZFP42*, and *ASCL4*) that could predict patient outcomes in OSCC. The risk score based on the signature was able to separate OSCC patients into high- and low-risk groups that showed significantly different overall survival ($p < 0.001$, log-rank test). The signature was further validated as an effective independent prognostic predictor of OSCC by multivariate Cox regression analysis (hazard ratio = 3.747, confidence interval: 2.279–5.677, $p < 0.001$) and ROC curve of the third year (area under the curve = 0.733). Functional analysis demonstrated that the key hub genes in the PPI network were mainly enriched in cell division, cell proliferation, and the p53 signaling pathway. GSEA results showed that the 5 mRNAs were significantly enriched in mismatch repair, DNA replication, and the NOTCH signaling pathway. Finally, qRT-PCR results showed that the 5 mRNAs were upregulated in OSCC tissue in agreement with the predictions from our bioinformatics analysis.
Conclusions: We identified a novel 5-mRNA signature that could predict the survival of patients with OSCC and may be a promising biomarker for personalized cancer treatments.

Keywords: oral squamous cell carcinoma, mRNA signature, bioinformatics analysis

Introduction

Oral cancer is the sixth most common malignancy and nearly 90% of oral cancers are oral squamous cell carcinoma (OSCC), which has a 5-year overall survival rate of <50% (Jemal *et al*., 2011). The number of new OSCC cases is around 450,000 annually around the world (Villagómez-Ortíz *et al*., 2016). The current standard-of-care treatment for OSCC is surgery with adjuvant radiotherapy and chemotherapy. While significant advances have been made in the diagnosis and treatment of OSCC, 5-year survival rates remain poor (Ferlay *et al.*, 2015).

[1]Department of Oral and Maxillofacial Surgery, College & Hospital of Stomatology, Guangxi Medical University, Nanning, P.R. China.
[2]Guangxi Key Laboratory of Oral and Maxillofacial Rehabilitation and Reconstruction, Nanning, P.R. China.
[3]Guangxi Clinical Research Center for Craniofacial Deformity, Nanning, P.R. China.
[4]Guangxi Key Laboratory of Oral and Maxillofacial Surgery Disease Treatment, Nanning, P.R. China.
[i]ORCID ID (https://orcid.org/0000-0002-1023-1019).

Currently, biopsy is the standard approach for the diagnosis of OSCC. Other clinical prognostic indicators that are commonly used in OSCC include tumor-node-metastasis stage, tumor margins, and tumor size. However, due to the heterogeneity of OSCC, patients with the same clinical features have very different outcomes. There is an urgent need to identify effective biomarkers for improved diagnosis and prognosis in OSCC, which can be used to develop personalized treatment programs and provide a better understanding of the underlying molecular mechanisms of OSCC.

The development of whole-exome and whole-genome sequencing technologies has led to a body of studies demonstrating that molecular markers have a major potential for the early diagnosis and prognosis of tumors. Brooks *et al.* (2016) demonstrated a significant association of high *COL1A1* and *COL1A2* messenger RNA (mRNA) expression in non-muscle invasive bladder cancer with poor progression-free and overall survival in a multicenter clinical cohort study (Michael *et al.*, 2016). Also, the expression of *CEA* mRNA in peripheral blood has been demonstrated as a prognostic marker for advanced non-small cell lung cancer as it closely correlates with *CK-18* and *CK-19* expression levels (Arrieta *et al.*, 2014).

Recent studies have identified mRNA-focused expression signatures to predict patient survival in several cancers, including esophageal adenocarcinoma (Dong *et al.*, 2018; Chen *et al.*, 2019) and early relapse hepatocellular carcinoma (Cai *et al.*, 2019). In oral cancers, p16 and HPV are recognized as predictive biomarkers (da Costa *et al.*, 2018). In addition, Li *et al.* collected unstimulated salivary RNA from primary OSCC patients and normal paired individuals and identified 7 mRNA biomarkers, including *IL8*, *IL1B*, and *DUSP1* based on microarray analysis and quantitative polymerase chain reaction (qPCR) validation (Li *et al.*, 2004). The combination of spermidine/spermine N1-acetyltransferase 1 and interleukin 8 has shown a 75.5% predictive ability as an early detection tool for OSCC (Michailidou *et al.*, 2016). Another study showed that *IGTA5* promotes the progress of OSCC by activating the PI3K/Akt signaling pathway, suggesting that *IGTA5* may be a novel biomarker in the treatment of OSCC (Fan *et al.,* 2019). However, these emerging biomarkers commonly have low specificity and sensitivity and have not been validated in large clinical cohorts. Therefore, it is necessary to develop more effective biomarkers for predicting OSCC survival and better understand the underlying molecular mechanisms of OSCC.

In this study, we analyzed data from mRNA expression profiles from OSCC patients deposited in The Cancer Genome Atlas (TCGA) to identify mRNA molecules capable of predicting overall survival by Cox regression analysis. We developed a 5-mRNA expression signature that was an effective, independent predictor of survival in OSCC patients using Kaplan–Meier and multivariate Cox regression analysis. To further verify the results from the bioinformatics analysis, the expression levels of the 5-mRNAs were determined in 20 paired OSCC and adjacent healthy tissues by quantitative real-time polymerase chain reaction (qRT-PCR).

Methods

Data processing

The mRNA expression data and corresponding clinical characteristics of the OSCC patients were downloaded from TCGA in May 2020. The data set consisted of 280 samples, including 19 normal and 261 OSCC samples from 257 patients with OSCC. Differentially expressed mRNAs (DEmRNAs) were identified between the OSCC and normal tissues. mRNA expression was analyzed using the exact test performed using the edgeR package (bioconductor.org/packages/release/bioc/html/edgeR.html) in R (software version 4.0.0; r-project.org) (Robinson *et al.*, 2010). (1) $p < 0.05$ and $\log2|FC| > 1$, and (2) false discovery rate <0.05 were used as the cutoff criteria for the DEmRNA analysis.

Patient recruitment

Twenty tissue samples from patients diagnosed with OSCC in the Affiliated Stomatology Hospital of Guangxi Medical University (ASHGMU, Nanning, China) between May 2020 and August 2020 were analyzed. This study was approved by the Ethical Review Committee of Guangxi Medical University. Approval number: 2018-082), and all subjects participating in this study signed written informed consent.

Statistical analysis

The DEmRNA expression profiles were combined with paired survival prognostic information and univariate Cox regression analysis was performed to identify the DEmRNAs that were significantly related to survival ($p < 0.001$). Multivariate Cox regression analysis was performed on all candidate DEmRNAs to identify the optimum mRNA signature that had independent prognostic value. In multivariate Cox analysis, the correlation coefficient, hazard ratio (HR), and 95% confidence interval (CI) of each candidate DEmRNAs were calculated and the model performance was assessed using the Akaike information criterion (AIC). In both univariate and multivariate Cox regression analyses, the mRNA expression level was considered an independent variable. A prognostic risk assessment model was carried out using the linear combination of mRNA expression values weighted by the coefficient. These data were used to calculate a risk score as follows:

$$\text{Risk score} = \sum_{i=1}^{n} (Ci * Ei)$$

The risk score is an mRNA-focused risk score for OSCC patients where n is the number of prognostic mRNAs and Ci is the regression coefficient that represents the contribution of mRNA in the prognostic risk scores. Ei represents the expression value of the mRNA. According to the median risk score, patients were divided into high-risk and low-risk groups.

Validation of the mRNA-focused signature

The Kaplan–Meier method was used to calculate the overall and median survival times for the high-risk and low-risk groups. Differences in the survival times between the two groups were analyzed using a log-rank test at the significance level of 1%. We integrated the risk score with the survival time and survival status of patients. Then the R package ''survivalROC'' (Patrick and Saha-Chaudhuri, 2013) was used to analyze the sensitivity and specificity of the 5-mRNA expression signature. The Kaplan–Meier method (multiplication of the positive limit) was selected to construct the survival function and the endpoint defined at 3 years. The predictive time-dependent receiver operating characteristic (ROC) curve of the third year was plotted and the area

under the curve (AUC) was then calculated. A chi-square test was used to evaluate correlations between the mRNA-focused signature and the clinicopathological variables ($p < 0.05$). Then the mRNA-focused signature and the clinical variables associated with overall survival were subjected to multivariate Cox regression analysis to verify that the mRNA-focused signature can be used as an independent index for prognosis.

Bioinformatics analysis of mRNA function

The online tool string was used to obtain the DEmRNA-encoded proteins and the protein–protein interaction (PPI) network. Based on the string data, genes were selected that had comprehensive scores ≥0.9. Cytoscape was then used to build a visual PPI network in which the function modules were identified using the cytoHubba plug-in. The parameters of the cluster search in cytoHubba were the top 20 nodes ranked by Maximal Clique Centrality (MCC). The corresponding proteins in the function modules may be the core proteins and key candidate genes that have important physiological regulatory functions. To infer the biological process and function of the DEmRNAs, the corresponding DEGs were analyzed using gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) through DAVID Bioinformatics Resources. *p*-Values <0.05 and gene counts >2 were used as the screening criteria.

Gene set enrichment analysis (GSEA) created an ordered list of all genes indicated by their connection with the expression of *HOXA1*, *CELSR3*, *HIST1H3J*, *ZFP42*, and *ASCL4*. Annotated gene sets (c2.cp.kegg.v7.1.symbols.gmt) were selected as the reference gene sets that included terms with $p < 0.05$. A permutation number of 1000 was adopted.

qRT-PCR verification

The mRNA expression levels of *HOXA1*, *CELSR3*, *HIST1H3J*, *ZFP42*, and *ASCL4* were quantified by qRT-PCR. The PCR conditions were as follows: stage 1, 95°C/30 s for 1 cycle; stage 2, 95°C/5 s, and 60°C/30 s for 40 cycles; and stage 3, 95°C/15 s, 60°C/60 s, and 95°C/15 s for 1 cycle. Each sample was tested in triplicate. The expression of the mRNAs was normalized against the level of *GAPDH* expression. The levels of mRNA expression were calculated using the $2^{-\Delta\Delta Ct}$ method. The statistical significance of the expression differences was calculated using a Student's *t*-test. The sequence information of all primers (5′->3′) is as shown below:

HOXA1-F: ACAGCCCCTACGCGTTAAAT.
HOXA1-R: ATGTATTGAGGCGAGCCCAC.
CELSR3-F: CCTGCCAGCCAGGTTACTAC;
CELSR3-R: ATCCTGTGCTCACAGTGGTG.
HIST1H3J-F: AACTGGCGGAAAGTCGTCAA.
HIST1H3J-R: CCCTTGAAGCGGATTGACCT.
ZFP42-F: ACTGGAGAGAAGCCGTTTCG.
ZFP42-R: TGCGTTAGGATGTGGGCTTT.
ASCL4-F: TCATGCACCGTTTCCCTGAA.
ASCL4-R: CATTTGCCGGAAAGCACACA.
GAPDH-F: CAGGAGGCATTGCTGATGAT.
GAPDH-R: GAAGGCTGGGGCTCATTT.

Results

Patient characteristic and DEmRNA analysis

In our study, a total of 280 samples were analyzed, which included 261 OSCC samples and 19 normal tissue samples. All 280 samples were obtained from 257 patients with OSCC and so some of the patients in the cohort provided more than one tissue sample. The clinical characteristics and survival data from the samples are summarized in Table 1 and Supplementary Table S1, respectively. The clinical variables included the age, gender, American Joint Committee on Cancer (AJCC) stage, clinical grade, AJCC-T, AJCC-N, and the survival data for each patient. A total of 280-mRNA expression files were extracted and analyzed using R language (EdegR). A total of 1769 DEmRNAs were identified from 280 mRNA expression files. These included 1048 upregulated and 721 downregulated mRNAs in the OSCC tissue compared to normal tissues (data are shown in Fig. 1). The Supplementary Fig. S1 showing the scheme of identifying DEmRNAs.

Table 1. Clinical Characteristics of Oral Squamous Cell Carcinoma Patients Used in this Study

Characteristic	*Number of patients*
Age (years) <60/≥60	121/136
Sex Male/female	180/77
Grade 1/2/3/4/NA	39/159/51/2/6
Stage I/II/III/IV/NA	17/35/50/130/25
T T1/T2/T3/T4/NA	26/74/57/79/21
N N0/N1/N2/N3/NA	95/38/88/2/34
Vital status Alive/dead	161/96

NA, not available; T, tumor; N, lymph node status.

Identification of a 5-mRNA expression signature associated with survival in OSCC patients

DEmRNAs are independent variables that were subjected to univariate COX regression analysis to identify mRNAs whose expression was strongly associated with overall survival ($p < 0.001$). Initially, a total of six mRNAs (*HOXA1*, *CELSR3*, *HIST1H3J*, *ZFP42*, *ASCL4*, and *IQCN*) were identified, and then multivariate COX regression analysis was used to choose the optimum independent mRNAs for survival prediction using overall survival as an independent variable and the six candidate mRNA expression level as independent covariates. Due to a significant correlation with the expression of *HOXA1* (correlation coefficient = –0.468, $p < 0.001$), *IQCN* was excluded from further analysis (shown in Supplementary Fig. S2). The signature was generated from 5 mRNAs that performed equally well. Also, as the signature consisted of only five genes, the medical costs of the patients may be significantly reduced.

A total of 5 mRNAs were retained as independent prognostic mRNAs in OSCC. The results showed that *HOXA1*, *CELSR3*, *HIST1H3J*, *ZFP42*, and *ASCL4* were independent prognostic indicators of OSCC and the AIC value was 1027.12. The detailed prognostic analysis is shown in Table 2 and Supplementary Figure S3. We found that *CELSR3* and

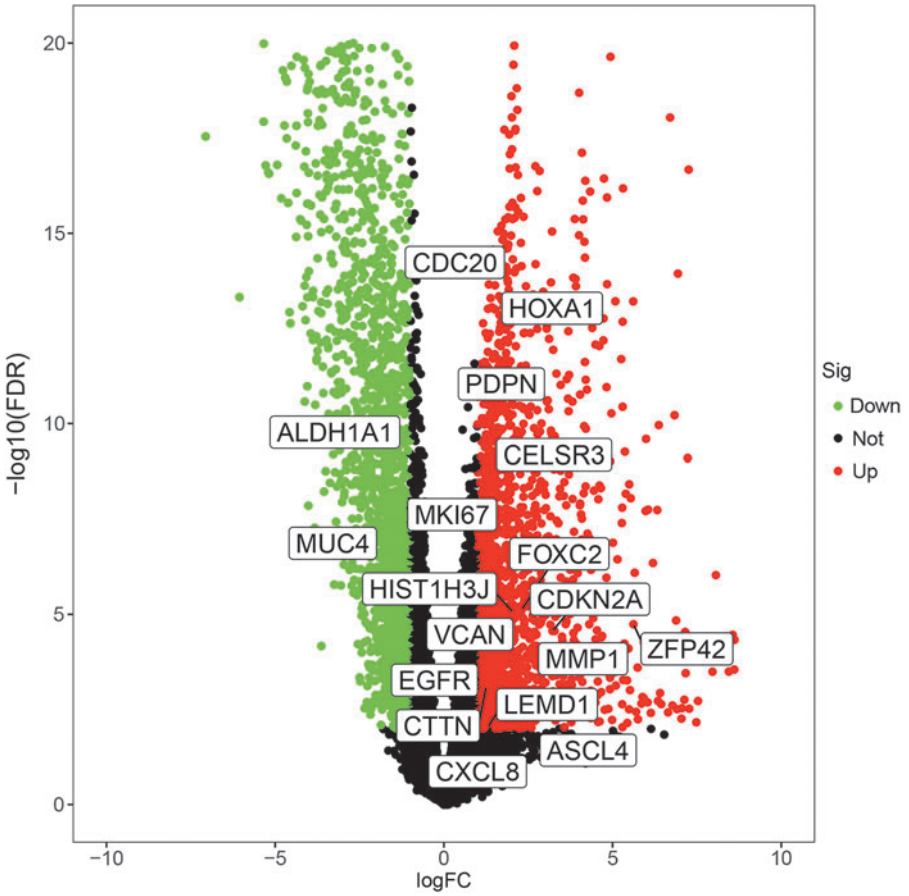

FIG. 1. Volcano plot of DEmRNAs. *Red dots* represent upregulated mRNAs and *green dots* represent downregulated mRNAs. FC, fold change; FDR, false discovery rate; DEmRNAs, differentially expressed mRNAs. Color images are available online.

ASCL4 had regression coefficients <0 and the HR was <1, and so these genes were regarded as protective factors. The upregulation of protective mRNAs correlated with a good prognosis. In contrast, *HOXA1*, *HIST1H3J*, and *ZFP42* were viewed as risk factors as the regression coefficients were >0 and the HR was >1. The upregulation of high-risk mRNAs correlated with poor overall survival.

The 5 mRNAs were used to develop a signature to predict patient survival. A risk score model was developed using a weighted scoring method based on the regression coefficients and mRNA expression levels as follows: risk score = (0.1979 × expression value of *HOXA1*)+(−0.1727 × expression value of *CELSR3*)+(0.2235 × expression value of *HIST1H3J*)+(0.0868 × expression value of *ZFP42*)+(−0.2690 × expression value of *ASCL4*).

Using the risk formula, the risk score for each patient was calculated and all patients were split into high- (129 cases) and low-risk groups (128 cases) using the median risk score (median score = 1.048, shown in Fig. 2a, b). The Kaplan–Meier method was used to measure the survival rate of the two groups and a risk heat map was used to visualize differences in the expression levels of 5 mRNAs in the two groups. *ASCL4* and *CELSR3* were expressed at higher levels in the low-risk group, while the *HOXA1*, *HIST1H3J*, and *ZFP42* were significantly higher in the high-risk groups (shown in Fig. 2c). A log-rank test was used to compare the survival curves (shown in Fig. 2d). The difference in the overall survival between the high-risk and low-risk groups was statistically significant ($p < 0.001$). As shown in Figure 2d, patients in the low-risk group had a longer median survival time compared to patients

Table 2. The Detailed Information of Five Prognostic Messenger RNAs Significantly Associated with Overall Survival in Oral Squamous Cell Carcinoma

Gene symbol	*Chromosome*	p[a]	*Coefficient*[b]	*HR*[b]	*Regulation*[c]
HOXA1	Chr7:27092993–27096000	0.0008	0.1979	1.2188	Upregulated
CELSR3	Chr3:48636463–48662886	0.0011	−0.1727	0.8414	Upregulated
HIST1H3J	Chr6:27890315–27890826	0.0008	0.2235	1.2504	Upregulated
ZFP42	Chr4:187994044–188005046	0.0007	0.0868	1.0906	Upregulated
ASCL4	Chr12:107774385–107776644	0.0009	−0.2690	0.7641	Upregulated

[a]Derived from the univariable Cox proportional hazards regression analysis in 257 OSCC patients.
[b]Derived from the multivariate Cox proportional hazards regression analysis in 257 OSCC patients.
[c]Type of regulation in OSCC compared to normal tissue.
HR, hazard ratio; OSCC, oral squamous cell carcinoma.

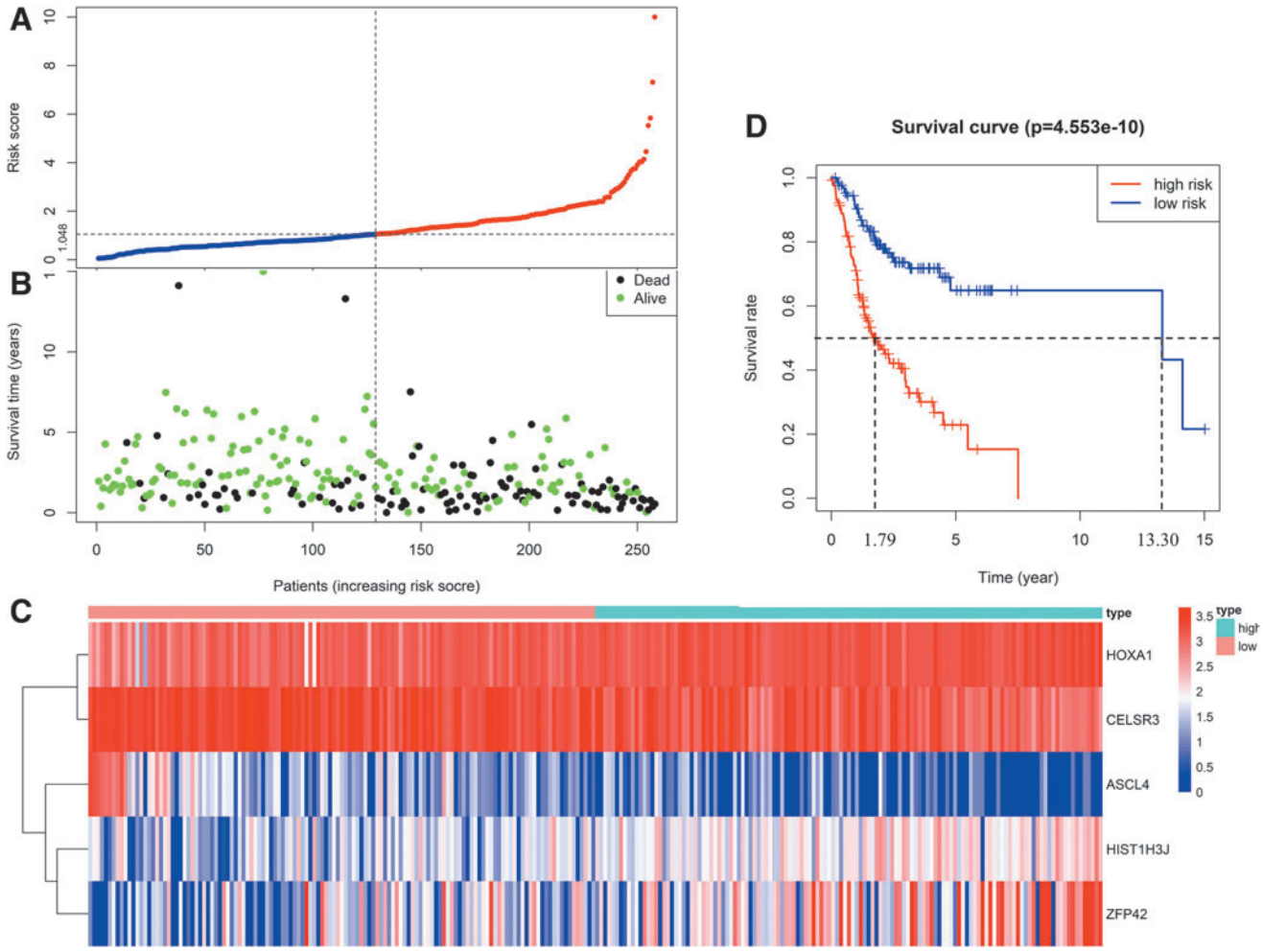

FIG. 2. 5-mRNA risk score analysis of 257 OSCC patients. **(A)** 5-mRNA risk score distribution. **(B)** Patient survival status along with risk score. The *dotted line* represents the 5-mRNA signature cutoff dividing patients into low-risk and high-risk groups. **(C)** Heat map of 5-mRNA expression profiles of OSCC patients. **(D)** Kaplan–Meier curves for low- and high-risk patients. OSCC, oral squamous cell carcinoma. Color images are available online.

in the high-risk group (median survival time 1.79 years vs. 13.30 years). The overall survival in the high-risk group was 47.5% at 2 years, 36.5% at 3 years, and 20.6% at 5 years, while the corresponding levels of overall survival in the low-risk group were 80.5%, 73.3%, and 63.9%, respectively. The data clearly showed that survival was worse in the high-risk group compared to the low-risk group. Moreover, the HR of the high-risk group versus the low-risk for overall survival was 3.497 (CI: 2.279–5.677, $p<0.001$) by univariate Cox regression analysis (shown in Table 4). It is interesting to find that two protective mRNAs are upregulated in OSCC in comparison to normal. The specificity and sensitivity of 5-mRNA diagnostic indicators were comprehensively evaluated by the ROC curve of the third year, and the AUC reached 0.733 (shown in Fig. 3).

Correlation between the 5-mRNA signature and the clinical characteristics of OSCC patients

We further analyzed the 5-mRNA signature to determine whether the signature based on risk score was associated with the clinical characteristics of OSCC patients. Statistical analysis showed that patients with high-risk scores were

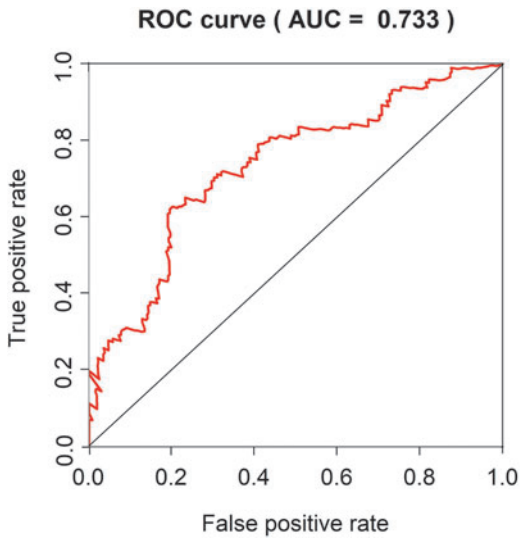

FIG. 3. The ROC curve for the 5-mRNA signature representing 3-year prediction. ROC, receiver operating characteristic. Color images are available online.

Table 3. Correlation of Risk Score and Clinicopathological Characteristics in Oral Squamous Cell Carcinoma Patients

Characteristic	Number of patients			
Total patients	Low risk 128	High risk 129	χ^2	p
Age (years)				
<60	58 (60.3)	63 (60.7)	0.320	0.571
≥60	70 (67.7)	66 (68.3)		
Sex				
Male	93 (89.6)	87 (90.4)	0.832	0.362
Female	35 (38.4)	42 (38.6)		
Grade				
G1+G2	87 (98.6)	111 (99.4)		
G3+G4	37 (26.4)	16 (26.6)	11.951	0.002
Gx	4 (3)	2 (3)		
Clinical-stage				
I+II	36 (25.9)	16 (26.1)		
III+IV	78 (89.6)	102 (90.4)	11.249	0.004
NA	14 (12.5)	11 (12.5)		
Clinical-T				
T1+T2	64 (49.8)	36 ((50.2)		
T3+T4	52 (67.7)	84 (68.3)	15.794	0.000
Tx	12 (10.5)	9 (10.5)		
Clinical-N				
N0	54 (47.3)	41 (47.7)		
N1+N2	53 (62.8)	73 (63.2)	6.009	0.050
Nx	20 (16.9)	14 (17.1)		

NA, not available; T, tumor status; N, lymph node status.

significantly different from those low-risk scores relating to the AJCC stage, clinical grade, and AJCC-T. However, there was no difference in age, sex, and AJCC-N observed between the patients with high- and low-risk scores. Patients with high-risk scores had worse clinical stage disease (III+IV) and larger tumor sizes (T3+T4) compared to those with low-risk scores. Also, G1 and G2 histology were more commonly seen in the high-risk group instead of the low-risk group (shown in Table 3).

To verify the independent prognostic value of the 5-mRNA signature, univariate Cox regression analysis was performed to test the performance of the signature. Our results demonstrated that risk score (HR = 3.497, CI: 2.279–5.677, $p < 0.001$), age (HR = 1.020, CI: 1.002–1.038, $p < 0.05$), AJCC-T (HR = 1.483, CI: 1.096–2.006, $p < 0.05$), and AJCC stage (HR = 1.466, CI: 1.030–2.087, $p < 0.05$) were associated with overall survival. Subsequently, risk scores, age, and AJCC-T were used as covariates and overall survival as the dependent variable in multivariate Cox regression analysis. Our results indicated that only the risk score was significant (HR = 3.747, CI: 2.279–5.677, $p < 0.001$) compared to the clinical characteristics (shown in Table 4). In summary, our 5-mRNA signature can be used as an independent prognostic indicator in OSCC and patients with high-risk scores tend to have worse outcomes.

Construction and analysis of the PPI network and GSEA

To provide important bioinformatics evidence for investigating the regulatory molecular mechanism of OSCC, functional enrichment analysis and PPI network analysis were performed using the STRING database. GO and KEGG pathway enrichment analysis revealed that the identified upregulated DEGs in the OSCC tissues were enriched in pathways relating to mitosis nuclear division, cell division, positive regulation of ubiquitin-protein ligase activity, and microRNAs in cancer. The downregulated DEGs were significantly enriched in GO terms such as axoneme and extracellular exosomes, and KEGG terms, including saliva secretion and the calcium, PPAR, glucagon, and cGMP-PKG signaling pathways (shown in Supplementary Table S2). The PPI network of the DEGs was developed using 224 node interactions with combined scores ≥0.9. The functional modules were evaluated using the cytoHubba plug-in. We calculated the top 20 hubba nodes by MCC for further analysis. The module consisted of 20 nodes and 156 edges, including *CDC20*, *BUB1B*, *AURKB*, and *CDK1* (shown in Fig. 4). Further functional analysis showed that genes in this module were mainly enriched in cell division, the anaphase-promoting complex-dependent catabolic process, and cell proliferation. The genes were significantly related to the cell cycle and the p53 and FoxO signaling pathways in KEGG analysis (shown in Table 5).

To identify the signaling pathways that may be activated in OSCC, GSEA was conducted among high and low *HOXA1*, *CELSR3*, *HIST1H3J*, and *ZFP42* expression data sets ($p < 0.05$). The top 5 enriched tumor-associated signaling pathways dependent on the normalized enrichment scores were selected. The NOTCH and WNT signaling pathways

Table 4. Univariate and Multivariate COX Regression Analyses of the Messenger RNA Signature and Survival

Variables	Unfavorable/favorable	Univariate analysis		Multivariate analysis	
		HR (95% CI)	p	HR (95% CI)	p
Risk score	High/low	3.497 (2.279–5.677)	3.864E-8	3.506 (2.214–5.554)	8.992E-8
Age	≥60/<60	1.020 (1.002–1.038)	0.026	1.378 (0.917–2.072)	0.123
Sex	Male/female	1.220 (0.802–1.856)	0.352		
Grade	G3+G4/G1+G2	1.061 (0.819–1.373)	0.654		
Clinical stage	III+IV/I+II	1.466 (1.030–2.087)	0.034		
Clinical-T	T3+T4/T1+T2	1.483 (1.096–2.006)	0.011	1.365 (0.985–1.892)	0.062
Clinical-N	N1+N2/N0	1.289 (0.969–1.716)	0.082		

CI, confidence interval.

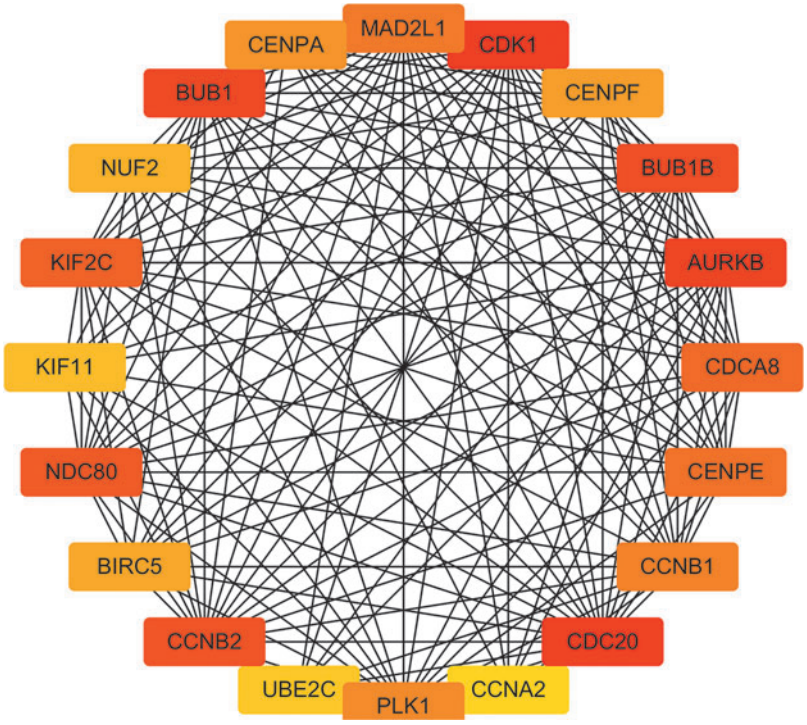

FIG. 4. The first 20 genes network. The first 20 genes of the MMC method were chosen using CytoHubba plug-in. The more forward ranking is represented by a redder color. Color images are available online.

were differentially enriched in the *HOXA1* high expression phenotype. The VEGF signaling pathway was differentially enriched in the *CELSR3* high expression phenotype. Cell cycle and base excision repair were differentially enriched in the *HIST1H3J* low expression phenotype. Primary immunodeficiency was differentially enriched in the *ZFP42* high expression phenotype. DNA replication was differentially enriched in the *ASCL4* high expression phenotype. Of particular interest, it was observed that mismatch repair was enriched in CELSR3 and ASCL4 (shown in Fig. 5).

Validation of the 5-mRNA signature in OSCC tissues

The expression levels of the 5 mRNAs (HOXA1, CELSR3, HIST1H3J, ZFP42, andASCL4) were determined in 20 paired OSCC and adjacent healthy tissues by qRT-PCR. HOXA1, CELSR3, and HIST1H3J were significantly upregulated in tumor tissues ($p < 0.05$), while ZFP42 and ASCL4 showed an upregulated trend that was not statistically significant (shown in Fig. 6). The clinical characteristics of 20 samples are summarized in Supplementary Table S3.

Discussion

In this study, we established a 5-mRNA signature by analyzing the expression profiles downloaded from TCGA. This signature included three high-risk mRNAs (HOXA1, HIST1H3J, and ZFP42) and two protective mRNAs (CELSR3 and ASCL4). To improve the accuracy of the mRNA signature, we used a combination of multiple mRNAs associated

Table 5. Significantly Enriched Gene Ontology Terms and Kyoto Encyclopedia of Genes and Genomes Pathways for Top 20 Hub Genes in the Protein–Protein Interaction Networks

	Description	*Number of enriched genes*	p
GO terms			
GO:0007062	Sister chromatid cohesion	14	2.08E-25
GO:0051301	Cell division	17	8.23E-25
GO:0031145	Anaphase-promoting complex-dependent catabolic process	8	1.87E-12
GO:0051437	Positive regulation of ubiquitin-protein ligase activity involved in regulation of mitotic cell cycle transition	7	1.82E-10
GO:0007059	Chromosome segregation	5	9.11E-07
GO:0008283	Cell proliferation	7	2.20E-06
KEGG terms			
hsa04110	Cell cycle	9	3.88E-13
hsa04115	p53 signaling pathway	3	0.0040
hsa04068	FoxO signaling pathway	3	0.0153
hsa05203	Viral carcinogenesis	3	0.0340

GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

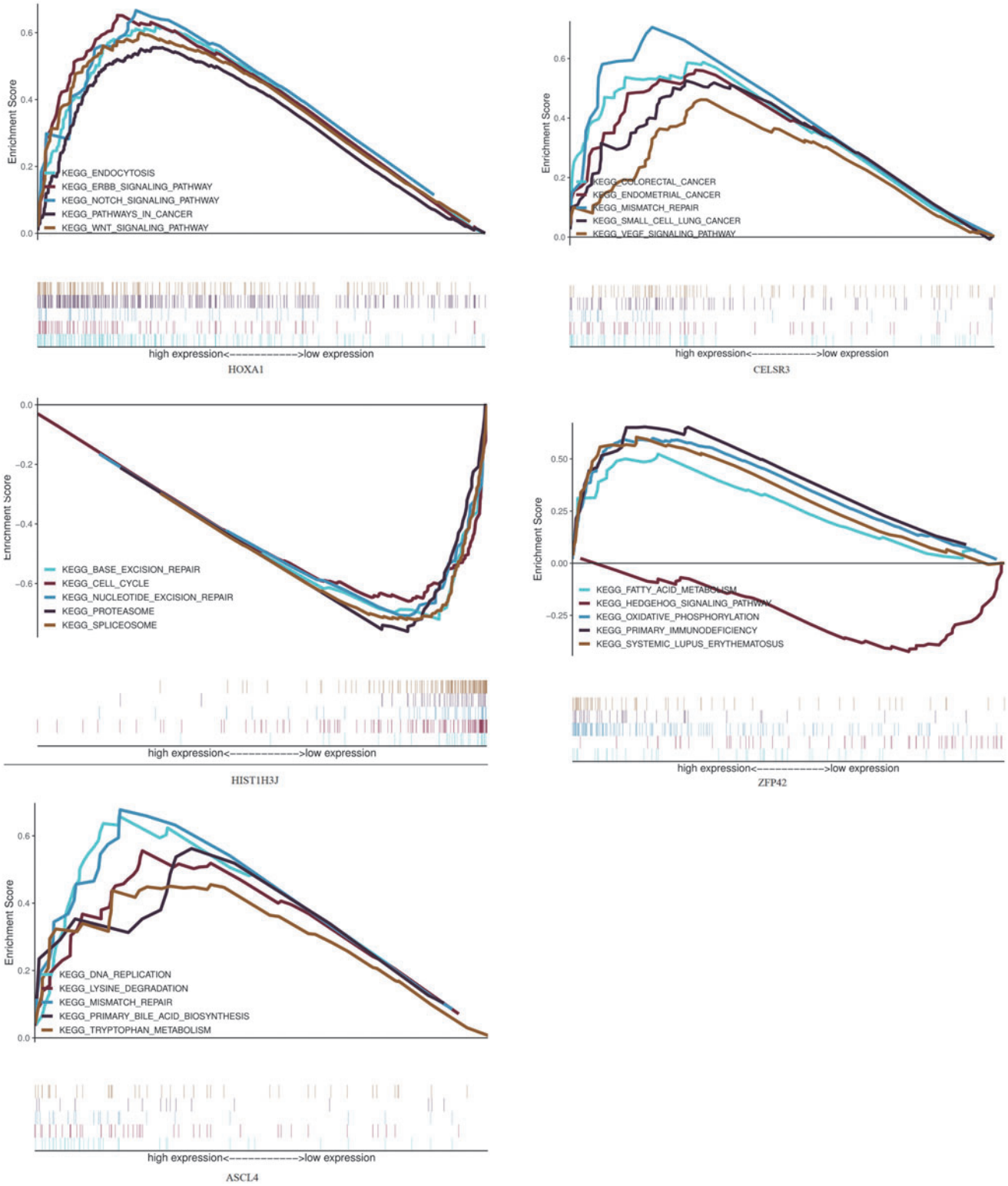

FIG. 5. Enrichment plots of the five signature-related mRNAs from GSEA. GSEA, gene set enrichment analysis. Color images are available online.

with overall survival. The molecular functions of these abnormal mRNAs were identified by GSEA, GO, and KEGG pathway analysis and our results were verified by RT-PCR.

Emerging evidence shows that molecular markers can be used in the diagnosis and prognosis of cancer (Lin *et al.*, 2010; Arun *et al.*, 2018; He *et al.*, 2018; Giraldez *et al.*, 2019). Major efforts have been made to find suitable biomarkers that can predict survival in OSCC patients. It has been reported that miR-196b contributes to the progression of OSCC by accelerating tumor cell migration and invasion

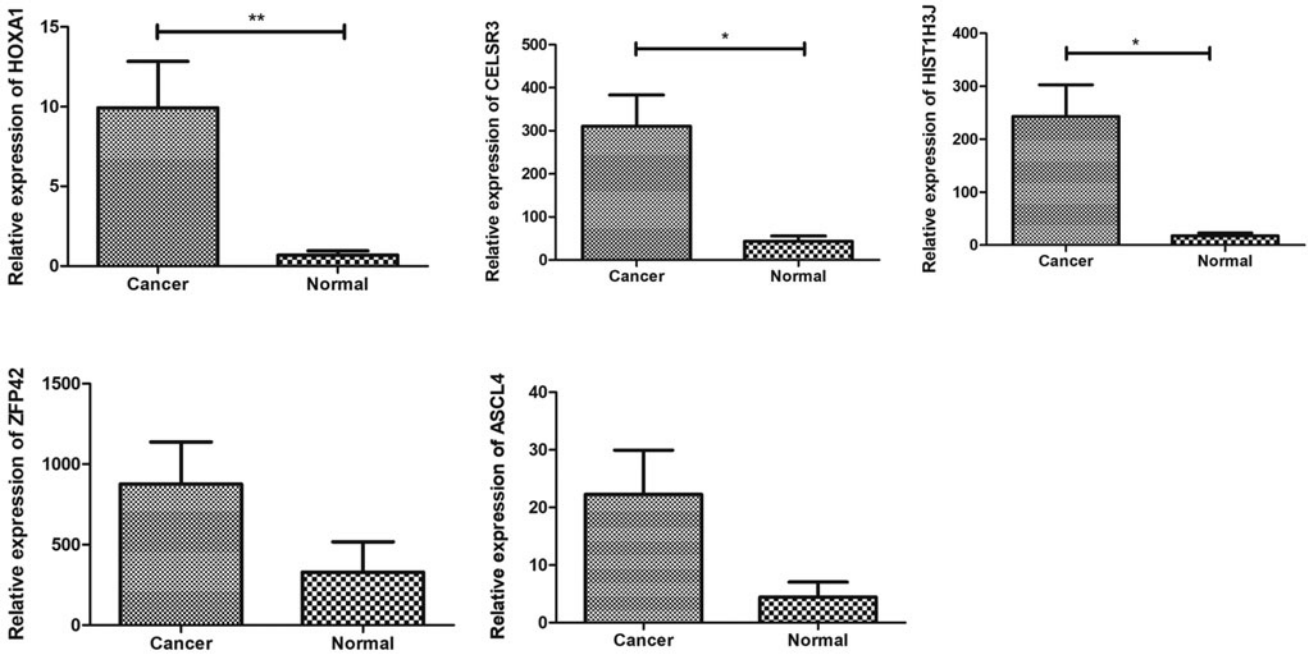

FIG. 6. qRT-PCR results of the five signature-related mRNAs. Expression of these mRNAs was normalized against GAPDH expression. *$p < 0.05$; **$p < 0.001$. qRT-PCR, quantitative real-time polymerase chain reaction.

(Hou *et al.*, 2016). Wang *et al.* (2015) analyzed the expression profiles of OSCC patients with extracapsular spread and obtained a prognostic signature based on 11 genes that were used to predict the prognosis of patients without nodal metastases. However, predicting the prognosis of OSCC patients still requires stable biomarkers and reliable prognostic models. The expression levels of mRNA are higher than long noncoding RNA and so mRNA biomarkers may not introduce bias.

In agreement with our data, a previous study of HOXA1 showed that it can contribute to oral carcinogenesis by increasing proliferation and may have potential as a prognostic marker in OSCC (Bitu *et al.*, 2012). The precise role of HOXA1 in tumors remains unclear. Several studies have reported that HOXA1 functions as an oncogene and is an independent prognostic indicator in various tumors, including gastric cancer and hepatocellular carcinoma (Zha *et al.*, 2012; Yuan *et al.*, 2016). In contrast, studies have demonstrated a tumor suppressor role of HOXA1. Inhibition of HOXA1 expression promotes the invasiveness of pancreatic cancer cells (Ohuchida *et al.*, 2012) and low HOXA1 expression is associated with a poor prognosis in small cell lung cancer (Xiao *et al.*, 2014). Taken together, these data suggest that HOXA1 may play different roles in carcinogenesis and cancer progression depending on the type of tumor.

CELSR3 is a member of the flamingo protein subfamily that is correlated with the WNT/PCP signaling pathway. Khro *et al.* (2016) reported that hypermethylation of the CELSR3 promoter decreases gene expression in OSCC and may have an important role in oral carcinogenesis. Also, dysregulation of CELSR3 has been reported as a biomarker for prognosis in various cancers (Karpathakis *et al.*, 2016; Gu *et al.*, 2019). A study has shown that aberrant methylation of HISTIH3J is a prognostic biomarker in human papillary thyroid cancer (Kikuchi *et al.,* 2013), yet the biological role of HISTIH3J in tumorigenesis remains to be fully determined.

The ZFP42 gene is currently widely used as a marker of embryonic stem cells. ASCL4 is crucial in determining cell fates as well as in the development and differentiation of many tissues (Jonsson *et al.*, 2004). To verify the results of bioinformatics analysis, we used qRT-PCR to analyze the expression levels of these genes in 20 pairs of OSCC and adjacent normal tissues. HOXA1, CELSR3, and HIST1H3J showed the same trends in expression as predicted, verifying the accuracy of our method. However, the expression of ZFP42 and ASCL4 was not significantly different between OSCC and normal tissues, which may potentially be due to the small sample size used in this study.

We further investigated the molecular mechanism of DEmRNAs and the 5 mRNAs in OSCC by performing pathway enrichment analysis. The GO analysis showed that the DEmRNAs were mainly enriched in the extracellular exosome and extracellular space of cell component category indicating that cell-to-cell communication is critical for OSCC progression. Interestingly, muscle contraction had the highest enrichment score in the biological process category. We assume that muscle-related genes may play an important role in the control of cellular locomotion and cytoplasmic streaming, and cytokinesis in non-muscle cells. Further studies are needed to elucidate the role of muscle-related genes in OSCC carcinogenesis.

The KEGG pathway analysis showed that the most affected pathways were related to some cancer pathways. For example, the calcium signaling pathway, which has an important role in cell proliferation, death, invasion, and metastasis, has been characterized in various cancer types. These data suggest potential opportunities for targeting altered calcium signaling during tumorigenesis (Monteith *et al.*, 2017).

Previous studies have demonstrated that the AMPK signaling pathway participates in the regulation of cell growth and reprogramming of metabolism, and has recently been connected to autophagy and cell polarity (Mihaylova and Shaw, 2011). Chang *et al*. (2013) showed that the AMPK signaling pathway controls apoptotic and autophagic cell death in OSCC cell lines. To further investigate the functions of the 5 mRNAs in OSCC, we performed GSEA using TCGA data. GSEA showed that the NOTCH and WNT signaling pathways, the VEGF signaling pathway, cell cycle, and mismatch repair were significantly enriched ($p < 0.05$). Previously, a progressive reduction in DNA mismatch repair proteins (*hMLH1*, *hMSH2*, and *hPMS2*) has been reported from mild, moderate, and severe dysplasia to OSCC (Jessri *et al.*, 2015). Therefore, we hypothesize that the 5 mRNAs may have interactions with these signaling pathways and cell cycle and mismatch repair, and play important roles in OSCC.

In summary, this study established a 5-mRNA signature (HOXA1, CELSR3, HIST1H3J, ZFP42, and ASCL4) using comprehensive bioinformatics analyses to identify potential biomarkers to predict progression in OSCC. Further analysis revealed the function and mechanism of these mRNAs. The data presented in this study are limited due to the small sample size of the control group that was used for DEmRNA analysis. Our data require further validation in larger patient cohorts and should be explored in future studies.

Ethics Approval and Consent to Participate

This study was approved by the Ethics Committee of Guangxi Medical University. Informed consent was obtained from all individual participants included in the study.

Consent for Publication

Written informed consent for publication was obtained from each participant.

Availability of Data and Materials

The datasets used and/or analyzed during this study are available from the corresponding author on reasonable request

Authors' Contributions

X.S. and X.H. designed the experiments and analyzed the data; H.G. analyzed the data; H.G. performed mRNA functional predictions; C.L. performed survival analysis; and all authors read and approved the article and agree to be accountable for all aspects of the research in ensuring that the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgments

The authors would like to thank all the reviewers who participated in the review and MJEditor for its linguistic assistance during the preparation of this article.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

This work was supported by grants from the Youth Science Foundation of Guangxi Medical University (GXMUY SF201927).

Supplementary Material

Supplementary Table S1
Supplementary Table S2
Supplementary Table S3
Supplementary Figure S1
Supplementary Figure S2
Supplementary Figure S3

References

Arrieta O, Pineda B, Muñiz-Hernández S, *et al*. (2014) Molecular detection and prognostic value of epithelial markers mRNA expression in peripheral blood of advanced non-small cell lung cancer patients. Cancer Biomark 14:215–223.

Arun G, Diermeier SD, Spector DL (2018) Therapeutic targeting of long non-coding RNAs in cancer. Trends Mol Med 24:257–277.

Bitu CC, Destro MF, Carrera M, *et al*. (2012) HOXA1 is overexpressed in oral squamous cell carcinomas and its expression is correlated with poor prognosis. BMC Cancer 12: 146.

Cai J, Tong Y, Huang L, Xia L, *et al*. (2019) Identification and validation of a potent multi-mRNA signature for the prediction of early relapse in hepatocellular carcinoma. Carcinogenesis 40:840–852.

Chang HW, Lee YS, Nam HY, *et al*. (2013) Knockdown of β-catenin controls both apoptotic and autophagic cell death through LKB1/AMPK signaling in head and neck squamous cell carcinoma cell lines. Cell Signal 25:839–847.

Chen SL, Qin ZY, Hu F, *et al*. (2019) The role of the HOXA gene family in acute myeloid leukemia. Genes (Basel). 10: 621.

da Costa AABA, Costa FD, Araújo DV, *et al*. (2018) The roles of PTEN, cMET, and p16 in resistance to cetuximab in head and neck squamous cell carcinoma. Med Oncol 36:8.

Dong Z, Wang J, Zhan T, Xu S (2018) Identification of prognostic risk factors for esophageal adenocarcinoma using bioinformatics analysis. Onco Targets Ther 11:4327–4337.

Fan QC, Tian H, Wang Y, Liu XB (2019) Integrin-α5 promoted the progression of oral squamous cell carcinoma and modulated PI3K/AKT signaling pathway. Arch Oral Oncol 101:85.

Ferlay J, Soerjomataram I, Dikshit R, *et al*. (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 136:E359–E386.

Giraldez MD, Spengler RM, Etheridge A, *et al*. (2019) Phospho-RNA-seq: a modified small RNA-seq method that reveals circulating mRNA and lncRNA fragments as potential biomarkers in human plasma. EMBO J 38:e101695.

Gu X, Li H, Sha L, *et al*. (2019) CELSR3 mRNA expression is increased in hepatocellular carcinoma and indicates poor prognosis. PeerJ 7:e7816.

He B, Lin X, Tian F, *et al*. (2018) miR-133a-3p inhibits oral squamous cell carcinoma (OSCC) proliferation and invasion by suppressing COL1A1. J Cell Biochem 119:338–346.

Hou YY, You JJ, Yang CM, *et al*. (2016) Aberrant DNA hypomethylation of miR-196b contributes to migration and invasion of oral cancer. Oncol Lett 11:4013–4021.

Jemal A, Bray F, Center MM, *et al*. (2011) Global cancer statistics. CA Cancer J Clin 61:69–90.

Jessri M, Dalley AJ, Farah CS (2015) hMSH6: a potential diagnostic marker for oral carcinoma in situ. J Clin Pathol 68: 86–90.

Jonsson M, Björntorp Mark E, Brantsing C, *et al*. (2004) Hash4, a novel human achaete-scute homologue found in fetal skin. Genomics 84:859–866.

Karpathakis A, Dibra H, Pipinikas C, *et al*. (2016) Prognostic impact of novel molecular subtypes of small intestinal neuroendocrine tumor. Clin Cancer Res 22:250–258.

Khor GH, Froemming GR, Zain RB, *et al*. (2016) Involvement of CELSR3 hypermethylation in primary oral squamous cell carcinoma. Asian Pac J Cancer Prev 17:219–223.

Kikuchi Y, Tsuji E, Yagi K, *et al*. (2013) Aberrantly methylated genes in human papillary thyroid cancer and their association with BRAF/RAS mutation. Front Genet 4:271.

Li Y, St John MA, Zhou X, *et al*. (2004) Salivary transcriptome diagnostics for oral cancer detection. Clin Cancer Res 10: 8442–8450.

Lin SC, Liu CJ, Lin JA, *et al*. (2010) miR-24 up-regulation in oral carcinoma: positive association from clinical and in vitro analysis Oral Oncol 46:204–208.

Brooks M, Mo Q, Krasnow R, *et al*. (2016) Positive association of collagen type I with non-muscle invasive bladder cancer progression. Oncotarget 7:82609–82619.

Michailidou E, Tzimagiorgis G, Chatzopoulou F, *et al*. (2016) Salivary mRNA markers having the potential to detect oral squamous cell carcinoma segregated from oral leukoplakia with dysplasia. Cancer Epidemiol 43:112–118.

Mihaylova MM, Shaw RJ (2011) The AMPK signalling pathway coordinates cell growth, autophagy and metabolism. Nat Cell Biol 13:1016–1023.

Monteith GR, Prevarskaya N, Roberts-Thomson SJ (2017) The calcium-cancer signalling nexus. Nat Rev Cancer 17:367–380.

Ohuchida K, Mizumoto K, Lin C, *et al*. (2012) MicroRNA-10a is overexpressed in human pancreatic cancer and involved in its invasiveness partially via suppression of the HOXA1 gene. Ann Surg Oncol 19:2394–2402.

Patrick JH, Saha-Chaudhuri P (2013) SurvivalROC: Time-dependent ROC curve estimation from censored survival data. R package version 1.0.3. Available online at: https://CRAN.R-project.org/package=survivalROC (accessed October 20, 2019).

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.

Villagómez-Ortíz VJ, Paz-Delgadillo DE, Marino-Martínez I, *et al*. (2016) [Prevalence of human papillomavirus infection in squamous cell carcinoma of the oral cavity, oropharynx and larynx]. Cir Cir 84:363–368.

Wang W, Lim WK, Leong HS, *et al*. (2015) An eleven gene molecular signature for extra-capsular spread in oral squamous cell carcinoma serves as a prognosticator of outcome in patients without nodal metastases. Oral Oncol 51:355–362.

Xiao F, Bai Y, Chen Z, *et al*. (2014) Downregulation of HOXA1 gene affects small cell lung cancer cell survival and chemoresistance under the regulation of miR-100. Eur J Cancer 50:1541–1554.

Yuan C, Zhu X, Han Y, *et al*. (2016) Elevated HOXA1 expression correlates with accelerated tumor cell proliferation and poor prognosis in gastric cancer partly via cyclin D1. J Exp Clin Cancer Res 35:15.

Zha TZ, Hu BS, Yu HF, *et al*. (2012) Overexpression of HOXA1 correlates with poor prognosis in patients with hepatocellular carcinoma. Tumour Biol 33:2125–2134.

Address correspondence to:
Xiaoping Su, PhD
Department of Oral and Maxillofacial Surgery
The Affiliated Stomatology Hospital
of Guangxi Medical University
Guangxi Medical University
Number 22 Shuangyong Road
Nanning 530021
P.R. China

E-mail: 417863731@qq.com

Xuanping Huang, MD
Department of Oral and Maxillofacial Surgery
The Affiliated Stomatology Hospital
of Guangxi Medical University
Guangxi Medical University
Number 22 Shuangyong Road
Nanning 530021
P.R. China

E-mail: hxp120@126.com