



OPEN

A ceRNA-associated risk model predicts the poor prognosis for head and neck squamous cell carcinoma patients

Yuzi Xu^{1,5}, Fengqin Xu^{3,5}, Yiming Lv^{4,5}, Siyuan Wang¹, Jia Li¹, Chuan Zhou¹, Jimin Jiang¹, Binbin Xie²✉ & Fuming He¹✉

Head and neck squamous cell carcinoma (HNSCC) is one of the most malignant cancers with poor prognosis worldwide. Emerging evidence indicates that competing endogenous RNAs (ceRNAs) are involved in various diseases, however, the regulatory mechanisms of ceRNAs underlying HNSCC remain unclear. In this study, we retrieved differentially expressed long non-coding RNAs (DElncRNAs), messenger RNAs (DEmRNAs) and microRNAs (DEmiRNAs) from The Cancer Genome Atlas database and constructed a ceRNA-based risk model in HNSCC by integrated bioinformatics approaches. Functional enrichment analyses showed that DEmRNAs might be involved in extracellular matrix related biological processes, and protein–protein interaction network further selected out prognostic genes, including MYL1 and ACTN2. Importantly, co-expressed RNAs identified by weighted co-expression gene network analysis constructed the ceRNA networks. Moreover, AC114730.3, AC136375.3, LAT and RYR3 were highly correlated to overall survival of HNSCC by Kaplan–Meier method and univariate Cox regression analysis, which were subsequently implemented multivariate Cox regression analysis to build the risk model. Our study provides a deeper understanding of ceRNAs on the regulatory mechanisms, which will facilitate the expansion of the roles on the ceRNAs in the tumorigenesis, development and treatment of HNSCC.

Head and neck squamous cell carcinoma (HNSCC) is one of the most common and malignant cancers with a yearly occurrence of 750,000 cases worldwide and 40–50% mortality^{1,2}. HNSCC is characterized by a high propensity of recurrence and frequent metastasis developing in more than 65% of patients³ and second primary cancers at an annual rate of approximately 2–3%⁴. HNSCC can be staged according to tumor-node-metastasis (TNM) staging method, and the mainstays of therapeutic strategy is surgery in combination with radio- and chemotherapy for early-stage tumors⁵. Nevertheless, the survival rate of patients sharply declines once the disease has metastasized⁶. Systemic therapy with active agents including platinum, taxanes, antifolates, and cetuximab has remarkably improved the prognosis for a limited number of patients, thanks to the highlights of signature molecular markers⁷. For instance, Pavon et al. reported that overexpressed uPA/uPAR and SERPINE1 in HNSCC patients enhanced tumor cell proliferation, migration, and invasion⁸. And Hersi et al. discovered that plerixafor reversed the downregulation of miR-9, hence suppressing cellular proliferation, cell cycle progression and colony formation⁹. However, the clinical efficacy of these biomarkers is hardly convincing, with low 5-year overall survival rates and high mortality rate³. As a result, it is of great importance to further investigate more effective biomarkers, signaling the incidence of HNSCC metastasis and poor prognosis, thus improving long-term curative effect.

¹Department of Oral Implantology and Prosthodontics, The Affiliated Hospital of Stomatology, School of Stomatology, Zhejiang University School of Medicine, and Key Laboratory of Oral Biomedical Research of Zhejiang Province, 395# Yanan Road, Hangzhou 310006, Zhejiang, People's Republic of China. ²Department of Medical Oncology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, 3# East Qingchun Road, Hangzhou 310016, Zhejiang, People's Republic of China. ³The First Affiliated Hospital of Kangda College of Nanjing Medical University, The First People's Hospital of Lianyungang, The Affiliated Lianyungang Hospital of Xuzhou Medical University, Lianyungang 222000, Jiangsu, People's Republic of China. ⁴Department of Colorectal Surgery, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou 310016, Zhejiang, People's Republic of China. ⁵These authors contributed equally: Yuzi Xu, Fengqin Xu and Yiming Lv. ✉email: 387609530@qq.com; hfm@zju.edu.cn

The application of sequencing technology and the large-scale database have recently brought about novel approaches of cancer diagnosis, treatment and prognosis. Investigations about non-coding RNAs (ncRNAs) challenge the conventional perspectives on the predominant role of transcribed RNAs during protein synthesis¹⁰. Due to absence of open reading frames, ncRNAs such as long non-coding RNA (lncRNA), pseudogenes and circular RNA don't have the ability to translate into proteins, but can act as competing endogenous RNAs (ceRNAs) to perform the post-transcriptional regulation by communicate with or co-regulate each other^{11,12}. The milestone conception of ceRNA was put forward by Salmena et al. in 2011¹³. In this hypothesis, miRNA response elements (MREs) play an integral role for miRNA competitively binds to ncRNAs, by which miRNA also binds to mRNAs¹³. More generally, any RNA transcript with MREs has the potential to function as a ceRNA and repress other RNAs' activity with similar MREs. Ongoing researches have recently confirmed the role of ceRNA network in comprehensively clarifying complex gene interactions and identifying potential biomarkers for cancer diagnosis, treatment and prognosis of gastric¹⁴, breast¹⁵, pancreas¹⁶, bladder¹⁷ and HNSCC^{18,19}. Fang et al. proposed a regulatory mechanism and indicated lncRNAs might have an important impact on the survival and prognosis of HNSCC patients¹⁸. Pan et al. comprehensively analyzed lncRNA-associated ceRNA network and identified several mRNAs and miRNAs as prognostic biomarkers of HNSCC¹⁹. These studies provide a better understanding how the ceRNAs contribute to improving the diagnostic and prognostic efficiency for HNSCC patients. However, the co-expression among ceRNAs has been overlooked. Weighted co-expressed gene network analysis (WGCNA) clusters highly correlated genes into one module and relates it to clinical traits, which has already been a widely-accepted tool to identify clinical biomarkers for diagnosis and therapy²⁰. Previous studies have so far identified effective and reliable biomarkers for diagnosis and prognosis of HNSC patients using WGCNA approach, including CNFN²¹, APP and COL1A2²², which could be exploited as novel therapeutic targets for HNSCC. Consequently, we believe integral analysis of ceRNA in combination with WGCNA can be more beneficial in underlying mechanisms of cancer metastasis and prognosis, which has been rarely reported yet.

In the present study, we retrieved HNSCC-related RNA-seq data and miRNA-seq data from The Cancer Genome Atlas (TCGA) database and screened out differentially expressed lncRNAs (DElncRNAs), mRNAs (DEmRNAs), and miRNAs (DEmiRNAs) to successfully develop a ceRNA-based risk model for HNSCC patients. Though the therapeutic efficiency of HNSCC-specific prognostic signatures temporarily lacks validation by large-scale studies, our findings may provide novel insights toward developing a promising predictive tool for the metastasis and prognosis of HNSCC and lay the groundwork for further research in unveiling mechanisms of tumor progression and improving overall survival in HNSCC patients.

Results

Identification of DElncRNAs, DEmRNAs and DEmiRNAs. A schematic of the workflow of this work is shown in Fig. 1. The clinical features of 502 HNSCC patients were demonstrated in Fig. 2. We collected level 3 HNSCC-related RNA-seq count data of 502 HNSCC samples and 44 normal samples and level 3 miRNA-seq count data of 525 HNSCC samples and 44 normal samples from the National Cancer Institute's Genomic Data Commons data portal. In total, 25,295 lncRNAs, 19,601 mRNAs and 1880 miRNAs were extracted to implement normalization and variance-stabilizing transformation by "DESeq2" package. Using "DESeq2" package to perform differentially expressed gene analysis (DEGA), we identified DElncRNAs, DEmRNAs and DEmiRNAs between HNSCC samples and normal samples with the cut-off of $|\log_2(\text{foldchange})| (|\log_2\text{FC}|) \geq 1$ and adjusted P value < 0.05 . In total, 5749 DElncRNAs were screened out, including 2243 downregulated and 3506 upregulated lncRNAs in HNSCC samples. Of the 4790 DEmRNAs, 2418 were downregulated and 2372 were upregulated. Among 303 DEmiRNAs, there were 126 downregulated miRNAs and 177 upregulated miRNAs. The volcano plots displayed the distributions of DElncRNAs, DEmRNAs and DEmiRNAs (Fig. 3a) and the heat maps manifested the expression levels of these genes (Fig. 3b).

Functional enrichment analyses of DEmRNAs. In order to shed light on biological function of DEmRNAs ($|\log_2\text{FC}| \geq 1$, adjusted P value < 0.05) in HNSCC, we implemented Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses with a cut-off of P value < 0.05 . 716 BPs, 140 CCs and 153 MFs were involved in GO enrichment results, and 51 KEGG pathways were significantly enriched, of which the top 10 GO terms and KEGG pathways are listed in Table 1. "Extracellular matrix" related GO terms were significantly enriching most DEmRNAs, and the most significant biological process (BP), cellular component (CC), and molecular function (MF) were Extracellular matrix organization (GO:0030198, $P = 7.24\text{E}-22$), Extracellular matrix (GO:0031012, $P = 5.10\text{E}-23$) and Extracellular matrix structural constituent (GO:0005201, $P = 6.84\text{E}-19$), respectively. Several studies have previously reported that bi-directional interactions between cancer-associated fibroblast and cancer cell depend on ECM remodeling mechanisms in HNSCC^{23,24}. In the absence of natural control and balance, malignant cells are free to manipulate the ECM and eventually lead to metastasis²⁵. Most DEmRNAs fell into neuroactive ligand-receptor interaction (hsa04080, $P = 4.27\text{E}-11$) pathway. Neuronal activity is regulated by neuroactive steroids that function as mediators of neurotransmitter receptors²⁶. Previous researches have reported neuroactive steroids to modulate γ -aminobutyric acid (GABA) receptors that have been suggested to affect oncogenesis by controlling cell proliferation²⁷. GABA_A receptors, one of the subclasses of GABR, are a class of transmembrane ligand-gated chloride channels, which are named based on their subunits composition: alpha (GABRA1-6), beta (GABRB1-3), gamma (GABRG1-3), delta (GABRD), epsilon (GABRE), theta (GABRQ), pi (GABRP) and rho (GABRR1-3)²⁸. Among the significantly differentially expressed genes, our pathway enrichment analysis suggested that eight GABRs were involved in the neuroactive ligand-receptor interaction pathway, including GABRA4 ($\log_2\text{FC} = -3.09$), GABRB2 ($\log_2\text{FC} = -1.09$), GABRG2 ($\log_2\text{FC} = 4.80$), GABRD ($\log_2\text{FC} = 1.86$), GABRE ($\log_2\text{FC} = 1.52$), GABRP ($\log_2\text{FC} = -2.49$), GABRR1 ($\log_2\text{FC} = 1.08$) and GABRR2 ($\log_2\text{FC} = 1.04$).

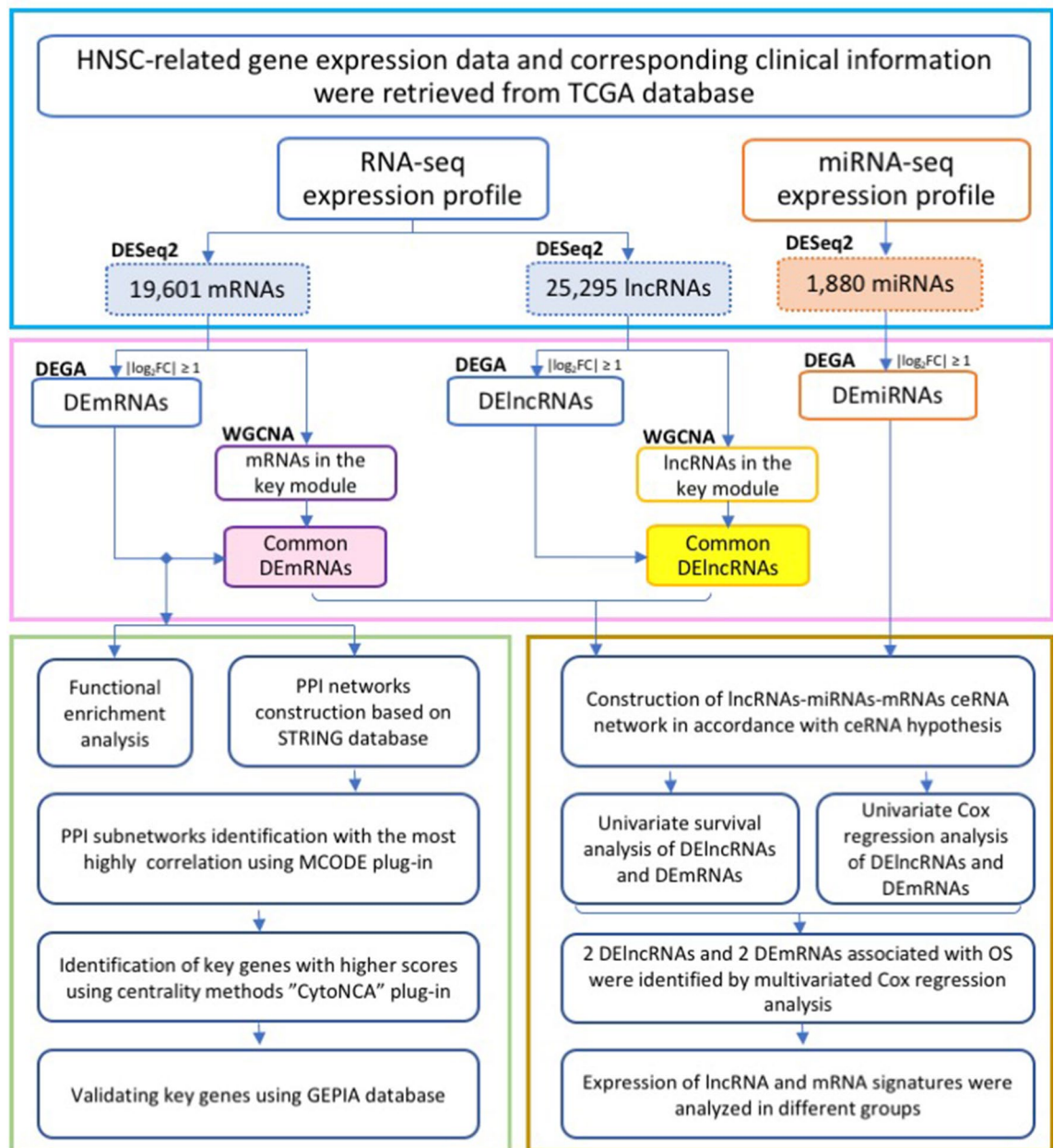


Figure 1. Flow chart of data preparation, processing, analysis and validation in this study.

Construction of PPI networks and validation of key DEmRNAs. To select robust HNSCC-specific mRNAs, we used downregulated and upregulated DEmRNAs with $|\log_2FC| > 2$ to perform PPI analysis, and chose the gene pairs with the highest confidence (the combined score > 0.9) for further analysis, separately. In total, downregulated PPI network included 343 nodes and 1188 edges (Fig. 4a), while upregulated included 256 nodes and 1032 edges (Fig. 4b). Subnetworks selected by MCODE algorithm were displayed and clustered together based on the same cluster number, whose nodes were colored in different shades of colors according to cluster scores. The most significantly correlated module of downregulated PPI network included 22 downregulated DEmRNAs (Fig. 4c), and upregulated module covered 23 upregulated DEmRNAs (Fig. 4d). Functional pathways for each subnetwork were simultaneously established based on the "Wikipathways" database. Next, Centrality analysis of DEmRNAs in each PPI network indicated that top 1% genes obtained by the Centrality methods were MYL1 ($\log_2FC = -3.92$, $P = 2.03E-9$), ACTN2 ($\log_2FC = -3.57$, $P = 1.10E-11$), MYH8 ($\log_2FC = -2.01$, $P = 0.000548$), MYH6 ($\log_2FC = -4.12$, $P = 1.07E-7$), COL1A1 ($\log_2FC = 2.90$, $P = 1.44E-27$), COL1A2 ($\log_2FC = 2.39$, $P = 7.76E-20$) and COL3A1 ($\log_2FC = 2.53$, $P = 7.65E-21$). The seven genes were all clustered in the most significantly correlated module and considered as key DEmRNAs in HNSCC patients. Then as shown in Fig. 4e, the expression calculation based on TCGA database showed that MYL1, ACTN2, MYH8 and MYH6 were significantly downregulated in HNSCC samples, and COL1A1, COL1A2 and COL3A1 were over-expressed. In Fig. 4f, 4 downregulated mRNAs, especially MYL1, demonstrated lower expressions in

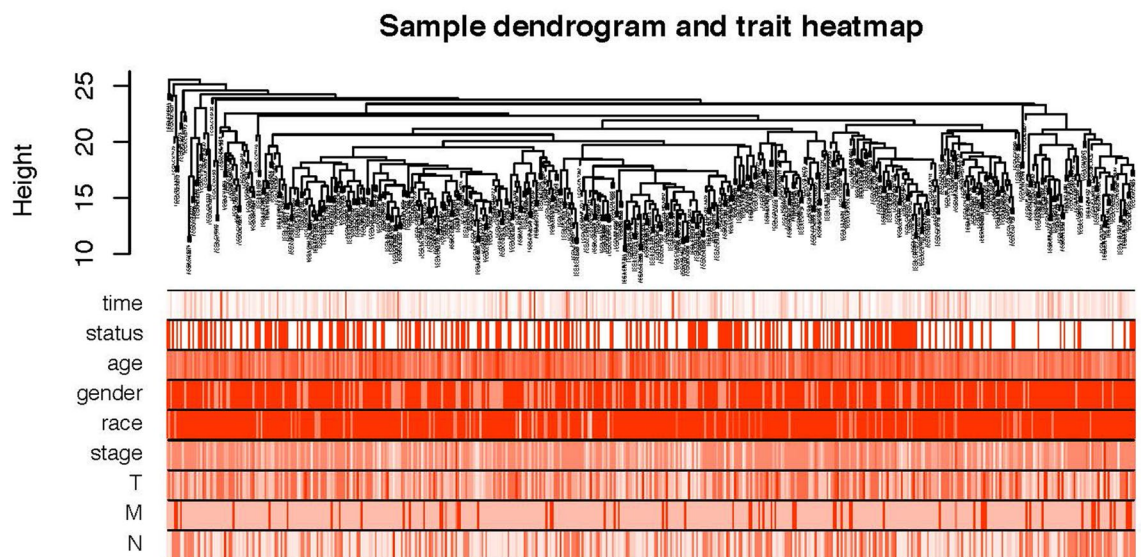


Figure 2. Clustering dendrogram of 502 HNSCC samples and the clinical traits. The red color represented death, female and metastasis. The color intensity was proportional to longer survival time, older age as well as higher tumor stage.

advanced clinical stages when compared to normal samples. On the contrary, higher expressions of 3 upregulated mRNAs were observed in the advanced stages. Consistent with results from our analysis of the TCGA database, the subsequent expression analysis from Gene Expression Profiling and Interactive Analyses (GEPiA) database revealed that MYL1 and ACTN2 were decreasingly expressed in HNSCC patients and COL1A1, COL1A2 and COL3A1 were upregulated (Fig. 4g). Survival analysis showed that high expression of MYL1 (Log-Rank (LR) $P=0.0064$), ACTN2 (LR $P=0.0039$) and MYH8 (LR $P=0.007$) presented worse OS than low expression (Fig. 4h). Finally, MYL1 and ACTN2 were confirmed their prognostic association with HNSCC in TCGA and GEPiA database, respectively, and were subsequently included in further analysis.

Identification of co-expression module by WGCNA. 19,601 mRNAs and 25,295 lncRNAs retrieved from TCGA RNA-seq data were used to establish co-expression networks, respectively. 489 HNSCC patients with complete clinical information were selected to be analyzed using WGCNA algorithm. Pearson correlation matrix among mRNAs were converted into an adjacency matrix that was strengthened by the best power ($\beta=4$) with scale-free topology criterion of $R^2=0.84$ (Fig. 5a). Through minimizing interference from noise and spurious associations by topological overlap measure (TOM), mRNAs with identical expression profiles were grouped together into gene modules. Then, gene modules with 70% similarity were merged into one module by dynamic cutting algorithm (Fig. 5b). From this analysis, 12 gene modules were identified for downstream analysis. Module-trait relationship (MTR) analysis results showed that the magenta module was found to have the highest association with metastasis ($P=2E-13$) and prognosis ($P=0.00002$) in HNSCC patients (Fig. 5c). The module membership in the magenta module (including 307 mRNAs) suggested the correlation with gene significance (GS) was 0.73 ($P=2.4E-52$) for metastasis (Fig. 5d) and 0.5 ($P=8E-21$) for prognosis (Fig. 5e). Parallel processing and analysis were synchronously implemented among lncRNAs, of which the results were showed in Supplementary Figure 1. The yellow module was most significantly correlated with metastasis ($P=1E-14$) and prognosis ($P=0.00003$), including 743 lncRNAs. The metastatic correlation between the module membership of the yellow module and GS of 743 lncRNAs was 0.82 with $P=9.7E-182$, and the prognostic correlation was 0.61 with $P=6.3E-77$. Finally, the magenta module in mRNAs co-expression network and the yellow module in lncRNAs co-expression network were selected for further analysis.

Construction of the ceRNA networks. According to the results of WGCNA and DEGA, 67 out of 307 mRNAs in the magenta module and 241 out of 743 lncRNAs in the yellow module were differentially expressed in HNSCC patients. Then 67 DEmRNAs (17 downregulated and 50 upregulated), 241 DElncRNAs (33 downregulated and 208 upregulated) and 303 DEmiRNAs (126 downregulated and 177 upregulated) were required to construct lncRNA-miRNA-mRNA ceRNA networks. Dependent upon the hypothesis of ceRNA and expression levels of DERNA, ceRNA networks divided into two types: under-expressed and over-expressed networks. In all, the under-expressed ceRNA network included 5 downregulated DElncRNAs, 6 downregulated DEmRNAs and 6 miRNAs (1 upregulated DEmiRNAs) and in combination with 20 edges (Fig. 6a). In the over-expressed ceRNA network, 55 upregulated DElncRNAs and 12 upregulated DEmRNAs were connected by 18 common miRNAs (5 downregulated DEmiRNAs) and 282 edges were involved (Fig. 6b). Among the 6 downregulated DEmRNAs and 12 upregulated DEmRNAs, we found that GABBR1 was involved in the neuroactive ligand-receptor interaction pathway and RYR3 was associated with ion channel activity.

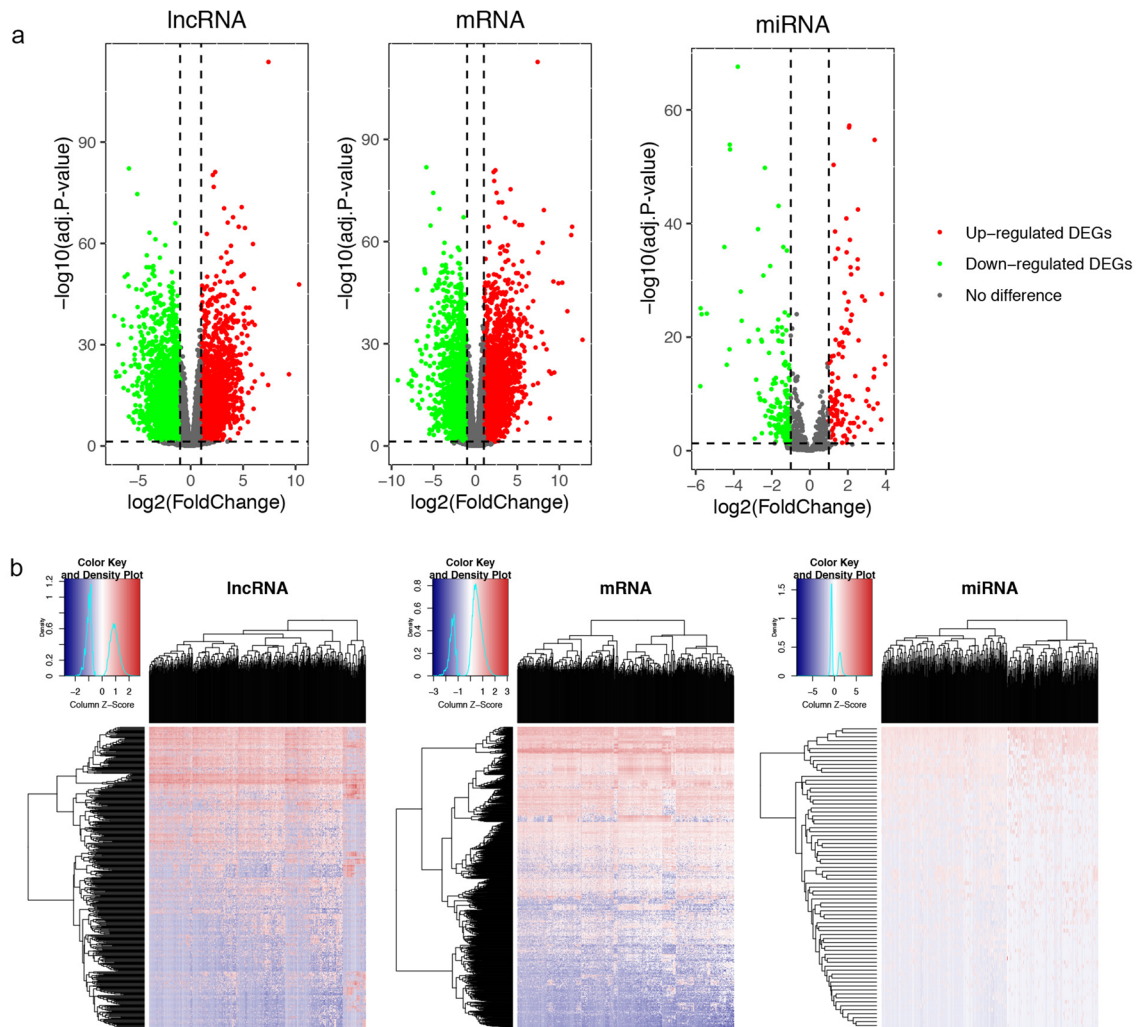


Figure 3. Differentially expressed genes analysis in HNSCC. **(a)** Volcano plots of expressions of lncRNAs, mRNAs and miRNAs. X axis represented the expression differences between HNSCC samples and normal samples, and Y axis represented log transformed adjusted P value; **(b)** heat maps of expression levels of differentially expressed lncRNAs, mRNAs and miRNAs. X axis represented HNSCC samples, and Y axis represented gene markers.

Survival analysis of lncRNA and mRNA signatures. 489 HNSCC patients were divided into high- and low-expression groups according to the median value of gene expression level. The association of overall survival (OS) in HNSCC patients and DELncRNAs and DEMRNAs in ceRNA networks were evaluated by Kaplan–Meier (KM) curve analysis and LR test. Low expression of 11 lncRNAs and 6 mRNAs predicted poorer OS than high expression in HNSCC patients (Supplementary Figure 2). Among them, 1 out of 11 lncRNAs and 4 out of 6 mRNAs were downregulated, while the others were upregulated in HNSCC patients.

Establishment of HNSCC-specific prognostic model for lncRNA and mRNA signatures. Using the univariate regression analysis, we identified potential lncRNAs and mRNAs from ceRNA networks with prognostic values for HNSCC. Among them, 10 lncRNAs and 10 mRNAs were detected to be significantly associated with OS of HNSCC (Table 2). Meanwhile, we found that 6 lncRNAs (AC009148.1, AC073573.1, AC114730.3, AC136475.3, AL139130.1 and AL513320.1) and 6 mRNAs (CAPS, MUC19, RIC3, RYR3, LAT and PRRT2) were simultaneously identified to have prognostic values in LR test and univariate Cox regression analysis. Subsequently, in order to establish HNSCC-specific prognostic model, the multivariate Cox regression analysis were performed among the lncRNAs (AC114730.3 and AC136475.3) and mRNAs (LAT, RYR3) with stronger correlation to OS ($P < 0.01$). As shown in Fig. 7a, the prognostic model was established ($P = 0.00055$). High lncRNA expression of AC114730.3 and AC136375.3, and high mRNA expression of LAT and RYR3 contributed to better OS in HNSCC patients (Fig. 7b). Based on the median value of risk score for all patients, HNSCC patients were divided into high- and low-risk groups (Fig. 7c). In the high-risk group, the mortality rate was significantly higher ($P = 0.0016$) and the prognosis was worse when compared to the low-risk group (Fig. 7d). Furthermore, in Fig. 7e, the area under curve (AUC) of the prognostic model was 0.67 and 0.657 that

ID	Description	Adj. P value	Count
Biological process (BP)			
GO:0030198	Extracellular matrix organization	7.24E-22	159
GO:0043062	Extracellular structure organization	7.24E-22	176
GO:0006936	Muscle contraction	2.15E-17	152
GO:0003012	Muscle system process	3.97E-15	179
GO:0034765	Regulation of ion transmembrane transport	4.43E-12	172
GO:0043588	Skin development	2.09E-10	155
GO:0070252	Actin-mediated cell contraction	2.40E-10	61
GO:0030049	Muscle filament sliding	2.71E-10	30
GO:0033275	Actin-myosin filament sliding	2.71E-10	30
GO:0060047	Heart contraction	3.27E-10	113
Cellular component (CC)			
GO:0031012	Extracellular matrix	5.10E-23	204
GO:0043292	Contractile fiber	4.30E-19	109
GO:0062023	Collagen-containing extracellular matrix	5.36E-19	172
GO:0044449	Contractile fiber part	2.87E-18	101
GO:0030016	Myofibril	6.84E-18	102
GO:0030017	Sarcomere	6.09E-17	92
GO:0005788	Endoplasmic reticulum lumen	3.52E-15	133
GO:1902495	Transmembrane transporter complex	6.73E-15	121
GO:1990351	Transporter complex	3.28E-14	121
GO:0016324	Apical plasma membrane	3.97E-13	126
Molecular function (MF)			
GO:0005201	Extracellular matrix structural constituent	6.84E-19	92
GO:0022838	Substrate-specific channel activity	6.84E-19	156
GO:0022803	Passive transmembrane transporter activity	1.25E-18	166
GO:0015267	Channel activity	1.88E-18	165
GO:0030545	Receptor regulator activity	7.47E-18	181
GO:0048018	Receptor ligand activity	7.47E-18	172
GO:0005216	Ion channel activity	4.19E-17	148
GO:0008324	Cation transmembrane transporter activity	4.96E-13	186
GO:0046873	Metal ion transmembrane transporter activity	1.00E-12	140
GO:0022836	Gated channel activity	2.31E-12	117
KEGG pathway			
hsa04080	Neuroactive ligand-receptor interaction	4.27E-11	137
hsa05410	Hypertrophic cardiomyopathy (HCM)	5.36E-10	51
hsa04512	ECM-receptor interaction	1.71E-09	49
hsa05414	Dilated cardiomyopathy (DCM)	1.71E-09	52
hsa04974	Protein digestion and absorption	3.33E-09	51
hsa04060	Cytokine-cytokine receptor interaction	3.36E-08	113
hsa05412	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	6.31E-08	42
hsa00830	Retinol metabolism	5.52E-06	35
hsa05322	Systemic lupus erythematosus	6.51E-06	57
hsa04970	Salivary secretion	1.53E-05	42

Table 1. Top 10 pathways enriched by GO and KEGG analyses.

were calculated by NNE and KM method, respectively, and which indicated the 2 lncRNAs and 2 mRNAs could be effective prognostic biomarkers in predicting OS for HNSCC patients.

To further confirm the metastatic and prognostic roles of the lncRNAs and mRNAs, we analyzed their expression levels between different groups. Firstly, the expression levels between HNSCC samples and normal samples were evaluated, and the results showed AC114730.3, AC136375.3 and LAT were significantly over-expressed in HNSCC patients, whereas RYR3 was downregulated (Fig. 7f). When compared to the low-risk group, all the four RNAs were significantly decreased expression in the high-risk group, demonstrating higher risk scores (Fig. 7g). Furthermore, we evaluated their expression levels when took the clinical stages into account (Fig. 7h). AC136475.3 and LAT obviously over-expressed in stage I, II, III, and IV HNSCC samples compared with normal samples, respectively. AC114730.3 expressed higher only in the stage IV than in normal samples. RYR3 showed lower expression level except stage I. It's surprising that gene expression levels in the high-risk group

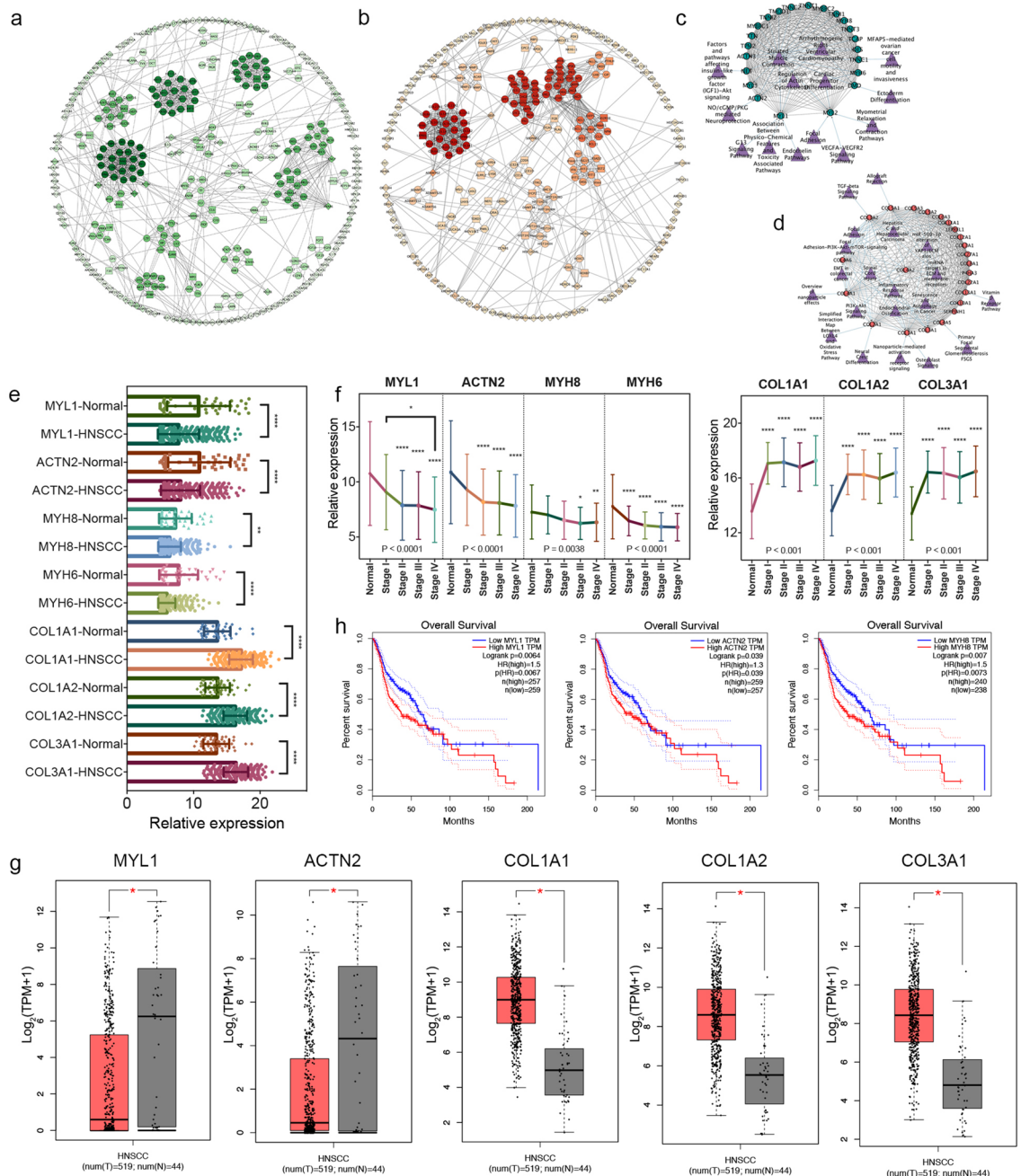


Figure 4. PPI networks and key mRNAs analysis in HNSCC. **(a)** Downregulated PPI network with 343 nodes and 1188 edges. The green color intensity was proportional to cluster scores. **(b)** Upregulated PPI network with 256 nodes and 1032 edges. The red color intensity was proportional to cluster scores. **(c, d)** Pathway-gene networks. Green circles represented downregulated DEMRNAs, and red circles represented upregulated DEMRNAs. Purple triangles represented pathways. **(e)** Bar plots of expression levels of seven key mRNAs between HNSCC samples and normal samples. MYL1, ACTN2, MYH8 and MYH6 were significantly downregulated in HNSCC samples, and COL1A1, COL1A2 and COL3A1 were over-expressed. **(f)** The expressions of seven key mRNAs had significant difference between clinical stages. **(g)** Validation of gene expressions from GEPIA database. **(h)** Survival analysis. Higher expression of MYL1, ACTN2 and MYH8 resulted in worse overall survival of HNSCC patients. *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$; ****: $P < 0.0001$.

were lower than that in the low-risk group at different clinical stages, which suggested some of the patients with clinical early-stage (stage I or II) were supposed to belong to the high-risk group and were likely to have a poor prognosis (Fig. 7i). Moreover, the expressions of the RNAs (except RYR3) between alive- and dead-group were also observed, which were consistent with the above expression patterns (Fig. 7j).

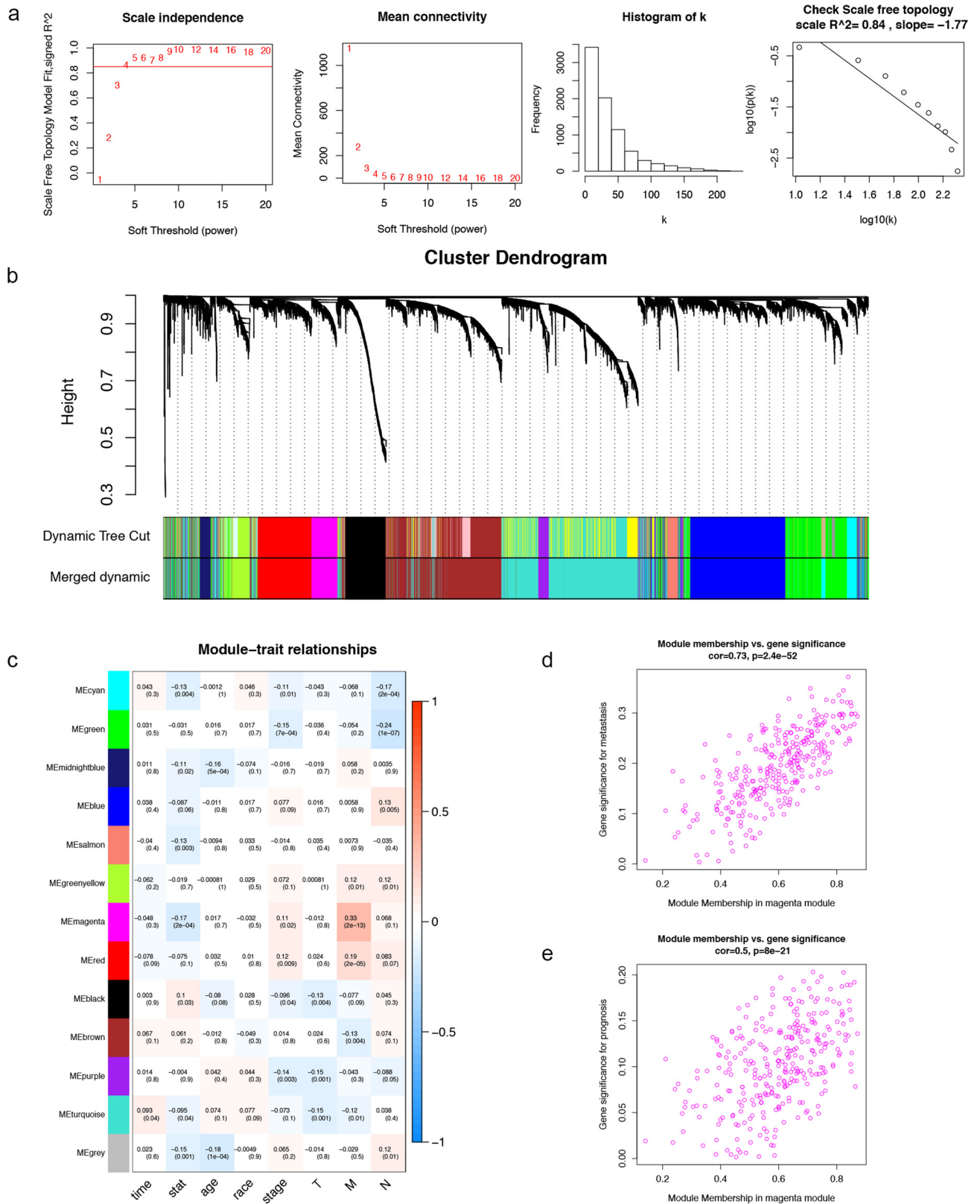


Figure 5. Co-expression network analysis by WGNCA. **(a)** Soft-threshold power analysis for network topology and the test of property of scale-free network. **(b)** Dendrogram of all DEMRNAs clustered based on a dissimilarity measure (1-TOM). **(c)** The relationships between modules and clinical traits. Magenta module was identified to be the most significant association with metastasis and prognosis in HNSCC patients. **(d)** The correlation between the magenta module membership and metastasis. **(e)** The correlation between the magenta module membership and prognosis.

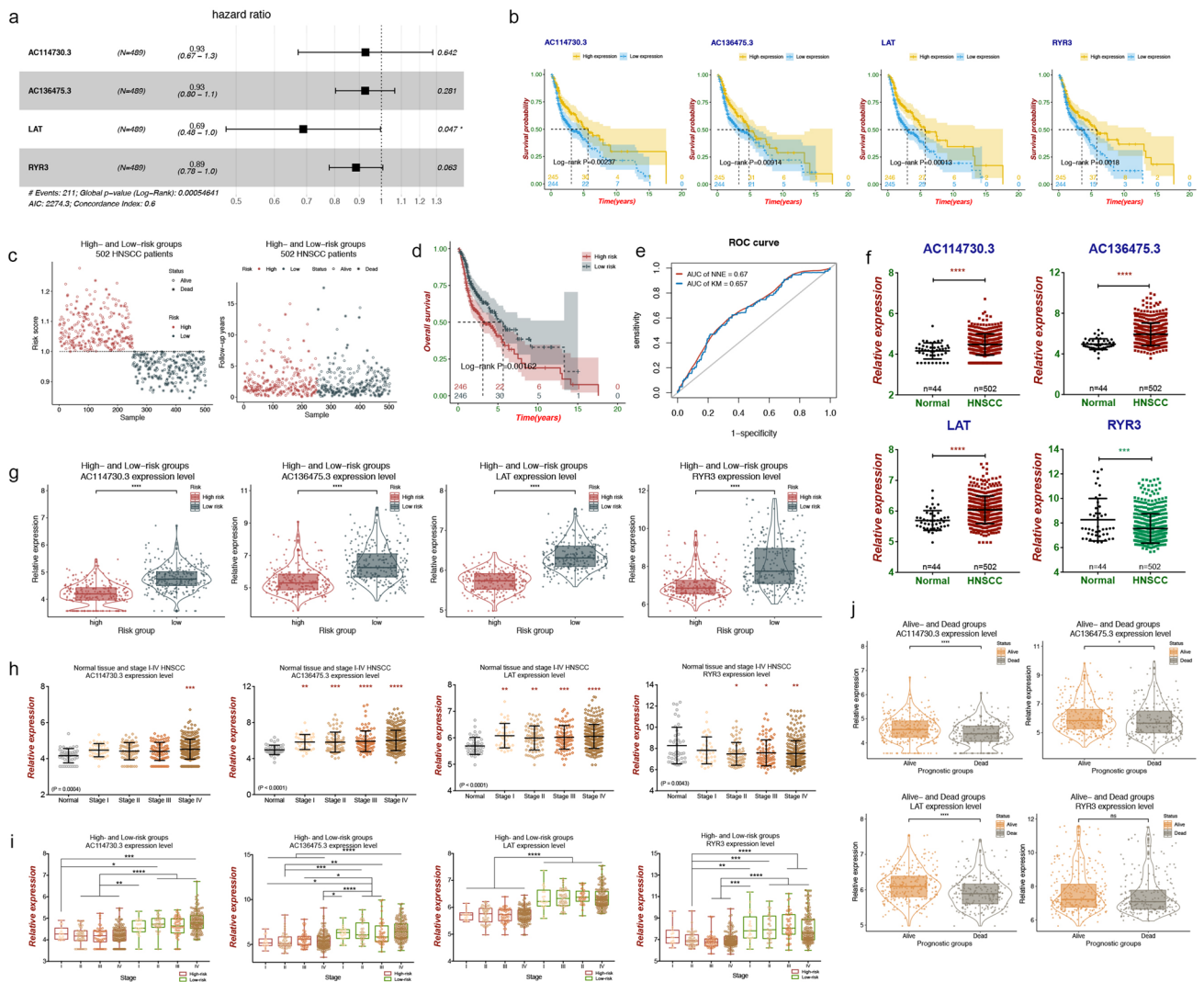


Figure 7. HNSCC-specific prognostic risk model and validation. **(a)** The prognostic model including 2 lncRNAs and 2 mRNAs. **(b)** Kaplan–Meier survival curves for 2 lncRNAs and 2 mRNAs associated with OS of HNSCC patients by Log-rank test. X axis represented overall survival time (years) and Y axis represented survival probability. **(c)** The HNSCC patients were divided into high- and low-risk groups based on the risk scores of HNSCC patients. **(d)** Survival analysis suggested high risk group had worse prognosis in HNSCC patients. **(e)** The ROC analysis showed the AUC of the prognostic model was 0.67 and 0.657 that were calculated by NNE and KM method, respectively. **(f)** Expression patterns of the 2 lncRNAs and 2 mRNAs between HNSCC samples and normal samples. **(g)** Violin plots of expression patterns between high-risk group and low-risk group. **(h)** Comparison of expression levels for 2 lncRNAs and 2 mRNAs between normal samples and stage I, II, III, and IV HNSCC samples. **(i)** Comparison of expression levels for 2 lncRNAs and 2 mRNAs at different clinical stages between high-risk and low-risk groups. **(j)** Violin plots of expression patterns between alive and dead patients. *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$; ****: $P < 0.0001$.

Verification of MYL1, ACTN2 and LAT protein expression levels. To explore the protein expression patterns of the prognosis related genes in HNSCC samples, we examined the expression of MYL1, ACTN2 and LAT in The Human Protein Atlas database (Fig. 8a). RYR3 protein expression level was not evaluated because its information was not available in the database. Besides, cancerous and para-cancerous tissues from oral squamous cell carcinoma (OSCC) patients were collected from The Affiliated Hospital of Stomatology, Zhejiang University School of Medicine. Verification of gene expression levels and protein expression levels were demonstrated in Fig. 8b,c. Consistent with our bioinformatic analysis results, the protein expression patterns demonstrated that MYL1, ACTN2 and LAT were strongly over expressed in OSCC tissues than in normal oral mucosa tissues. IHC staining revealed the cytoplasmic membranous location of the MYL1, ACTN2 and LAT proteins in cancerous tissues. Atypia of cancerous cell was shown in the nests. The boundary between epithelial tissue and connective tissue becomes blurred in cancerous tissues, which was clearly well-structured in the normal oral mucosa tissues. All patients’ information is provided in Tables 3 and 4.

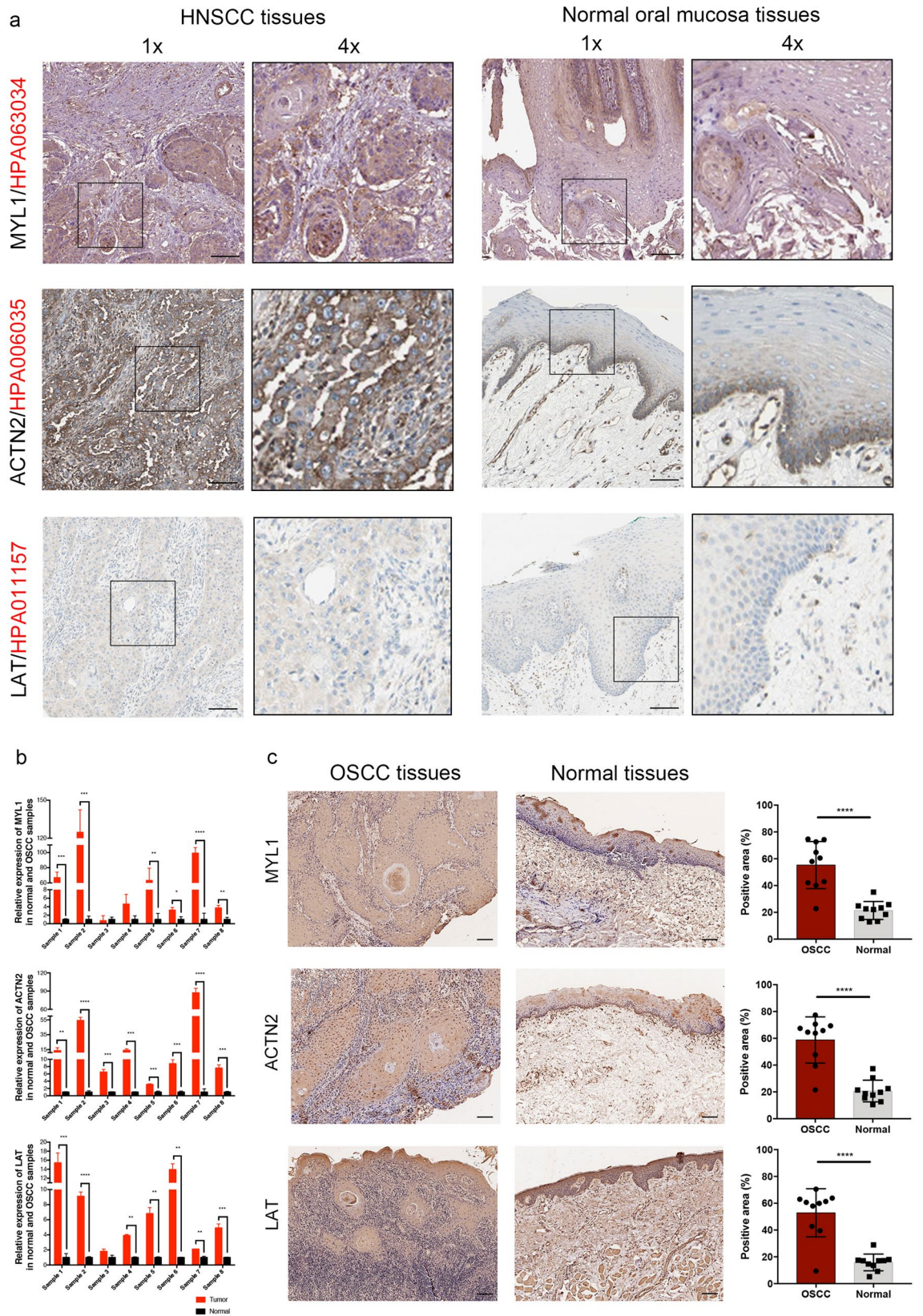


Figure 8. MYL1, ACTN2 and LAT expression validation. (a) Protein expression levels in HNSCC as compared to those in normal tissues by IHC staining from the Human Protein Atlas database. Black square boxes represent typical normal and HNSCC tissues, respectively. (b) Comparison of gene expression level by using qPCR. (c) IHC staining of the proteins in OSCC tissues and normal oral mucosa tissues. Scale bars represent 100 μ m. *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$; ****: $P < 0.0001$.

Gene	Patient ID	Gender	Age	Staining	Intensity	Quantity	Location
Normal oral mucosa tissue							
MYL1	5345	Male	9	Low	Weak	75–25%	Cytoplasmic membranous
ACTN2	1711	Male	71	Low	Moderate	< 25%	Cytoplasmic membranous
LAT	3135	Male	58	Not detected	Negative	None	None
Head and neck squamous cell carcinoma tissue							
MYL1	5286	Male	56	Medium	Moderate	> 75%	Cytoplasmic membranous
ACTN2	2624	Male	87	High	Strong	> 75%	Cytoplasmic membranous
LAT	2608	Male	51	Low	Weak	75–25%	Cytoplasmic membranous

Table 3. IHC staining characteristics of genes from The Human Protein Atlas database. *IHC* immunohistochemistry.

Order	Patient ID	Gender	Age (years)	Location	Tumor size (cm ³)	Differentiation [#]
1	00013232	Male	64	Gum	3.8*2.0*1.8	High
2	00013195	Male	76	Lip	3.0*2.0*0.7	High
3	00013006	Female	64	Tongue	2.5*1.5*0.9	Medium-high
4	00012796	Male	49	Tongue	3.8*1.6*1.0	Medium-high
5	00012792	Male	57	Tongue	1.0*1.0*0.2	Medium-low
6	00012679	Female	45	Tongue	4.5*2.2*1.0	Medium-high
7	00012380	Male	55	Tongue	1.5*1.0*0.8	Medium-high
8	00001858	Male	75	Gum	1.2*1.0*0.5	Medium-low
9	00011717	Female	58	Lip	1.5*1.0*0.9	High
10	00010176	Female	53	Gum	2.7*1.8*0.7	Low

Table 4. Histopathological characteristics of the OSCC patients. *IHC* immunohistochemistry. [#]Diagnosed by two experienced pathologists.

Discussion

HNSCC is a common malignant cancer with high genetic diversities, and per year causes hundreds of thousands of physical dysfunctions and aesthetic problems, even deaths¹. Especially, advanced HNSCC with high metastatic rate usually suggests poor overall survival because conventional therapeutic strategies have achieved unsatisfactory results². In order to promote outcomes of this disease, novel therapeutic targets with higher specificity and efficacy are urgently needed, which are also expected to be potential predicting biomarkers for HNSCC. It's well documented that lncRNAs play chief roles in many biological processes by ceRNA mechanism, such as tumorigenesis and progression. In this study, we systematically analyzed the HNSCC-related RNA-seq and miRNA-seq data from TCGA database by integrated bioinformatics analyses including GO and KEGG pathway analyses, PPI network construction, WGCNA, ceRNA network establishment and survival analysis, and screened out the HNSCC-specific prognostic lncRNA and mRNA signatures to establish a novel ceRNA-based risk model. Finally, four mRNAs (MYL1, ACTN2, LAT and RYR3) and two lncRNAs (AC114730.3 and AC136375.3) were identified to associate with OS in HNSCC patients.

MYL1 (myosin light chain 1) encodes an alkali light chain for myosin that is a hexameric ATPase cellular motor protein. It was well studied that MYL1 played a key role in congenital myopathy and knockdown of the gene could cause severe muscle-related disorders in zebrafish²⁹. Recently, MYL1 has been identified as a potential biomarker in various cancers^{30,31}. Sajnani et al. found that MYL1 was experimentally proved to be downregulated in buccal cancer and be involved in the "actin mediated cell contraction" biological progress³¹, which was consistent with our results. ACTN2 (actin alpha 2) is a member of the spectrin gene superfamily that includes varying groups of cytoskeletal proteins. In breast cancer patients, mutated ACTN2 was found to be related to invasive ductal carcinoma and suggested a worse OS than ductal carcinoma in situ³². Sun et al. and his colleagues reported ACTN2 was one of the hub genes selected by bioinformatics methods in PTEN mutation prostate cancer³³. However, so far researches on the associations between aberrantly expressed ACTN2 and tumorigenesis and progression of HNSCC are barely reported. The current study firstly indicated negative MYL1 and ACTN2 expression contributed to a better OS of HNSCC patients. Perhaps the low expression of the genes was to inhibit some activities in the tumor cells, so patients with lower expression of these genes showed better prognosis.

With the development of bioinformatics analysis methodology, the single method to find a limited group of genes as therapeutic targets has been less convincing, so the comprehensive bioinformatics analyses are more recognized by researchers. The ceRNA theory points out that lncRNAs could function as ceRNAs and competitively bind to miRNAs by acting as "sponges" of miRNAs, which might derepress activity or expression of those miRNA-mediated target with similar binding sites³⁴. This analysis may shed light on regulatory networks that have been underestimated by conventional protein-coding studies. Heretofore, some studies have investigated

ceRNAs in HNSCC and identified several RNAs associated with OS of HNSCC patients^{35,36}. For instance, Gao et al. comprehensively investigated TCGA data to find HOXC13-AS was highly expressed in HNSCC, and further confirmed knockdown of HOXC13-AS impaired proliferation, migration and invasion of nasopharyngeal carcinoma cell via targeting miR-383-3p/HMGA2 axis³⁵. Lately, Cai et al. revealed the prognostic roles of miRNA-204/211 in HNSCC by integrally analyzing sequencing data from TCGA database to construct ceRNA networks³⁶. These studies provide a better understanding how the ceRNA contributes to improving the diagnostic and prognostic efficiency of HNSCC patients. However, rare studies have focused on the co-expression among ceRNAs, let alone establishing risk models. In the present study, we innovatively used WGCNA to identify co-expressed modules in HNSCC, and further constructed lncRNA-miRNA-mRNA ceRNA network based on the co-expressed genes in the most significant module. Moreover, LR test and univariate Cox regression analysis selected 12 RNAs (including 6 lncRNAs and 6 mRNAs) that were simultaneously having significant prognostic values in HNSCC patients. LAT, RYR3, AC114730.3 and AC136375.3 were anteriorly implemented multivariate Cox regression analysis to build a HNSCC-specific prognostic model, since their stronger correlations to OS of HNSCC.

The protein encoded by LAT (linker for activation of T cells) can recruit multiple downstream molecules after phosphorylation and activation, and is involved in immunology related pathways³⁷. T cell is an important component of acquired immunity for hosts, which is able to be activated by phosphorylated LAT through TCR signaling pathway³⁸. Wang et al. elucidated prognostic risk models containing 3 genes by comprehensively analyzing methylation data and RNA-seq data for clear cell renal cell carcinoma (ccRCC) from TCGA database, and concluded that hypomethylation and over-expression of LAT resulted in poor OS of ccRCC³⁹. Nevertheless, no studies so far have reported any correlation between LAT and HNSCC. In our study, we found that LAT was highly expressed in HNSCC patients and correlated with tumor metastasis and prognosis, and those patients with higher expression of LAT surprisingly presented lower risk scores and better survival. According to our findings and previous literatures, we speculate that LAT, as a protective gene, takes a good effect on preventing tumor metastasis and deterioration in HNSCC. On the contrary, RYR3 (ryanodine receptor 3), one of the isoforms of RYRs, encodes protein to release calcium from intracellular storage⁴⁰. The present study demonstrated that RYR3 was apparently downregulated in HNSCC tissues when compared to normal samples, and lower expression of RYR3 always led to higher risk scores and worse OS in HNSCC patients. However, none of the previous studies has been elucidated its biological role in HNSCC. Schmitt K et al. held that reduced expressed RYR2, another isoform of RYRs, might serve as risk factor for unfavorable prognosis and upcoming malignant conversion in head and neck cancer⁴¹. Using siRNA or miR-367 to knockdown the endogenous RYR3 apparently inhibited growth and migration of breast cancer cells, which led cell-cell contacts became more weakened because of their rounder morphology⁴². Therefore, tumor suppressor role of RYR3 on HNSCC still need to be further clarified.

Between the identified 2 lncRNAs in the risk model, AC114730.3 and AC136375.3 were remarkably over-expressed in HNSCC tissues, and HNSCC patients with higher expressions presented lower risk scores and better survival, which were similar to LAT expression pattern. Although no studies so far have reported any correlation between the 2 lncRNAs and cancer, our study firstly detected the upregulated AC114730.3 in ceRNA network competed with downregulated miR-375 and miR-338-3p to indirectly regulate downstream mRNAs. Dependent upon various cancer types, miR-375 tended to have dramatically suppressive or oncogenic effect on cancers⁴³⁻⁴⁵. MiR-375 could inhibit the growth and metastasis of ovarian cancer cells, of which the suppressive effect was reversed by MLK7-AS1 lncRNA that indirectly upregulated YAP1 expression⁴³. The MLK7-AS1 regulatory role towards miR-375 was observed in gastric cancer as well⁴⁴. However, lower miR-375 expression level was evaluated in malignant basal-like breast cancer than in luminal-like breast cancer, suggesting malignant development of tumor⁴⁵. The current study found AC114730.3 was aberrantly expressed in HNSCC and functioned as a novel regulatory lncRNA of miR-375, which triggered our speculation that over-expressed AC114730.3 inhibited miR-375 expression, but might as similar to breast cancer, the miR-375 would tend to act as a tumorigenesis molecular in HNSCC. Besides, among the potential targeting mRNAs for miR-375, KCNAB3 was surprisingly upregulated in HNSCC, composing AC114730.3/miR-375/KCNAB3 regulatory axis. Thus, it is reasonable to understand why highly expressed AC114730.3 contributed to better OS in HNSCC patients. In addition, miR-338-3p was predicted to be another targeting miRNA of AC114730.3 and no researches have so far focused on its effect on HNSCC patients. Not coincidentally, such a miRNA deserves our attention since its reported association with non-small cell lung cancer (NSCLC)⁴⁶, cervical cancer⁴⁷ and glioma⁴⁸. Yang et al. found over-expressed LINC00525 suggested poor prognosis and could modulate miR-338-3p that endogenously targeted IRS2 in NSCLC⁴⁶. Two circular RNA, HIPK3 and SMO, were reported to promote tumor proliferation and metastasis through sponging miR-338-3p in cervical cancer⁴⁷ and glioma⁴⁸, respectively. Based on our findings and published literatures, however, we are weak in understanding and explaining the mechanisms of AC114730.3 and miR-338-3p about their prognostic effect on HNSCC. In brief, findings in the current study indicate that MYL1, ACTN2, LAT, AC114730.3 and AC136375.3 may play roles as HNSCC suppressors, while RYR3 may function as a HNSCC promoter.

Although ceRNA interactions have been described by multiple studies, the molecular conditions for optimal activity of ceRNA remain ambiguous¹². The ceRNAs were required to compete with the entire target pool of miRNAs to bind targets. In order to win the "one versus multiple" battle, the abundance of ceRNA must to be increased to an aberrantly high level. A previous study has revealed up to 23-fold increase of ALDOA abundance did not lead to detectable miR-122 inhibition in hepatocytes⁴⁹. Except for the abundance of ceRNA, other factors might also have profound effects on ceRNA crosstalk, including localization of ceRNA¹³, binding affinity of the shared MREs¹³, secondary structure of RNA⁵⁰, RNA editing⁵¹ and so on. Therefore, current ceRNA research is still in its infancy so that one of the exciting avenues for future work is to uncover the mystery of their widespread crosstalk. Some limitations must be noted despite the construction the risk model in clinical implication. Firstly, the ceRNA-based risk model must be verified by experimental approaches to further elucidate the precise

molecular mechanisms underlying HNSCC. Secondly, the predictive and therapeutic efficiency of HNSCC-specific prognostic signatures needs to be evaluated by large-scale studies.

In conclusion, four mRNAs and two lncRNAs were identified as HNSCC-specific prognostic signatures for HNSCC patients. Moreover, gene functions, co-expressed modules and ceRNA-based prognostic risk model were also elucidated, which facilitated the expansion of the current study on the roles of ceRNAs in the tumorigenesis, development and treatment of head and neck squamous cell carcinoma. Further experimental studies are required to biologically validate these findings.

Materials and methods

Data acquisition. HNSCC-related RNA-sequencing (RNA-seq) data and miRNA-sequencing (miRNA-seq) data that were derived from the IlluminaHiSeq_RNASeq and IlluminaHiSeq_miRNASeq sequencing platforms, were retrieved from TCGA database (<https://www.cancer.gov/tcga>), respectively. The RNA and miRNA expression profiles (level 3) were free to downloaded from the National Cancer Institute's Genomic Data Commons data portal (<https://portal.gdc.cancer.gov>). The RNA-seq data included 502 HNSCC samples and 44 normal samples, while the miRNA-seq data included 525 HNSCC samples and 44 normal samples. Among them, 44 normal samples of RNA-seq and miRNA-seq data were from normal tissues nearby tumorous tissues of 44 HNSCC patients. We also simultaneously downloaded detailed clinical and follow-up information of all HNSCC patients from TCGA database. Patients who met the following inclusion criteria were involved in TCGA-HNSCC cohort: (1) histologically diagnosed HNSCC; (2) no other malignancy except HNSCC; (3) patients with detailed clinical and follow-up information.

RNA-seq and miRNA-seq data preprocessing and differentially expressed gene analysis (DEGA). Subsequently, mRNAs and lncRNAs were annotated based on GENCODE Release 35 (GRCh38.p13) (<https://www.genecodegenes.org/human/>) and miRNAs were encoded according to miRbase v22 (<http://www.mirbase.org/index.shtml#opennewwindow>). Level 3 RNA-seq and miRNA-seq count data were normalized by size factor⁵² and transformed via variance-stabilizing transformation (VST) implemented using "DESeq2" package⁵³. This transformation makes the expression values homoscedastic by fitting the dispersion to a negative binomial distribution. "DESeq2" package in Bioconductor project (version 3.10, <http://www.biocductor.org>) was utilized to perform DEGA. Then, we screened out DElncRNAs, DEMRNAs and DEMiRNAs between HNSCC samples and normal samples. $|\log_2\text{FC}| \geq 1$ and adjusted P value < 0.05 were set as cut-off criteria. The volcano plots of lncRNA, mRNA and miRNA expression were constructed using the "ggplot2" package and the heat maps of DElncRNAs, DEMRNAs and DEMiRNAs were plotted using the "pheatmap" package in R software (version 3.6.0)⁵⁴.

Functional enrichment analyses of DEMRNAs. We next implemented functional enrichment analyses so as to shed light on biological function of DEMRNAs. To do this, GO and KEGG pathway⁵⁵ analyses of all DEMRNAs was conducted using the "clusterProfiler" package⁵⁶ in R software. For this analysis, the statistical cutoff threshold was set at adjusted P value < 0.05 . GO analysis included three categories: BP, CC, and MF.

PPI network construction and analysis of DEMRNAs. In accordance with expression levels, DEMRNAs were separated into two groups and then separately used to build PPI networks. DEMRNAs with cut-off of $|\log_2\text{FC}| \geq 2$ and adjusted P value < 0.05 were uploaded to the Search Tool for the Retrieval of Interacting Genes (STRING, version 11, <https://string-db.org>) database. The gene pairs with a combined score ≥ 0.9 (the highest confidence) were retrieved from the results and visualized in Cytoscape software (version 3.7.0). Next, the highly correlated module was screened out from whole PPI network using the MCODE plugin in Cytoscape software to perform Molecular Complex Detection algorithm. The most significantly correlated module of each group was selected for further investigation. Then, we performed functional pathway enrichment analysis among clustered DEMRNAs in the most significantly correlated module based on the "Wikipathways" database using CyTargetLinker plugin in Cytoscape software.

Identification of key DEMRNAs and validation. Another plugin in Cytoscape software, CytoNCA, was used to perform Centrality analyses that included Subgraph Centrality, Degree Centrality, Eigenvector Centrality, Information Centrality, Betweenness Centrality, Closeness Centrality and Network Centrality⁵⁷. Key DEMRNAs in each PPI network were identified by Centrality analyses. We selected the DEMRNAs with top 1% Centrality scores as key DEMRNAs. Next, we calculated expression levels of key DEMRNAs between HNSCC samples and normal samples, and further evaluated the relationship between HNSCC samples and normal samples in different stages. Moreover, GEPIA (<http://gepia.cancer-pku.cn>) database has been previously and publicly adopted to analyze gene expression levels of normal and tumorous samples from TCGA. Thus, we also performed the expression analysis between normal and HNSCC samples using $\log_2(\text{TPM} + 1)$ in GEPIA database. Besides, survival analysis of key DEMRNAs was implemented using Log-Rank test. The P value, hazard ratio and the 95% confidence interval were calculated as well.

Weighted gene co-expression network analysis. In order to construct the co-expression network and identify co-expression gene modules, we performed WGCNA among lncRNAs and mRNAs retrieved from the TCGA RNA-seq data, respectively. After removing normal samples and HNSCC samples without clinical information, the expression matrix was firstly assembled with the gene symbols as column and HNSCC samples as the row names. Following this, "WGCNA" package in R software was adopted to perform WGCNA⁵⁸.

The expression matrix was converted into an adjacency matrix that was further used to build an unsupervised co-expression relationship by using Pearson's correlation coefficients for all pair-wise genes. Based on scale-free topology criterion, we selected the best power ($\beta = 4$) as a soft-thresholding parameter to strengthen correlation adjacency matrix⁵⁹. Secondly, the adjacency matrix was transformed into a topological matrix, and TOM was used to minimize interference from noise and spurious associations⁶⁰. Any genes exhibiting identical expression profiles were grouped together into gene modules using hierarchical clustering. We set minimum module size as 30 genes. Finally, gene modules were identified from the system cluster tree by dynamic cutting algorithm and those with 70% similarity were merged into a module. Module eigengenes (MEs) were considered as the major component for each module, and together with the clinical information, they were used for MTR analysis. The MTR heat map was plotted to clearly visualize clinical trait related modules. Module significance (MS) was defined as the mean GS that was calculated based on the samples' clinical traits. Finally, the genes in the most significant module with strongest correlation to HNSCC clinical features were selected for further analysis.

Construction of ceRNA networks. According to the ceRNA theory, we constructed lncRNA-miRNA-mRNA ceRNA networks with the following steps: (1) DE miRNAs, co-expressed DE lncRNAs and DE mRNAs in the most significant module were kept; (2) the interaction relationships between DE lncRNAs and DE miRNAs were predicted by miRcode (<http://www.mircode.org>); (3) the interaction relationships between DE miRNAs and DE mRNAs were predicted by TargetScan (<http://www.targetscan.org>), miRTarBase (<http://mirtarbase.mbc.ntcu.edu.tw>) and miRBase (<http://www.mirbase.org>); (4) overlapped DE miRNAs which were negatively interacted with DE lncRNAs and DE mRNAs were selected to construct ceRNA networks. Then, the networks were visualized in Cytoscape software.

Identification of HNSCC-specific prognostic lncRNA and mRNA signatures. In order to identify specific prognostic genes for HNSCC, we firstly performed KM curve analysis and LR test of DE lncRNAs and DE mRNAs in ceRNA networks using "survival" package in R software. HNSCC patients were divided into high- and low-expression groups according to gene expression levels and LR P value < 0.05 was defined as the significant prognostic standard for lncRNAs and mRNAs. Then, univariate Cox proportional hazards regression analysis was implemented to evaluate the prognostic association of DE lncRNAs and DE mRNAs in ceRNA networks in HNSCC patients by using "survival" package in R software as well. Statistical P value < 0.05 was set as the cut-off criteria. Next, multivariate Cox hazards regression model was constructed using overlapped DE lncRNAs and DE mRNAs in the above two methods with "survival" package in R software, which were strongly significantly associated with OS of HNSCC patients. The multivariate Cox hazards regression model was built to evaluate the HNSCC-specific prognostic lncRNA and mRNA signatures. The model was established based on the following equation:

$$\text{Risk score} = \sum (\text{Coe} * \text{Exp}_{\text{lncRNA/mRNA}})$$

where "Coe" represented the regression coefficient of the lncRNA or mRNA retrieved from the multivariate Cox regression model and "Exp" referred to the expression level of the lncRNA or mRNA. Subsequently, HNSCC patients were divided into high- and low-risk groups based on the median risk score for all patients. A receiver operating characteristic (ROC) curve was plotted and AUC was calculated to access the risk prediction rate of risk model adopting "survivalROC" package in R software.

Patients and tissue sample collection. Human OSCC samples were collected at The Affiliated Hospital of Stomatology, Zhejiang University School of Medicine between January 2018 and December 2020, which were applied to validate the expression difference of the genes. There were 6 males and 4 females (average age, 59.6 years old; range 45–76 years old). Table 4 shown the details. The tissue samples were immediately fixed in 4% paraformaldehyde, embedded in paraffin, sectioned, and processed for the following steps. Each experimental sample was separated into two parts, one was used for IHC analysis and one for total RNA extraction. Research was authorized by the Ethics Committee of The Affiliated Hospital of Stomatology, Zhejiang University School of Medicine. Informed consent was obtained from all participating patients. All experiments were carried out in accordance with the approved study and relevant guidelines.

qPCR validation. Total RNA of tissue samples was extracted by TRIzol reagent (Invitrogen, USA) with following the manufacturer's instructions, quantified by applying NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, Inc., USA), and synthesized into cDNA by using RNeasy Mini Kit (Takara, Japan). qPCR analysis was performed through utilizing SYBR Green Master Mix (Takara, Kyoto, Japan) under the following conditions: 95 °C for 5 min; 45 cycles of 95 °C for 5 s and 60 °C for 30 s; one cycle of 95 °C for 5 s, 60 °C for 1 min and 95 °C for 15 s; and finally, 50 °C for 30 s. The primer sequences were listed in Supplementary Table 1. Relative expression levels of targeted mRNAs were measured using β -actin as the internal control and analyzed using the $2^{-\Delta\Delta C_q}$ method.

Immunohistochemistry analysis. The Human Protein Atlas (<https://www.proteinatlas.org/>) database was preliminarily employed to map tissue protein expression pattern for the identified genes⁶¹. Subsequently, paraffin embedded sections were deparaffinized in xylene and rehydrated in graded ethanol. After washing with distilled water, the sections were heated in 37 °C with pepsin for antigen recovery, then washed and incubated in 3% (v/v) hydrogen peroxide for 25 min to block endogenous peroxidase activity, and finally incubated in 3% (w/v) BSA. Sections were incubated with primary antibodies against MYL1 (1:100; SAB, 47440), ACTN2

(1:250; Proteintech, 14221-1-AP) and LAT (1:100; Proteintech, 11326-1-AP) overnight. Then, washing in PBS, the sections were incubated with horseradish peroxidase (HRP)-labeled secondary antibodies for 1 h at room temperature. Diaminobenzidine (DAB) was used for visualizing immunolabeling.

Statistical analysis. The expression levels of lncRNAs and mRNAs in normal samples and HNSCC samples were firstly evaluated, and the association between expression levels and clinical stages were subsequently accessed. Besides, we also compared the genes' expression levels in high-risk group and low-risk group, and performed survival analysis by the LR test, of which the result was visualized by plotting KM curve. According to prognostic information of HNSCC patients, the comparison of expression in alive group and dead group was also analyzed. qPCR results were presented by using GraphPad Prism (version 7.0; GraphPad Software, Inc., USA). Student's t-test was used to compare two groups, and one-way ANOVA for more groups. All values are presented as the mean \pm standard deviation and $P < 0.05$ was considered to indicate a statistically significant difference.

Data availability

The HNSCC RNA-seq and miRNA-seq data were deposited in the TCGA database. Besides, please contact author for data and material requests.

Received: 21 August 2020; Accepted: 9 March 2021

Published online: 18 March 2021

References

- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA Cancer J. Clin.* **69**, 7–34. <https://doi.org/10.3322/caac.21551> (2019).
- Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424. <https://doi.org/10.3322/caac.21492> (2018).
- Chow, L. Q. M. Head and neck cancer. *N. Engl. J. Med.* **382**, 60–72. <https://doi.org/10.1056/NEJMr1715715> (2020).
- Guidi, A., Codeca, C. & Ferrari, D. Chemotherapy and immunotherapy for recurrent and metastatic head and neck cancer: a systematic review. *Med. Oncol.* **35**, 37. <https://doi.org/10.1007/s12032-018-1096-5> (2018).
- Gan, S. J. *et al.* Incidence and pattern of second primary malignancies in patients with index oropharyngeal cancers versus index nonoropharyngeal head and neck cancers. *Cancer* **119**, 2593–2601. <https://doi.org/10.1002/ncr.28107> (2013).
- Duprez, F. *et al.* Distant metastases in head and neck cancer. *Head Neck* **39**, 1733–1743. <https://doi.org/10.1002/hed.24687> (2017).
- Leemans, C. R., Snijders, P. J. F. & Brakenhoff, R. H. The molecular landscape of head and neck cancer. *Nat. Rev. Cancer* **18**, 269–282. <https://doi.org/10.1038/nrc.2018.11> (2018).
- Pavon, M. A. *et al.* uPA/uPAR and SERPINE1 in head and neck cancer: role in tumor resistance, metastasis, prognosis and therapy. *Oncotarget* **7**, 57351–57366. <https://doi.org/10.18632/oncotarget.10344> (2016).
- Hersi, H. M., Raulf, N., Gaken, J., Folarin, N. & Tavassoli, M. MicroRNA-9 inhibits growth and invasion of head and neck cancer cells and is a predictive biomarker of response to plerixafor, an inhibitor of its target CXCR4. *Mol. Oncol.* **12**, 2023–2041. <https://doi.org/10.1002/1878-0261.12352> (2018).
- Sanchez-Mejias, A. & Tay, Y. Competing endogenous RNA networks: tying the essential knots for cancer biology and therapeutics. *J. Hematol. Oncol.* **8**, 30. <https://doi.org/10.1186/s13045-015-0129-1> (2015).
- Qu, J., Li, M., Zhong, W. & Hu, C. Competing endogenous RNA in cancer: a new pattern of gene expression regulation. *Int. J. Clin. Exp. Med.* **8**, 17110–17116 (2015).
- Tay, Y., Rinn, J. & Pandolfi, P. P. The multilayered complexity of ceRNA crosstalk and competition. *Nature* **505**, 344–352. <https://doi.org/10.1038/nature12986> (2014).
- Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?. *Cell* **146**, 353–358. <https://doi.org/10.1016/j.cell.2011.07.014> (2011).
- Chen, X. *et al.* Long noncoding RNA LINC01234 functions as a competing endogenous RNA to regulate CBFβ expression by sponging miR-204-5p in gastric cancer. *Clin. Cancer Res.* **24**, 2002–2014. <https://doi.org/10.1158/1078-0432.Ccr-17-2376> (2018).
- Fan, C. N., Ma, L. & Liu, N. Systematic analysis of lncRNA-miRNA-mRNA competing endogenous RNA network identifies four-lncRNA signature as a prognostic biomarker for breast cancer. *J. Transl. Med.* **16**, 264. <https://doi.org/10.1186/s12967-018-1640-2> (2018).
- Li, H. *et al.* Long noncoding RNA NORAD, a novel competing endogenous RNA, enhances the hypoxia-induced epithelial-mesenchymal transition to promote metastasis in pancreatic cancer. *Mol. Cancer* **16**, 169. <https://doi.org/10.1186/s12943-017-0738-0> (2017).
- Zhu, N. *et al.* Integrated analysis of a competing endogenous RNA network reveals key lncRNAs as potential prognostic biomarkers for human bladder cancer. *Medicine (Baltimore)* **97**, e11887. <https://doi.org/10.1097/md.00000000000011887> (2018).
- Fang, X. N. *et al.* Comprehensive analysis of competitive endogenous RNAs network associated with head and neck squamous cell carcinoma. *Sci. Rep.* **8**, 10544. <https://doi.org/10.1038/s41598-018-28957-y> (2018).
- Pan, Y., Liu, G., Wang, D. & Li, Y. Analysis of lncRNA-mediated ceRNA crosstalk and identification of prognostic signature in head and neck squamous cell carcinoma. *Front. Pharmacol.* **10**, 150. <https://doi.org/10.3389/fphar.2019.00150> (2019).
- Zhang, H. *et al.* Co-expression network analysis identified gene signatures in osteosarcoma as a predictive tool for lung metastasis and survival. *J. Cancer* **10**, 3706–3716. <https://doi.org/10.7150/jca.32092> (2019).
- Liu, B. *et al.* Analysis of gene coexpression network reveals prognostic significance of CNFN in patients with head and neck cancer. *Oncol. Rep.* **41**, 2168–2180. <https://doi.org/10.3892/or.2019.7019> (2019).
- Jin, Y. & Yang, Y. Identification and analysis of genes associated with head and neck squamous cell carcinoma by integrated bioinformatics methods. *Mol. Genet. Genomic Med.* **7**, e857. <https://doi.org/10.1002/mgg3.857> (2019).
- Chien, M. H., Lin, C. W., Cheng, C. W., Wen, Y. C. & Yang, S. F. Matrix metalloproteinase-2 as a target for head and neck cancer therapy. *Expert Opin. Ther. Targets* **17**, 203–216. <https://doi.org/10.1517/14728222.2013.740012> (2013).
- Dean, T., Li, N. T., Cadavid, J. L., Ailles, L. & McGuigan, A. P. A TRACER culture invasion assay to probe the impact of cancer associated fibroblasts on head and neck squamous cell carcinoma cell invasiveness. *Biomater. Sci.* **8**, 3078–3094. <https://doi.org/10.1039/c9bm02017a> (2020).
- Chaffer, C. L. & Weinberg, R. A. A perspective on cancer cell metastasis. *Science* **331**, 1559–1564. <https://doi.org/10.1126/science.1203543> (2011).
- Smith, S. S. Female sex steroid hormones: from receptors to networks to performance—actions on the sensorimotor system. *Prog. Neurobiol.* **44**, 55–86. [https://doi.org/10.1016/0301-0082\(94\)90057-4](https://doi.org/10.1016/0301-0082(94)90057-4) (1994).

27. Watanabe, M. *et al.* Gamma-aminobutyric acid (GABA) and cell proliferation: focus on cancer cells. *Histol. Histopathol.* **21**, 1135–1141. <https://doi.org/10.14670/HH-21.1135> (2006).
28. Olsen, R. W., Sieghart, W. & International Union of Pharmacology. LXX. Subtypes of gamma-aminobutyric acid(A) receptors: classification on the basis of subunit composition, pharmacology, and function. Update. *Pharmacol. Rev.* **60**, 243–260. <https://doi.org/10.1124/pr.108.00505> (2008).
29. Ravenscroft, G. *et al.* Bi-allelic mutations in MYL1 cause a severe congenital myopathy. *Hum. Mol. Genet.* **27**, 4263–4272. <https://doi.org/10.1093/hmg/ddy320> (2018).
30. Ohmori, H. *et al.* Evaluation of parameters for cancer-induced sarcopenia in patients autopsied after death from colorectal cancer. *Pathobiology* **86**, 306–314. <https://doi.org/10.1159/000503037> (2019).
31. Sajjani, M. R. *et al.* Identification of novel transcripts deregulated in buccal cancer by RNA-seq. *Gene* **507**, 152–158. <https://doi.org/10.1016/j.gene.2012.07.036> (2012).
32. Yang, M. *et al.* A breast one-patient panel of heterogeneous genomes reveals genetic alterations driving DCIS into invasive lesions. *Future Oncol.* **15**, 1565–1576 (2019).
33. Sun, J., Li, S., Wang, F., Fan, C. & Wang, J. Identification of key pathways and genes in PTEN mutation prostate cancer by bioinformatics analysis. *BMC Med. Genet.* **20**, 191. <https://doi.org/10.1186/s12881-019-0923-7> (2019).
34. Wang, P. *et al.* miRSponge: a manually curated database for experimentally supported miRNA sponges and ceRNAs. *Database (Oxford)* **2015**, bav098. <https://doi.org/10.1093/database/bav098> (2015).
35. Gao, C., Lu, W., Lou, W., Wang, L. & Xu, Q. Long noncoding RNA HOXC13-AS positively affects cell proliferation and invasion in nasopharyngeal carcinoma via modulating miR-383-3p/HMGA2 axis. *J. Cell Physiol.* **234**, 12809–12820. <https://doi.org/10.1002/jcp.27915> (2019).
36. Cai, J. *et al.* Exploring the role of Mir204/211 in HNSCC by the combination of bioinformatic analysis of ceRNA and transcription factor regulation. *Oral Oncol.* **96**, 153–160. <https://doi.org/10.1016/j.oraloncology.2019.07.024> (2019).
37. Balagopalan, L., Kortum, R. L., Coussens, N. P., Barr, V. A. & Samelson, L. E. The linker for activation of T cells (LAT) signaling hub: from signaling complexes to microclusters. *J. Biol. Chem.* **290**, 26422–26429. <https://doi.org/10.1074/jbc.R115.665869> (2015).
38. Balagopalan, L. *et al.* Plasma membrane LAT activation precedes vesicular recruitment defining two phases of early T-cell activation. *Nat. Commun.* **9**, 2013. <https://doi.org/10.1038/s41467-018-04419-x> (2018).
39. Wang, J. *et al.* LAT, HOXD3 and NFE2L3 identified as novel DNA methylation-driven genes and prognostic markers in human clear cell renal cell carcinoma by integrative bioinformatics approaches. *J. Cancer* **10**, 6726–6737. <https://doi.org/10.7150/jca.35641> (2019).
40. Meissner, G. The structural basis of ryanodine receptor ion channel function. *J. Gen. Physiol.* **149**, 1065–1089. <https://doi.org/10.1085/jgp.201711878> (2017).
41. Schmitt, K. *et al.* Somatic mutations and promotor methylation of the ryanodine receptor 2 is a common event in the pathogenesis of head and neck cancer. *Int. J. Cancer* **145**, 3299–3310. <https://doi.org/10.1002/ijc.32481> (2019).
42. Zhang, L. *et al.* Functional SNP in the microRNA-367 binding site in the 3' UTR of the calcium channel ryanodine receptor gene 3 (RYR3) affects breast cancer risk and calcification. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13653–13658. <https://doi.org/10.1073/pnas.1103360108> (2011).
43. Yan, H. *et al.* Long noncoding RNA MLK7-AS1 promotes ovarian cancer cells progression by modulating miR-375/YAP1 axis. *J. Exp. Clin. Cancer Res.* **37**, 237. <https://doi.org/10.1186/s13046-018-0910-4> (2018).
44. Quan, Y. *et al.* Knockdown of long non-coding RNA MAP3K20 antisense RNA 1 inhibits gastric cancer growth through epigenetically regulating miR-375. *Biochem. Biophys. Res. Commun.* **497**, 527–534. <https://doi.org/10.1016/j.bbrc.2018.02.072> (2018).
45. Liu, J. *et al.* An integrative bioinformatics analysis identified miR-375 as a candidate key regulator of malignant breast cancer. *J. Appl. Genet.* **60**, 335–346. <https://doi.org/10.1007/s13353-019-00507-w> (2019).
46. Yang, Z. *et al.* Long non-coding RNA LINC00525 promotes the non-small cell lung cancer progression by targeting miR-338-3p/IRS2 axis. *Biomed. Pharmacother.* **124**, 109858. <https://doi.org/10.1016/j.biopha.2020.109858> (2020).
47. Qian, W., Huang, T. & Feng, W. Circular RNA HIPK3 promotes EMT of cervical cancer through sponging miR-338-3p to up-regulate HIF-1 α . *Cancer Manag. Res.* **12**, 177–187. <https://doi.org/10.2147/CMAR.S232235> (2020).
48. Xiong, Z. *et al.* Circular RNA SMO sponges miR-338-3p to promote the growth of glioma by enhancing the expression of SMO. *Aging (Albany NY)* **11**, 12345–12360. <https://doi.org/10.18632/aging.102576> (2019).
49. Denzler, R., Agarwal, V., Stefano, J., Bartel, D. P. & Stoffel, M. Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Mol. Cell* **54**, 766–776. <https://doi.org/10.1016/j.molcel.2014.03.045> (2014).
50. Srikantan, S., Tominaga, K. & Gorospe, M. Functional interplay between RNA-binding protein HuR and microRNAs. *Curr. Protein Pept. Sci.* **13**, 372–379. <https://doi.org/10.2174/138920312801619394> (2012).
51. Maas, S. Posttranscriptional recoding by RNA editing. *Adv. Protein Chem. Struct. Biol.* **86**, 193–224. <https://doi.org/10.1016/B978-0-12-386497-0.00006-2> (2012).
52. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106. <https://doi.org/10.1186/gb-2010-11-10-r106> (2010).
53. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. <https://doi.org/10.1186/s13059-014-0550-8> (2014).
54. Team, R. C. R. *A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020).
55. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
56. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287. <https://doi.org/10.1089/omi.2011.0118> (2012).
57. Tang, Y., Li, M., Wang, J., Pan, Y. & Wu, F. X. CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *Biosystems* **127**, 67–72. <https://doi.org/10.1016/j.biosystems.2014.11.005> (2015).
58. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559. <https://doi.org/10.1186/1471-2105-9-559> (2008).
59. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, 17. <https://doi.org/10.2202/1544-6115.1128> (2005).
60. Yip, A. M. & Horvath, S. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinform.* **8**, 22. <https://doi.org/10.1186/1471-2105-8-22> (2007).
61. Thul, P. J. & Lindskog, C. The human protein atlas: a spatial map of the human proteome. *Protein Sci.* **27**, 233–244. <https://doi.org/10.1002/pro.3307> (2018).

Acknowledgements

This work was supported by the Key Research and Development Program of Science and Technology Department of Zhejiang Province (No. 2019C03081), and Natural Science Foundation of Zhejiang Province, China (LQ21H160027, LQ21H140001). We thank the institutions managing the TCGA and other databases for the freely accessible datasets without limitations.

Author contributions

F.H. designed the overall study with contributions from Y.X. Y.X. downloaded all datasets used in the current study and performed differentially expressed genes analysis between HNSCC and normal samples. F.X. and Y.L. analyzed the data and established the risk model. S.W., J.L., C.Z. and J.J. participated in interpretation of data and prepared figures in this study. Y.X., F.X. and B.X. drafted the main text. All authors contributed to manuscript writing, reviewing and final approval of manuscript. Y.X., F.X. and Y.L. contributed equally to this study.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-86048-x>.

Correspondence and requests for materials should be addressed to B.X. or F.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021