

Identification of Potential Prognostic Biomarkers for Tongue Squamous Cell Carcinoma

Mi Zhang

Fujian Medical University

Sihui Zhang

Fujian Medical University

Ling Wu

Fujian Medical University

Dexiong Li

Fujian Medical University

Jiang Chen (✉ jiangchen@fjmu.edu.cn)

Fujian Medical University

Research

Keywords: tongue squamous cell carcinomas, long noncoding RNA, messenger RNA, biomarkers, prognosis, methylation, overall survival.

Posted Date: November 6th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-102588/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Tongue squamous cell carcinoma (TSCC) is one of the most common types of oral cancer and has a poor prognosis owing to a limited understanding of its pathogenetic mechanisms. The purpose of this study was to explore and identify potential biomarkers in TSCC by integrated bioinformatics analysis.

Methods: The RNA sequencing data, methylation data, and clinical characteristics of TSCC patients were downloaded from The Cancer Genome Atlas (TCGA), and then differentially expressed RNAs (DERNAs), including differentially expressed long noncoding RNAs (DElncRNAs) and differentially expressed messenger RNAs (DEmRNAs), were identified in TSCC by bioinformatics analysis. Subsequently, Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Hallmark pathway analyses were used to analyze the DERNAs. Univariate and multivariate Cox regression analyses were used to develop four-lncRNA and two-mRNA signatures and predict survival in TSCC patients. We established a risk model to predict the overall survival (OS) of TSCC patients based on the DERNAs with Kaplan–Meier analysis and the log-rank p test. Furthermore, weighted gene coexpression network analysis (WGCNA) was performed in Cytoscape, and a protein-protein interaction (PPI) network was established in the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database.

Results: A total of 2,006 DEmRNAs and 1,001 DElncRNAs were found to be dysregulated in TSCC. A total of 417 DERNAs were used to construct the coexpression network, and the PPI network included 103 DEmRNAs. Univariate regression analysis of the DERNAs revealed 51 DElncRNAs and 90 DEmRNAs that were associated with OS in TSCC patients. Multivariate Cox regression analysis demonstrated that four of those lncRNAs (MGC32805, RP1-35C21.2, RP11-108K3.1, and RP11-109M17.2) and two mRNAs (CA9, GTSF1L) had significant prognostic value, and their cumulative risk score indicated that these four-lncRNA and two-mRNA signatures independently predicted OS in TSCC patients. Additionally, there was a positive correlation between the expression and methylation level of RP11-108K3.1, the OS significantly negatively correlated with hypermethylation and low expression of GTSF1L along with hypomethylation and high expression of CA9.

Conclusions: The current findings provide novel insights into the molecular mechanisms of TSCC and identify four lncRNAs and two mRNAs that are potential biomarkers that may be independent prognostic signatures for TSCC diagnosis and treatment.

1. Background

Tongue squamous cell carcinoma (TSCC) is one of the most common and lethal types of oral cancer^[1] and is characterized by poor prognosis, frequent lymphatic metastasis, and a high rate of regional recurrence. Currently, the most preferred treatment for TSCC is surgery combined with postoperative radiotherapy and chemotherapy. However, the 5-year overall survival (OS) rate of TSCC has not improved significantly over the past decades. It is pivotal to diagnose TSCC in the early stage. Therefore,

elucidating the molecular mechanisms and exploring effective molecular therapeutic targets play essential roles in improving the survival rate of patients with TSCC.

Less than 2% of the human genome contains protein-coding genes, while the majority of transcripts are non-protein-coding RNAs^[2]. Among them, long noncoding RNAs (lncRNAs) are RNA transcripts longer than 200 nucleotides and have poor protein-coding capacity. Compared to mRNAs, the expression patterns of cancer-specific lncRNAs seem to be more tissue- and stage-specific, indicating lncRNAs as promising alternative biomarkers and therapeutic targets^[3]. The dysregulated expression of lncRNAs has been reported in distinct pathologies, including TSCC and bladder, prostate, lung, breast, and gastric cancers^[4, 5]. lncRNAs affect various aspects of biological processes in TSCC, including cell proliferation, tumorigenesis, survival, apoptosis, and migration^[6–8]. For example, upregulation of the lncRNA FALEC suppresses malignant behaviors in TSCC^[8]. Overexpression of the lncRNA HOTTIP in TSCC patients indicates a poor clinical prognosis^[7]. Moreover, the lncRNAs KCNQ10T1 and SNHG17 act as competing endogenous RNAs (ceRNAs) to control cell proliferation and cancer progression in TSCC^[5, 9]. In addition, recent studies revealed that lncRNA combined with mRNA biomarkers could improve diagnosis^[10, 11]. Therefore, the identification of additional TSCC-related lncRNAs and investigations of their functions in TSCC are imperative.

N⁶-methyladenosine (m⁶A) is the most prevalent modification in eukaryotic cells^[12, 13] and the most abundant modification in eukaryotic mRNAs and noncoding RNAs (ncRNAs)^[14]. A large body of evidence has demonstrated that the m⁶A modification plays a vital role in RNA translation, stability and alternative splicing, and perturbations in m⁶A components are associated with human diseases, especially cancers^[15–18]. ncRNAs exhibit patterns of m⁶A modifications that are distinct from those of mRNAs^[19, 20]. However, the contribution of the m⁶A modification to ncRNA metabolism remains unexplored.

In this study, we aimed to analyze the differentially expressed profile of noncoding RNAs and mRNAs in TSCC and establish a Cox regression model to predict the OS of patients with TSCC. Moreover, we selected differentially expressed RNAs (DERNAs) related to methylation and then analyzed the correlation between their expression and methylation, as well as the correlation between their methylation level and OS. We also constructed a lncRNA-mRNA coexpression network and a protein-protein interaction (PPI) network. Our findings provide novel insights and may be conducive to understanding the molecular mechanisms. Further, verification of the lncRNAs and mRNAs could provide new therapeutic targets for TSCC.

2. Results

2.1 Identification of DElncRNAs and DEMRNAs

The RNA expression profiles and corresponding clinical data of 146 TSCC patients and 15 normal controls were downloaded from the TCGA database. $|\text{Log}_2\text{FC}| > 2$ and $\text{FDR} < 0.05$ were used as the

thresholds. A total of 1,001 DElncRNAs (680 upregulated and 321 downregulated, Fig. 1A) and 2,006 DEmRNAs (856 upregulated and 1,150 downregulated, Fig. 1B) were identified.

2.2 Functional analysis of the DEmRNAs

GO and KEGG enrichment analyses were performed to explore the potential functions of the DEmRNAs. The results indicated that these genes are enriched in extracellular structure organization, cell differentiation and transporter activity (Fig. 2C). KEGG pathway analysis demonstrated that the most significant pathways were the starch and sucrose metabolism, glycine, serine and threonine metabolism and glycolysis/gluconeogenesis pathways (Fig. 2D).

2.3 Establishment of the differentially expressed RNA (DERNA) prognostic model

First, univariate Cox regression analysis and Kaplan–Meier analysis (log-rank p test) were used to identify lncRNAs associated with the OS of TSCC patients. With the significance level cutoff threshold set at $p < 0.05$ (Table 1), a total of 51 lncRNAs were associated with the OS of TSCC patients. After Kaplan–Meier analysis was performed, as shown in Fig. 3, several lncRNAs including AC007879.5, FOXD2-AS1, LINC01615, MGC32805, PLS3-AS1, RP1-35C21.2, RP11-108K3.1, RP11-109M17.2, RP11-275N1.1, RP11-497D6.3 and RP11-502M1.2 were detected to have significant prognostic value.

Table 1
DElncRNA univariate Cox regression results

lncRNA	HR	z	p-value
RP11-865I6.2	1.422556	3.633421	0.00028
RP11-11N5.3	1.523196	3.357359	0.000787
AC007879.5	1.438161	3.090381	0.001999
RP11-109M17.2	1.196701	3.049139	0.002295
LINC00643	1.727934	2.930043	0.003389
MGC32805	1.367963	2.803139	0.005061
RP11-108K3.1	1.415545	2.583037	0.009793
RP1-35C21.2	1.376952	2.567905	0.010232
RP11-275N1.1	1.361493	2.548927	0.010806
RP3-449H6.1	0.470901	-2.52749	0.011488
SLC8A1-AS1	0.71846	-2.50322	0.012307
RP11-497D6.3	0.77128	-2.47703	0.013248
LINC01615	1.251519	2.459846	0.0139
FOXD2-AS1	1.520807	2.456913	0.014014
RP11-338H14.1	0.773312	-2.44269	0.014578
RP11-502M1.2	0.743464	-2.43872	0.014739
PLS3-AS1	1.331116	2.435826	0.014858
RP11-554D15.3	1.292266	2.423644	0.015366
AC005532.5	1.324041	2.420447	0.015501
SMC2-AS1	0.741963	-2.4047	0.016186
AC005281.1	1.292889	2.394904	0.016625
LINC01191	0.761373	-2.39095	0.016805
CTC-480C2.1	1.160654	2.350085	0.018769
RP11-190J1.3	1.171719	2.329582	0.019828
RP11-66N24.6	1.419531	2.323383	0.020159
PART1	0.867657	-2.30361	0.021245
RP11-11N5.1	1.29495	2.230697	0.025701

lncRNA	HR	z	p-value
RP11-284F21.10	1.155663	2.226429	0.025985
RP11-486A14.2	1.347946	2.220021	0.026417
FAM87A	1.266124	2.214125	0.02682
CTD-2527I21.14	0.75005	-2.21224	0.02695
RP11-63E9.1	1.197589	2.183795	0.028977
RP11-8L2.1	1.15177	2.181352	0.029157
RP11-100E13.1	0.781413	-2.17458	0.029661
AC122136.2	1.304731	2.172188	0.029841
RP11-631N16.4	0.78296	-2.15329	0.031296
RP11-284F21.9	1.137648	2.095205	0.036153
RP13-60M5.2	0.81282	-2.09177	0.036459
LINC00052	1.123159	2.09137	0.036495
RP11-10L7.1	0.759625	-2.08643	0.036939
FLJ43879	0.783353	-2.08438	0.037125
LINC01356	1.203655	2.063676	0.039048
RP11-608O21.1	0.788421	-2.0466	0.040698
RP11-284F21.7	1.152724	2.028537	0.042506
RP11-488P3.1	1.149669	2.025953	0.04277
RP11-430C1.1	1.213022	2.015911	0.043809
RP11-398G24.2	0.771309	-2.01524	0.043879
LINC01497	1.209287	2.006391	0.044815
LINC01555	1.269332	1.994386	0.04611
CTD-2184D3.5	1.265504	1.989985	0.046593
CTD-2337I7.1	1.216582	1.967779	0.049093

Then, 17 candidate DElncRNAs with p -value < 0.015 were all fitted into the multivariate Cox regression model. The results indicated that SLC8A1-AS1, FOXD2-AS1, MGC32805, RP11-108K3.1, PLS3-AS1, RP3-449H6.1, RP11-109M17.2, RP11-11N5.3, RP1-35C21.2 and RP11-865I6.2 had significant prognostic value

in TSCC (Table 2), and all candidate lncRNAs involved in the multivariate Cox regression model were used to develop a lncRNA prognostic model.

Table 2
DElncRNA multivariate Cox regression results

lncRNA	β-value	HR	z	p-value
MGC32805	0.77297	2.16619	4.893	9.91E-07
RP11-865I6.2	0.58052	1.78697	4.81	1.51E-06
RP11-11N5.3	0.47345	1.60552	3.665	0.000247
SLC8A1-AS1	-0.50666	0.6025	-3.417	0.000634
RP11-108K3.1	0.58357	1.79242	3.237	0.001209
FOXD2-AS1	0.61337	1.84665	3.033	0.002423
RP3-449H6.1	-1.03457	0.35538	-2.499	0.012451
RP11-109M17.2	0.12841	1.13702	2.131	0.033116
RP1-35C21.2	0.32227	1.38026	2.124	0.033678
PLS3-AS1	0.30076	1.35088	2.115	0.034406
RP11-502M1.2	-0.2267	0.79716	-1.646	0.099728
RP11-497D6.3	-0.20679	0.81319	-1.624	0.104482
Likelihood ratio test = 100.6 on 12 df, $p = 4.331E-16$				
n = 146, number of events = 51				

We also adopted the same approach to verify the mRNAs associated with the OS of TSCC patients (Table 3), and a total of 90 mRNAs were associated with the OS of TSCC patients. As shown in Fig. 4, many mRNAs, including ADH1B, ASCL4, C6, CA9, FAM65C, GNG8, GTSF1L, ODF4, OVOS2, PCK1, and SHOX2, were found to have significant prognostic value. Fifteen candidate DEmRNAs with p -value < 0.01 were all fitted into the multivariate Cox regression model, and the results showed that CA9, OVOS2, LRRTM3, C7, GTSF1L, PCK1, F2RL2 and ODF4 had significant prognostic value in TSCC (Table 4).

Table 3
DEmRNA univariate Cox regression results

mRNA	HR	z	p-value
GNG8	0.668118	-3.50266	0.000461
CA9	1.272047	3.323934	0.000888
OVOS2	0.788196	-3.32344	0.000889
C6	1.223323	3.273786	0.001061
FAM65C	0.703648	-3.07276	0.002121
PCK1	1.22433	3.068226	0.002153
GTSF1L	0.681065	-3.01555	0.002565
ODF4	0.632626	-2.85445	0.004311
C7	1.182026	2.733506	0.006266
ADH1B	1.140672	2.710891	0.00671
F2RL2	0.771353	-2.67225	0.007535
ADIPOQ	1.120641	2.65977	0.007819
LRRTM3	1.405741	2.61057	0.009039
SHOX2	0.742186	-2.60479	0.009193
ASCL4	0.726264	-2.57974	0.009888
RSPH4A	0.694298	-2.57261	0.010093
AC008810.1	0.617492	-2.56539	0.010306
HOXD9	0.727773	-2.55644	0.010575
SPINK7	0.908908	-2.52329	0.011626
KLK1	1.244047	2.516662	0.011847
RNF212B	0.75701	-2.48593	0.012921
RIBC1	0.666136	-2.47609	0.013283
ZG16	1.261833	2.465508	0.013682
C7orf33	1.307349	2.463732	0.01375
SFTPC	1.177867	2.44232	0.014593
CASP14	1.112717	2.428729	0.015152
AL590235.1	0.721791	-2.42216	0.015428

mRNA	HR	z	p-value
KLF7	1.563083	2.419143	0.015557
TTC16	0.778389	-2.40787	0.016046
CFAP73	0.726075	-2.38832	0.016925
MOV10L1	0.759161	-2.37089	0.017745
SOX1	1.133833	2.358584	0.018345
FGA	1.136625	2.347279	0.018911
KCNJ16	1.18101	2.346808	0.018935
CXCL13	0.842367	-2.33704	0.019437
DNAAF1	0.750386	-2.32597	0.02002
KRT33A	0.808128	-2.29248	0.021878
KRT39	1.284344	2.28228	0.022473
NXPH1	1.311622	2.270645	0.023168
SOHLH1	1.087993	2.268472	0.0233
CGNL1	0.815958	-2.26673	0.023407
KRT33B	0.825259	-2.26171	0.023715
RTP3	0.882223	-2.23521	0.025403
CLVS2	1.261847	2.228258	0.025863
DNER	1.156868	2.215978	0.026693
INSM1	0.81694	-2.21324	0.026881
TCEAL6	1.465081	2.212831	0.026909
AMTN	0.910094	-2.19567	0.028116
AQP7	1.177243	2.175438	0.029597
ALB	1.119201	2.172115	0.029847
ETNPPL	1.210195	2.169777	0.030024
GJB7	1.117709	2.166463	0.030276
TMEM232	1.347199	2.16267	0.030567
CST5	1.230666	2.144369	0.032003
FGG	1.13366	2.140215	0.032337

mRNA	HR	z	p-value
HIST1H3G	1.197317	2.136028	0.032677
C14orf180	1.316759	2.11975	0.034027
SLC6A4	0.825596	-2.11584	0.034358
PTCH2	0.747119	-2.1061	0.035196
PPP1R3C	1.179677	2.104002	0.035378
AC009060.2	0.814285	-2.09827	0.035881
PLIN4	1.1366	2.091326	0.036499
CAPS	0.745138	-2.08551	0.037023
SSX1	0.723928	-2.08103	0.037431
HIST1H2AG	1.27392	2.072023	0.038263
HTN3	1.125145	2.071847	0.03828
CGB8	0.868005	-2.07144	0.038318
TSNAXIP1	0.704588	-2.06564	0.038862
PLIN1	1.141516	2.061308	0.039274
JAML	0.800447	-2.05423	0.039954
GOLGA6L2	0.843113	-2.05132	0.040235
PAGE1	0.613576	-2.04904	0.040458
CASC1	0.79658	-2.04356	0.040997
PLPPR5	0.838986	-2.04324	0.041029
IL36A	0.913631	-2.03106	0.042249
CLEC3B	0.803616	-2.03096	0.042259
SERPINA9	0.874336	-2.02217	0.043159
RNF224	0.832738	-2.02016	0.043367
CFD	1.208165	2.00777	0.044668
SMR3B	1.119902	2.001821	0.045304
WDR38	0.799472	-1.98747	0.04687
SGIP1	0.789217	-1.98572	0.047065
RIMS4	1.157784	1.973463	0.048443

mRNA	HR	z	p-value
APELA	0.868882	-1.9734	0.04845
LAMB4	0.863059	-1.97128	0.048692
KLK12	0.917869	-1.9689	0.048965
ANGPTL5	1.234164	1.968716	0.048986
LAIR2	0.837717	-1.96821	0.049044
CFAP43	0.789604	-1.96655	0.049235
MMP19	0.768498	-1.96345	0.049594

Table 4
DEmRNA multivariate Cox regression results

mRNA	β-value	HR	z	p-value
CA9	0.35074	1.42012	4.573	4.81E-06
OVOS2	-0.30854	0.73452	-3.52	0.000432
LRRTM3	0.51937	1.68097	3.322	0.000893
C7	0.24123	1.27282	3.001	0.002687
GTSF1L	-0.4969	0.60841	-2.874	0.004059
PCK1	0.23855	1.26941	2.745	0.006049
F2RL2	-0.27071	0.76284	-2.318	0.020465
ODF4	-0.33474	0.71552	-2.153	0.03131
Likelihood ratio test = 80.93 on 8 df, p = 3.179E-14				
n = 146, number of events = 51				

Risk score analysis of the lncRNAs was performed for each patient, and based on the risk scores, the patients were divided into the low-risk and high-risk groups. The survival rate of the low-risk patients was significantly higher than that of the high-risk patients (Fig. 5E). The heatmap in Fig. 5F shows the relationships between the expression of 12 candidate lncRNAs and the risk status of each patient. The 5-year survival correlation of the 12-lncRNA signature was assessed by ROC analysis, and the area under the curve (AUC) was computed to assess the discriminatory capacity of the lncRNA signature (Fig. 5G). The AUC of the 12-lncRNA signature was 0.876, indicating its utility as a prognostic model for predicting the survival status of TSCC patients. Significantly, we found that FOXD2-AS1, MGC32805, PLS3-AS1, RP1-35C21.2, RP11-108K3.1, and RP11-109M17.2 were simultaneously identified to be associated with

OS in Kaplan–Meier and multivariate Cox regression analyses. Moreover, risk score analysis of the eight candidate mRNAs was performed for each patient (Fig. 5H) and a heatmap was constructed (Fig. 5I). Subsequently, the ROC curve was plotted, and the AUC was 0.84 (Fig. 5J). Likewise, CA9, GTSF1L, ODF4, OVOS2, and PCK1 were simultaneously verified to be associated with OS in Kaplan–Meier and multivariate Cox regression analyses.

The expression of FOXD2-AS1, MGC32805, PLS3-AS1, RP1-35C21.2, RP11-108K3.1 and RP11-109M17.2 was then analyzed in TSCC and normal tissues, as well as in the low- and high-risk paired groups (Fig. 6). FOXD2-AS1, PLS3-AS1, RP1-35C21.2, RP11-108K3.1, and RP11-109M17.2 were expressed at high levels, and MGC32805 was expressed at low levels in TSCC patients. Similarly, the expression of CA9, GTSF1L, ODF4, OVOS2, and PCK1 was analyzed in the tumor and normal tissues, as well as in the paired groups (Fig. 7). CA9, GTSF1L, and OVOS2 were expressed at high levels, and ODF4 and PCK1 were expressed at low levels in TSCC patients.

2.4 Prognostic value of the four-lncRNA and two-mRNA signatures in TSCC

Univariate regression analysis was then used to predict the potential risk factors for the long-term all-cause death of patients, including age, sex, grade, stage, T stage, N stage, and the expression of each lncRNA or mRNA. Multivariate regression models were conducted to assess the prognostic power of each lncRNA and mRNA signature separately. Potential risk factors with $p < 0.05$ were included in the multivariate Cox regression analysis to screen independent risk factors for predicting long-term death. Univariate analysis indicated that grade, stage, T stage, and N stage were significantly correlated with the OS of TSCC patients ($p < 0.05$) (Supplementary Table 1). As with the DElncRNAs, the DEmRNA results were obtained using the same approach (Supplementary Table 2). Of particular note, multivariate analysis found that the expression of MGC32805, RP1-35C21.2, RP11-108K3.1, RP11-109M17.2, CA9 and GTSF1L were significantly correlated with the OS of TSCC patients ($p < 0.05$) (Fig. 8).

2.5 GSEA of the four-lncRNA and two-mRNA signatures

The GSEA results showed that for patients with high signature scores, MGC32805 was mainly enriched in the GO terms of RNA binding and nuclease activity; RP1-35C21.2 was mainly enriched in the GO molecular function terms of regulator and regulation of lymphocyte apoptotic process; RP11-108K3.1 was mainly enriched in the GO terms of identical protein binding and positive regulation of cellular component biogenesis; RP11-109M17.2 was mainly enriched in the GO terms of RNA splicing and mRNA splice site selection; and CA9 was mainly enriched in the GO terms of pyruvate metabolic process, nucleotide phosphorylation, ADP metabolic process, steroid hydroxylase activity, and aromatase activity. Moreover, the mainly enriched KEGG terms included cell cycle, mTOR signaling pathway, glycolysis/gluconeogenesis, spliceosome, and RNA degradation. In addition, for patients with high signature scores, CA9 was mainly enriched in the Hallmark terms of glycolysis, mTORC1 signaling, hypoxia, MYC targets V2, and PI3K/AKT/MTOR signaling. For GTSF1L, those with high signature scores showed enrichment in the GO terms of immune receptor activity, negative regulation of lymphocyte activation, regulation of B cell activation, regulation of leukocyte proliferation, and regulation of T cell activation; mainly showed enrichment in KEGG terms including B cell receptor signaling pathway,

cytokine-cytokine receptor interaction, JAK/STAT signaling pathway, natural killer cell-mediated cytotoxicity, and T cell receptor signaling pathway; and mainly showed enrichment in the Hallmark terms of complement, IL2/STAT5 signaling, IL6/JAK/STAT3 signaling, inflammatory response, and MYC target V1 ($p < 0.05$; Fig. 9; Supplementary Table 3).

2.6 Validation of the expression and methylation levels of the potential signatures

The methylation levels of the six DERNA identified above were further validated in the TCGA database. The correlations between the expression of the potential signatures and their methylation levels are shown in Fig. 10. There is too little data on RP11-109M17.2 to analyze. It was noted that the upregulation of RP11-108K3.1 expression showed the most significant positive correlation with hypermethylation, indicating that methylation is important for its expression.

2.7 Correlation of the expression and methylation of the potential signatures with OS

The TCGA dataset was used to analyze the association of the potential signatures with OS. After combining the potential biomarker's expression and methylation, the OS significantly negatively correlated with hypermethylation and low expression of GTSF1L along with hypomethylation and high expression of CA9 (Fig. 11–12).

2.8 Weighted gene coexpression network construction

Moreover, a coexpression network was constructed successfully. The expression levels of 3,007 genes, including 2,006 mRNAs and 1,001 lncRNAs, were used for coexpression network construction by the WGCNA package. In the current study, to ensure a scale-free network, we selected $\beta = 12$ as the soft-thresholding power (Supplementary Fig. 1). Then, the coexpression network was constructed by 417 DERNA and visualized in Cytoscape. Subsequently, three subnetworks of the coexpression network were constructed by using the MCODE tool of Cytoscape (Supplementary Fig. 2).

2.9 PPI network analysis

The PPI network contained a total of 619 proteins and 2,639 edges. With the thresholds of confidence score > 0.9 , 103 genes were selected to construct three subnetworks with degree ≥ 20 by using the MCODE tool of Cytoscape (Supplementary Fig. 3). With the thresholds of confidence score > 0.7 , Fig. 13 shows the protein network interacting with CA9 (Fig. 13K) and GTSF1L (Fig. 13L). According to GO term analysis, the three modules were related to extracellular matrix component, G protein-coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger, actin binding, and posttranslational protein modification, showing that the genes in these modules play a pivotal role in cancer. Regarding the KEGG enrichment analysis, module 1 regulated metabolic pathways such as protein digestion and absorption, human papillomavirus infection, and the PI3K-Akt signaling pathway. Module 2 was significantly associated with hypertrophic cardiomyopathy (HCM). Moreover, module 3 was significantly related to the chemokine signaling pathway, calcium signaling pathway, and cAMP signaling pathway, which are associated with the genesis and progression of tumors (Supplementary Tables 4, 5, and 6).

2.10 Validation of the biomarkers

In summary, all the above results show that four lncRNAs, including MGC32805, RP1-35C21.2, RP11-108K3.1, and RP11-109M17.2, and two mRNAs, including CA9 and GTSF1L, were simultaneously identified to be associated with the OS of TSCC patients. Thus, they could act as potential prognostic biomarkers in TSCC. The differential expression information of the biomarkers is shown in Supplementary Table 7. qRT-PCR was performed to verify the expression levels of the candidate biomarkers in TSCC cell lines. The expression data on MGC32805, RP1-35C21.2, RP11-108K3.1, RP11-109M17.2, CA9, and GTSF1L were consistent with the sequencing data (Fig. 14M). In addition, we also verified the expression of CA9 and GTSF1L in GEPIA database (Fig. 14N). All primers are listed in Supplementary Table 8.

3. Discussion

TSCC is a one of lethal malignant cancer, and the lack of specific diagnostic and prognostic biomarkers contribute to the currently low 5-year survival rate of TSCC patients. Therefore, it is crucial to explore the comprehensive regulatory mechanisms of TSCC progression and to verify potential predictive biomarkers for the early diagnosis and prognosis of TSCC.

Emerging evidence indicates that most lncRNAs play essential roles in TSCC progression, with diverse underlying mechanisms. In this study, a total of 2,006 DEmRNAs and 1,001 DElncRNAs were identified. GO analysis demonstrated that the functions of the DERNAs were mainly associated with extracellular structure organization, cell differentiation and transporter activity, which play vital roles in tumorigenesis. In addition, KEGG pathway analyses revealed that the most significant pathways were the starch and sucrose metabolism, glycine, serine and threonine metabolism and glycolysis/gluconeogenesis pathways. All of these pathways are significantly associated with cancers. For example, the genetic activation of serine biosynthesis has been indicated to directly drive cancer proliferation; intriguingly, genetic and functional evidence supports the importance of the activity of serine and glycine biosynthesis as drivers of oncogenesis^[26]. Furthermore, differentially expressed proteins showed significant enrichment for the glycolysis/gluconeogenesis pathways in breast cancer patients^[27].

With the goal of identifying lncRNAs and mRNAs significantly associated with OS, we established a Cox regression model to predict the OS of patients with TSCC. Univariate regression analysis of the DElncRNAs identified 17 candidate lncRNAs associated with OS. Multivariate analysis showed that 12 of them were related to OS, and 10 of those lncRNAs were significantly associated with OS ($p < 0.05$). A cumulative risk score of the 12 lncRNAs was calculated, which indicated that they may independently predict OS in TSCC patients. Then, six of these lncRNAs were identified to have significant prognostic value after Kaplan–Meier analysis was carried out. The same approach was also applied for the DEmRNAs. Univariate and multivariate regression analyses were then used to predict the potential risk factors for long-term all-cause death in TSCC patients. Four of those lncRNAs (MGC32805, RP1-35C21.2, RP11-108K3.1, and RP11-109M17.2) and two mRNAs (CA9 and GTSF1L) were simultaneously identified to be associated with OS in TSCC patients. Moreover, we also established a lncRNA-mRNA coexpression network and a PPI network using the information obtained from the TCGA database. Our findings could

help improve the understanding of lncRNA-mRNA coexpression regulatory mechanisms and provide new insights for therapeutic targets in TSCC.

In the current study, among these four lncRNA biomarkers, a genome-wide association study (GWAS) showed that MGC32805 may be involved in paclitaxel in vitro drug response phenotypes, which are moderately associated with the time to epithelial ovarian cancer recurrence^[28]. Another GWAS revealed that the SNP rs146949893 in RP1-35C21.2 may be related to height^[29]. In addition, previous studies have shown that RP11-108K3.1 could serve as a potential independent biomarker for the prognostic prediction of colorectal cancer^[30]. RP11-108K3.1 is one of the DElncRNAs in lung adenocarcinoma^[31]. However, no study so far has reported any association of RP11-109M17.2 with cancer, which means that it may serve as a novel therapeutic target in TSCC. All of these studies confirm that the DElncRNAs screened in our research play critical roles in tumor pathogenesis. However, the underlying mechanisms remain unknown. In our study, the GO analysis of the four lncRNAs showed that they were mainly enriched in RNA binding and nuclease activity, regulator and regulation of lymphocyte apoptotic process, identical protein binding and positive regulation of cellular component biogenesis, and RNA splicing and mRNA splice site selection. All of these processes are significantly associated with cancers.

For the two-mRNA signature, research has shown that high CA9 expression is related to poor prognosis and tumor grade in TSCC^[32]; in addition, the GAA haplotype of 3 CA9 SNPs (rs2071676, rs3829078, and rs1048638) is correlated with a higher risk of oral cancer^[33]. Natural selection shapes cancer genomes, and the study revealed that negative selection is a hallmark of cell essentiality and immune response in cancer. Although there was no significant difference, GTSF1L may be related to negative selection^[34]. In addition, GTSF1L could also serve as a biomarker of common reproductive impairments in males^[35]. In our findings, the KEGG analysis of CA9 and GTSF1L showed that they were mainly enriched in glycolysis/gluconeogenesis, the B cell receptor signaling pathway, the JAK/STAT signaling pathway, natural killer cell-mediated cytotoxicity, and the T cell receptor signaling pathway. Furthermore, the Hallmark analysis of CA9 and GTSF1L showed that they were mainly enriched in mTORC1 signaling, MYC targets V2, PI3K/AKT/MTOR signaling, IL2/STAT5 signaling, and IL6/JAK/STAT3 signaling. These pathways all play essential roles in tumor immunity.

We also verified that m⁶A methylation, not DNA methylation or histone acetylation, is involved in these six signatures in TSCC. As one of the most common RNA modifications, m⁶A may regulate all stages of the RNA life cycle, such as RNA processing, nuclear export and translation^[36, 37]; and can therefore modulate cancer progression processes, such as cell proliferation^[38] and tumorigenesis^[39]. However, investigations on the functions of m⁶A in lncRNAs are lacking. One recent study revealed that the METTL3-mediated N⁶-methyladenosine modification led to LINC00958 upregulation by stabilizing its RNA transcript in hepatocellular carcinoma^[40]. A similar study showed that in colorectal cancer cells, METTL14 knockdown substantially abolished the m⁶A level of XIST and augmented XIST expression^[41]. Moreover, LINC00266-1, by encoding RBP to strengthen m⁶A recognition on RNAs, ultimately increases mRNA stability and expression^[42]. However, few studies have demonstrated that lncRNAs modulate

tumorigenesis by regulating the level of methylation in colorectal cancer^[43, 44]. Our study indicates that one of the six potential biomarkers, RP11-108K3.1, mediates tumor progression by regulating its methylation level. Besides, after combining the potential biomarker's expression and methylation, the OS significantly negatively correlated with hypermethylation and low expression of GTSF1L along with hypomethylation and high expression of CA9. Our report enriches the diversity of noncoding RNA functions in the m⁶A RNA pathway and in tumorigenesis, providing an important role for lncRNAs in RNA metabolism.

In our study, the expression of MGC32805, RP1-35C21.2, RP11-108K3.1, RP11-109M17.2, CA9, and GTSF1L was then analyzed in TSCC and normal tissues, and their expression in the paired groups was also found to be significantly associated with poor clinical outcomes in TSCC patients. This is the first study to indicate the aberrant expression of MGC32805, RP1-35C21.2, RP11-108K3.1, and RP11-109M17.2 in TSCC and indicates a potential prognostic role of this 4-lncRNA signature in TSCC. We also revealed a positive association between the expression and methylation of RP11-108K3.1. In addition, our findings also support that CA9 and GTSF1L could be potential biomarkers in TSCC. The bioinformatics-based investigation of lncRNAs will contribute to future experimental research.

4. Conclusion

Taking all the above results together, we identified the four-lncRNA and two-mRNA signatures as potential prognostic biomarkers for TSCC patients by analyzing genome-wide lncRNA and mRNA expression data from the TCGA database. Furthermore, we uncovered a connection between potential biomarkers and m⁶A methylation in TSCC. Additionally, we established a lncRNA-mRNA coexpression network that is beneficial to understanding the relationships between lncRNAs and mRNAs and provides effective strategies for further functional studies of lncRNAs. Moreover, construction of the PPI network provides novel insights for TSCC treatment, and the risk score model contributes to improving OS in TSCC.

5. Methods

5.1 Patient information collection and data preparation

Data associated with TSCC, including the RNA sequencing (RNA-Seq) data of transcriptome profiling, clinical data, and methylation data, were retrieved from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>) and were downloaded through the Genomic Data Commons (GDC) data transfer tool. In the predictive survival-related model, we excluded 27 samples because they lacked complete clinical data. Finally, 146 TSCC samples (119 samples for the predictive survival-related model) and 15 normal control samples were collected in our study.

5.2 Data processing and differentially expressed gene analysis

The edgeR package in R (version 3.6.3) was used to identify the differentially expressed mRNAs (DEmRNAs) with the thresholds of $|\log_2(\text{fold change [FC]})| > 2.0$ and false discovery rate (FDR) (adjusted

p -value) < 0.05^[21]. Then, we used the annotation file in GTF format (Homo_sapiens.GRCh38.100.chr.gtf) to identify and annotate differentially expressed lncRNAs (DElncRNAs) with the thresholds of $|\log_2FC| > 2.0$ and $FDR < 0.05$. A heatmap and volcano plot were constructed by the gplots package in R software. A scatter diagram and paired plot were generated by the Wilcox test in R software. We followed the methods of ^[22].

5.3 Cell lines

Two human TSCC cell lines, CAL-27 and SCC-9, were purchased from the American Type Culture Collection (ATCC; Manassas, VA, USA), and the human normal oral keratinocyte cell line hNOK was obtained from the Institute of Biochemistry and Cell Biology of the Chinese Academy of Sciences (Shanghai, China). CAL-27 and hNOK cells were grown in DMEM (Gibco) supplemented with 10% FBS (Invitrogen, Carlsbad, CA, USA). SCC-9 cells were cultured in DMEM/F-12 (Gibco) supplemented with 10% FBS and 0.4 $\mu\text{g/ml}$ hydrocortisone. All cells were maintained at 37 °C in an incubator supplied with 5% CO_2 .

5.4 RNA extraction and quantitative real-time PCR analysis

Total RNA was extracted from cells using TRIzol reagent (Invitrogen, CA, USA) according to the manufacturer's instructions. Total RNA was reverse transcribed into cDNA using a PrimeScript RT Reagent Kit (TaKaRa). RT-qPCR analyses were performed using SYBR Green Master Mix (TaKaRa). This process was performed using an Roche 480 Real-Time PCR System. The results were normalized to the expression of glyceraldehyde-3-phosphate dehydrogenase (GAPDH). The relative expression level was calculated by the $2^{-\Delta\Delta Ct}$ method. The primers were provided by the SunYa Company. In addition, we also verified the potential biomarkers in GEPIA database (<http://gepia.cancer-pku.cn/detail.php>)^[23].

5.5 Functional enrichment analysis

Gene Ontology (GO) analysis included biological processes, cellular components and molecular functions. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was used to annotate the potential functions. GO and KEGG functional analyses were conducted for the DElncRNAs. A significance level of $p < 0.05$ was set as the cutoff criteria, and the plots were constructed by the ggplot2 and clusterProfiler packages in R software^[24].

5.6 Gene set enrichment analysis (GSEA)

The presumptive biological roles of the DElncRNAs correspond to those of their associated mRNAs. The expression profile data were ranked according to the signature score, and the data were divided into a high-risk group and a low-risk group by the mean score. Then, GSEA (version 4.0.3) software was used for analysis.

The OS and hazard ratios (HRs) were calculated by the Kaplan–Meier algorithm and univariate Cox regression analysis, respectively. The log-rank method was used to test the differences between the survival curves. The univariate Cox proportional hazards regression method was applied to analyze the relationships between the DElncRNAs and OS when the significance level was set at 0.05 to determine

those with prognostic value in TSCC. We used the following formula to establish a prognostic risk score model: $\text{Risk score} = \exp_{\text{IncRNA1}} * \beta_{\text{IncRNA1}} + \exp_{\text{IncRNA2}} * \beta_{\text{IncRNA2}} + \dots + \exp_{\text{IncRNA}_n} * \beta_{\text{IncRNA}_n}$; $\text{Risk score} = \exp_{\text{Gene1}} * \beta_{\text{Gene1}} + \exp_{\text{Gene2}} * \beta_{\text{Gene2}} + \dots + \exp_{\text{Gene}_n} * \beta_{\text{Gene}_n}$ (where exp is the expression level of the prognostic DEIncRNA or DEmRNA and β is the corresponding multivariate Cox regression model regression coefficient). The gene signature score as a prognostic predictor for TSCC patients was evaluated in the model. We determined the significant variables through univariate Cox regression analysis. Candidate variables with a *p*-value of < 0.015 for DEIncRNAs and a *p*-value of < 0.01 for DEmRNAs in univariate analysis were included in the multivariable model. Setting the median risk score as the threshold, the TSCC patients were stratified into high- and low-risk groups. The survival ROC package in R was used to construct time-dependent receiver operating characteristic (ROC) curves within 5 years as the defining point and to calculate the risk prediction rate of specific lncRNAs and mRNAs between the two groups. Finally, univariate and multivariate analyses were implemented to evaluate the effects of other clinical variables of TSCC patients on the OS risk score. R software was used for all statistical analyses.

5.8 Validation of the expression and methylation levels of the identified DERNAs

Next, we integrated the expression data and methylation data on the DERNAs. DERNA expression was compared between TSCC samples and normal samples.

5.9 Correlation of the expression and methylation of the identified DERNAs with OS

First, patients with methylation data were classified in two different categories in accordance with the expression level of each DERNA (low vs. high). Second, an analysis was conducted on the patients with both transcriptome and methylation data. Finally, we performed Kaplan–Meier survival analysis with the log-rank test by adopting the “survival” package in R software.

5.10 Weighted gene coexpression network analysis (WGCNA)

The WGCNA package in R was used to construct a coexpression network utilizing the expression values of the mRNAs and lncRNAs screened above. In brief, we constructed the weighted adjacency matrix using a power function based on a soft-thresholding parameter β . In this study, we chose a β value of 12. The DEIncRNA and DEmRNA coexpression network was visualized in Cytoscape (Version 3.7.2) software. Subsequently, module analysis^[25] of the coexpression network was performed using the Molecular Complex Detection (MCODE) tool of Cytoscape software.

5.11 PPI analysis

The DEmRNAs with FDR < 0.001 and confidence score > 0.9 were included in the PPI network through the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (<https://string-db.org/>), confidence score > 0.7 for CA9 and GTSF1L, the PPI network was visualized in Cytoscape. The subnetworks were analyzed by the MCODE tool, and GO and KEGG analyses of the modules was carried out using R software.

5.12 Statistical analysis

All statistical analyses were performed using R software. For the survival analysis, the differences between two groups were assessed by the log-rank p test, and for the paired plot analysis, the differences between two groups were assessed by the Wilcoxon test. *P*-values < 0.05 were considered statistically significant.

Limitations of the study

Although the findings of our study have crucial clinical implications, the limitations should also be noted. First, a longer follow-up duration is needed to confirm our results, and second, the findings based on the TCGA database need to be verified using cell molecular experimental methods. In addition, subsequent functional studies are needed to elucidate the molecular mechanisms underlying the functions of the lncRNAs in TSCC, and the biological functions of MGC32805, RP1-35C21.2, RP11-108K3.1, RP11-109M17.2, CA9, and GTSF1L in TSCC also need to be further investigated.

Abbreviations

TSCC: tongue squamous cells carcinomas; lncRNA: long non-coding RNA; DEL: differentially expressed lncRNA; DEG: differentially expressed mRNA; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; WGCNA: weighted gene co-expression network analysis; PPI: protein-protein interaction.

Synonyms

RP1-35C21.2 (AL136114.1); RP11-108K3.1 (MIR4713HG); RP11-109M17.2 (AL162413.1).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analysed during the current study are available in The Cancer Genome Atlas (TCGA) repository (<https://portal.gdc.cancer.gov/>).

Competing interests

The authors declare no conflicts of interest.

Funding

This work was supported by the National Natural Science Foundation of China (Nos. 81771126) and Young Scholars Fund for scientific research, School and Hospital of Stomatology, Fujian Medical University (Grant number: 2016-1-71).

Authors' contributions

Conception and design: MZ, LW;

Administrative support: JC, SZ;

Provision of study materials or patients: MZ, LW, DL;

Collection and assembly of data: MZ, LW;

Data analysis and interpretation: MZ, SZ;

Writing of the manuscript: All authors;

Final approval of manuscript: All authors;

Acknowledgements

We would like to express our sincere thanks to all those who have helped us in the course of our writing this paper. We sincerely thank Fujian Key Laboratory of Oral Diseases of Fujian College and University (School and Hospital of Stomatology, Fujian Medical University, Fuzhou, China) and Public Technology Service Center (Fujian Medical University, Fuzhou, China) for technical guidance regarding the experiments described in this article.

References

- [1] Mannelli G, Arcuri F, Agostini T, et al. Classification of tongue cancer resection and treatment algorithm[J]. *Journal of Surgical Oncology*, 2018,117(5):1092-1099.
- [2] Sarah, Djebali, Carrie, et al. Landscape of transcription in human cells[J]. *Nature*, 2012,489:101-108.
- [3] Jun, Wang, Xuan, et al. Regulatory roles of long noncoding RNAs implicated in cancer hallmarks. [J]. *International Journal of Cancer*, 2019.
- [4] Gupta R A, Shah N, Wang K C, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis[J]. *Nature*, 2010,464(7291):1071-1076.
- [5] Zhang S, Ma H, Zhang D, et al. LncRNA KCNQ1OT1 regulates proliferation and cisplatin resistance in tongue cancer via miR-211-5p mediated Ezrin/Fak/Src signaling[J]. *Cell Death Dis*, 2018,9(7):742.

- [6] Fang Z, Zhang S, Wang Y, et al. Long non-coding RNA MALAT-1 modulates metastatic potential of tongue squamous cell carcinomas partially through the regulation of small proline rich proteins[J]. *BMC Cancer*, 2016,16:706.
- [7] Zhang H, Zhao L, Wang Y X, et al. Long non-coding RNA HOTTIP is correlated with progression and prognosis in tongue squamous cell carcinoma[J]. *Tumour Biol*, 2015,36(11):8805-8809.
- [8] Jia B, Xie T, Qiu X, et al. Long noncoding RNA FALEC inhibits proliferation and metastasis of tongue squamous cell carcinoma by epigenetically silencing ECM1 through EZH2[J]. *Aging (Albany NY)*, 2019,11(14):4990-5007.
- [9] Liu X, Zhang B, Jia Y, et al. SNHG17 enhances the malignant characteristics of tongue squamous cell carcinoma by acting as a competing endogenous RNA on microRNA-876 and thereby increasing specificity protein 1 expression[J]. *Cell Cycle*, 2020,19(6):711-725.
- [10] Shao T, Huang J, Zheng Z, et al. SCCA, TSGF, and the Long Non-Coding RNA AC007271.3 are Effective Biomarkers for Diagnosing Oral Squamous Cell Carcinoma[J]. *Cellular Physiology & Biochemistry*, 2018,47:26-38.
- [11] GZ H, QQ W, ZN Z, et al. Identification of Candidate Biomarkers and Analysis of Prognostic Values in Oral Squamous Cell Carcinoma. [J]. *Frontiers in oncology*, 2019,9:1054.
- [12] Fu Y, Dominissini D, Rechavi G, et al. Gene expression regulation mediated through reversible m(6)A RNA methylation[J]. *Nat Rev Genet*, 2014,15(5):293-306.
- [13] Stower, Hannah. Transcriptome-wide N6-methyladenosine analysis[J]. *Nature Reviews Genetics*, 2012,13(7):452.
- [14] Li S, Mason C E. The pivotal regulatory landscape of RNA modifications[J]. *Annu Rev Genomics Hum Genet*, 2014,15:127-150.
- [15] Schwartz S, Mumbach M R, Jovanovic M, et al. Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites[J]. *Cell Rep*, 2014,8(1):284-296.
- [16] Lin S, Choe J, Du P, et al. The m(6)A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells[J]. *Mol Cell*, 2016,62(3):335-345.
- [17] Zheng G, Dahl J A, Niu Y, et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility[J]. *Mol Cell*, 2013,49(1):18-29.
- [18] Yang X, Zhang S, He C, et al. METTL14 suppresses proliferation and metastasis of colorectal cancer by down-regulating oncogenic long non-coding RNA XIST[J]. *Mol Cancer*, 2020,19(1):46.

- [19] Zhou C, Molinie B, Daneshvar K, et al. Genome-Wide Maps of m6A circRNAs Identify Widespread and Cell-Type-Specific Methylation Patterns that Are Distinct from mRNAs[J]. *Cell Rep*, 2017,20(9):2262-2276.
- [20] Zhu S, Wang J Z, Chen, et al. An oncopeptide regulates m(6)A recognition by the m(6)A reader IGF2BP1 and tumorigenesis[J]. *Nat Commun*, 2020,11(1):1685.
- [21] Lun A T, Chen Y, Smyth G K. It's DE-licious: A Recipe for Differential Expression Analyses of RNA-seq Experiments Using Quasi-Likelihood Methods in edgeR[J]. *Methods Mol Biol*, 2016,1418:391-416.
- [22] Xiao H, He W, Chen P, et al. Identification of Seven Aberrantly Methylated and Expressed Genes in Adrenocortical Carcinoma[J]. *Front Endocrinol (Lausanne)*, 2019,10:472.
- [23] Tang Z, Li C, Kang B, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses[J]. *Nucleic Acids Res*, 2017,45(W1): W98-W102.
- [24] Yu G, Wang L G, Han Y, et al. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters[J]. *Omics A Journal of Integrative Biology*, 2012,16(5):284-287.
- [25] Bader G D, Hogue C W. An automated method for finding molecular complexes in large protein interaction networks[J]. *BMC Bioinformatics*, 2003,4:2.
- [26] Amelio I, Cutruzzola F, Antonov A, et al. Serine and glycine metabolism in cancer[J]. *Trends Biochem Sci*, 2014,39(4):191-198.
- [27] Zeidan B, Manousopoulou A, Garay-Baquero D J, et al. Increased circulating resistin levels in early-onset breast cancer patients of normal body mass index correlate with lymph node negative involvement and longer disease free survival: a multi-center POSH cohort serum proteomics study[J]. *Breast Cancer Res*, 2018,20(1):19.
- [28] Fridley B L, Ghosh T M, Wang A, et al. Genome-Wide Study of Response to Platinum, Taxane, and Combination Therapy in Ovarian Cancer: In vitro Phenotypes, Inherited Variation, and Disease Recurrence[J]. *Front Genet*, 2016,7:37.
- [29] Nagy R, Boutin T S, Marten J, et al. Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants[J]. *Genome Med*, 2017,9(1):23.
- [30] Zeng J H, Liang L, He R Q, et al. Comprehensive investigation of a novel differentially expressed lncRNA expression profile signature to assess the survival of patients with colorectal adenocarcinoma[J]. *Oncotarget*, 2017,8(10):16811-16828.
- [31] Shi X, Tan H, Le X, et al. An expression signature model to predict lung adenocarcinoma-specific survival[J]. *Cancer Manag Res*, 2018,10:3717-3732.

- [32] Guan C, Ouyang D, Qiao Y, et al. CA9 transcriptional expression determines prognosis and tumour grade in tongue squamous cell carcinoma patients[J]. *J Cell Mol Med*, 2020,24(10):5832-5841.
- [33] Chien M H, Yang J S, Chu Y H, et al. Impacts of CA9 gene polymorphisms and environmental factors on oral-cancer susceptibility and clinicopathologic characteristics in Taiwan[J]. *PLoS One*, 2012,7(12):e51051.
- [34] Zapata L, Pich O, Serrano L, et al. Negative selection in tumor genome evolution acts on essential cellular functions and the immunopeptidome[J]. *Genome Biol*, 2018,19(1):67.
- [35] Wahid B, Bashir H, Bilal M, et al. Developing a deeper insight into reproductive biomarkers[J]. *Clin Exp Reprod Med*, 2017,44(4):159-170.
- [36] Zhao B S, Roundtree I A, He C. Post-transcriptional gene regulation by mRNA modifications[J]. *Nat Rev Mol Cell Biol*, 2017,18(1):31-42.
- [37] Roundtree I A, Evans M E, Pan T, et al. Dynamic RNA Modifications in Gene Expression Regulation[J]. *Cell*, 2017,169(7):1187-1200.
- [38] Zhang S, Zhao B S, Zhou A, et al. m(6)A Demethylase ALKBH5 Maintains Tumorigenicity of Glioblastoma Stem-like Cells by Sustaining FOXM1 Expression and Cell Proliferation Program[J]. *Cancer Cell*, 2017,31(4):591-606.
- [39] Li Z, Weng H, Su R, et al. FTO Plays an Oncogenic Role in Acute Myeloid Leukemia as a N(6)-Methyladenosine RNA Demethylase[J]. *Cancer Cell*, 2017,31(1):127-141.
- [40] Zuo X, Chen Z, Gao W, et al. M6A-mediated upregulation of LINC00958 increases lipogenesis and acts as a nanotherapeutic target in hepatocellular carcinoma[J]. *J Hematol Oncol*, 2020,13(1):5.
- [41] Yang X, Zhang S, He C, et al. METTL14 suppresses proliferation and metastasis of colorectal cancer by down-regulating oncogenic long non-coding RNA XIST[J]. *Mol Cancer*, 2020,19(1):46.
- [42] Zhu S, Wang J Z, Chen, et al. An oncopeptide regulates m(6)A recognition by the m(6)A reader IGF2BP1 and tumorigenesis[J]. *Nat Commun*, 2020,11(1):1685.
- [43] Ni W, Yao S, Zhou Y, et al. Long noncoding RNA GAS5 inhibits progression of colorectal cancer by interacting with and triggering YAP phosphorylation and degradation and is negatively regulated by the m(6)A reader YTHDF3[J]. *Mol Cancer*, 2019,18(1):143.
- [44] Wu Y, Yang X, Chen Z, et al. m(6)A-induced lncRNA RP11 triggers the dissemination of colorectal cancer cells via upregulation of Zeb1[J]. *Mol Cancer*, 2019,18(1):87.

Figures

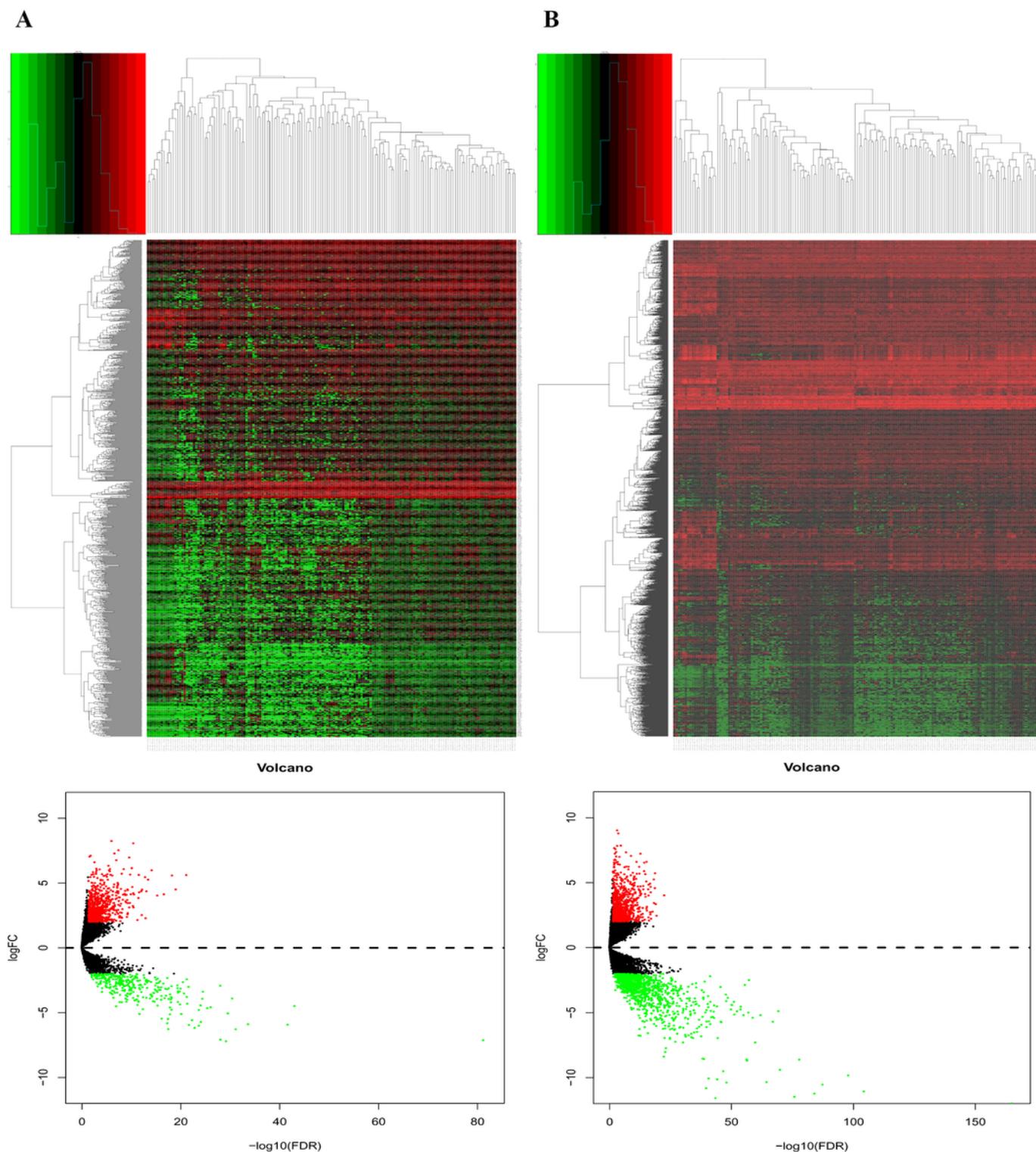
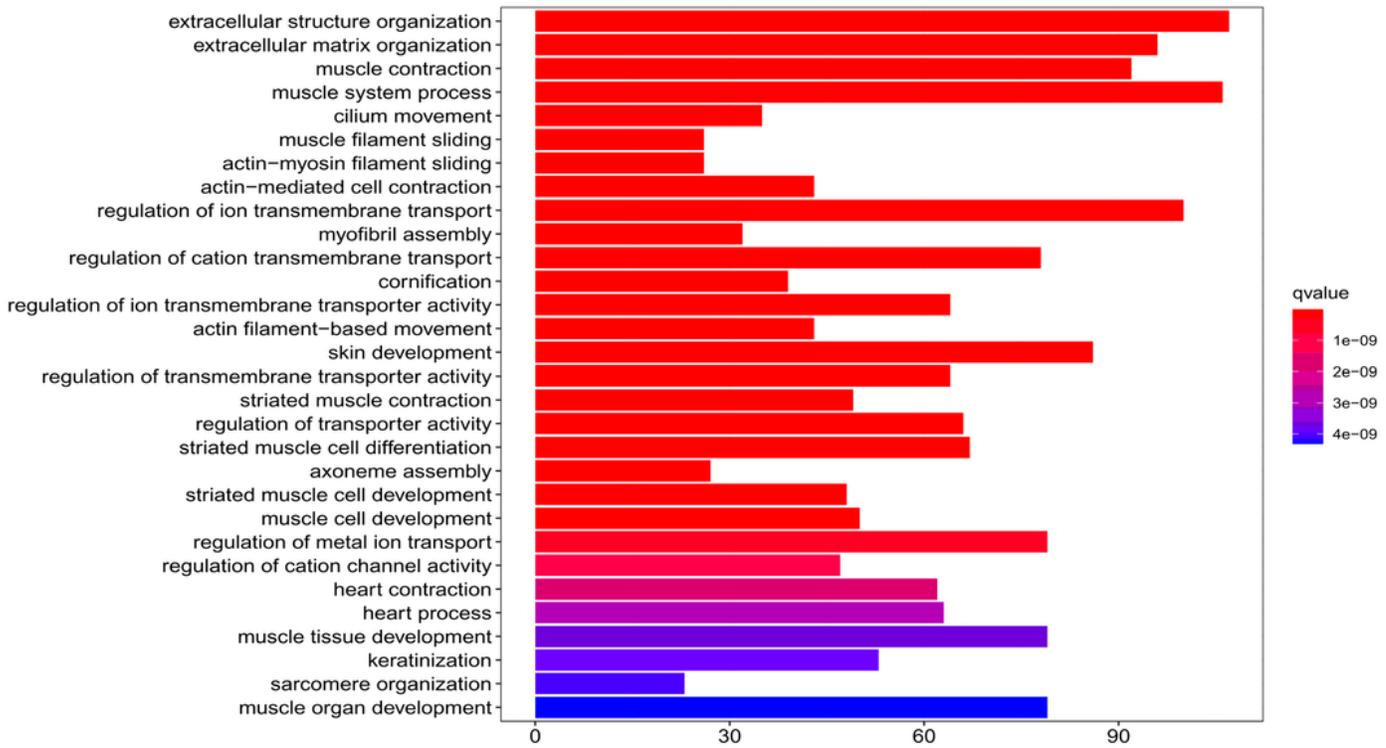


Figure 1

Heatmap of 1,001 DElncRNAs (A) and 2,006 DEmRNAs (B) ($|\log_{2}\text{FC}| > 2.0$ and adjusted $p < 0.05$). Red indicates upregulation, green indicates downregulation, and black indicates normal expression. Each column represents a sample, and each row represents a differentially expressed gene. Similar to that in the heatmap, red indicates upregulation, green indicates downregulation, and black indicates normal expression in the volcano plot. Each point represents a gene.

C



D

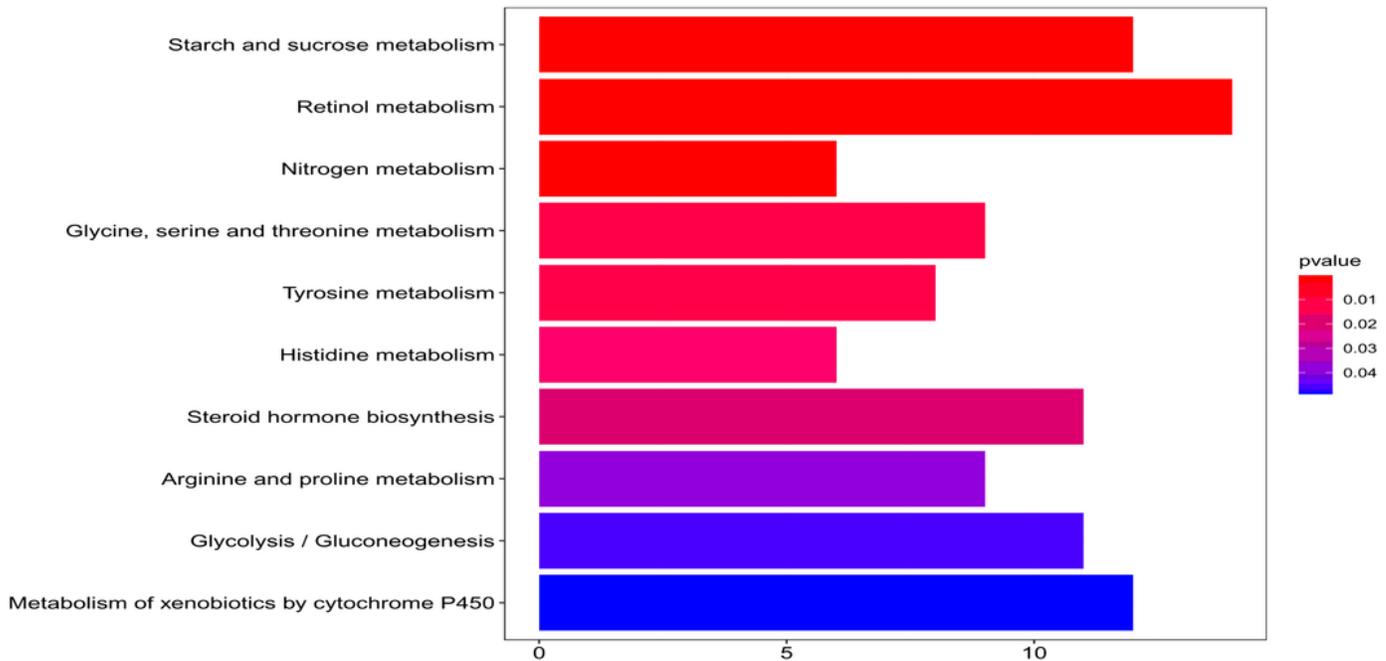


Figure 2

GO and KEGG pathway analysis of the DEmRNAs. (C) The numbers of genes enriched in each GO category. The y axis represents the GO categories including the biological process, cellular component and molecular function categories, and the x axis represents the enrichment score. Furthermore, the color represents the q-value. (D) The top 10 pathways enriched in the DEmRNAs. The y axis represents the pathways, the x axis represents enriched gene numbers, and the color indicates the p-value.

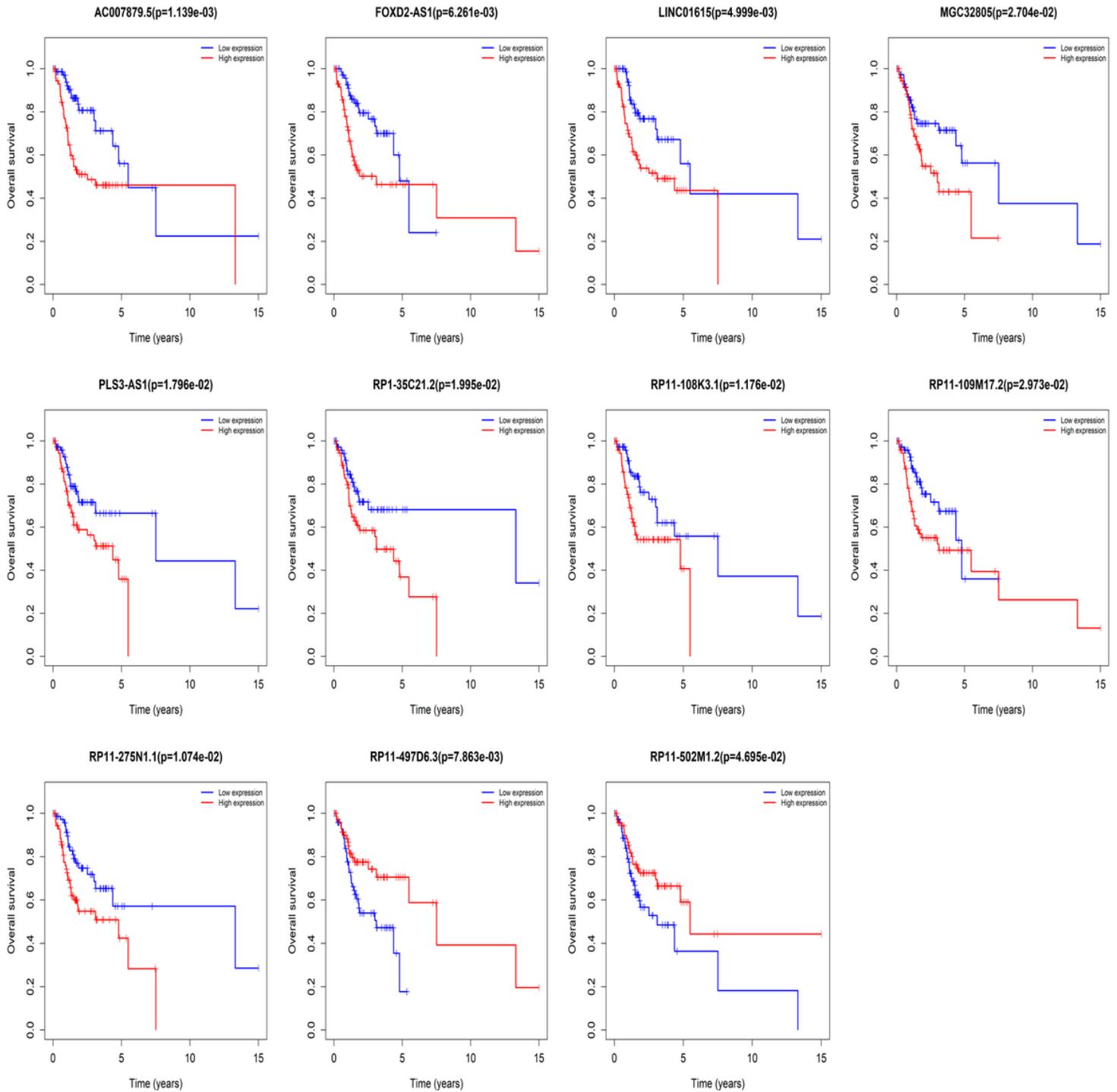


Figure 3

The significant prognostic value of DElncRNAs, including AC007879.5, FOXD2-AS1, LINC01615, MGC32805, PLS3-AS1, RP1-35C21.2, RP11-108K3.1, RP11-109M17.2, RP11-275N1.1, RP11-497D6.3, and RP11-502M1.2.

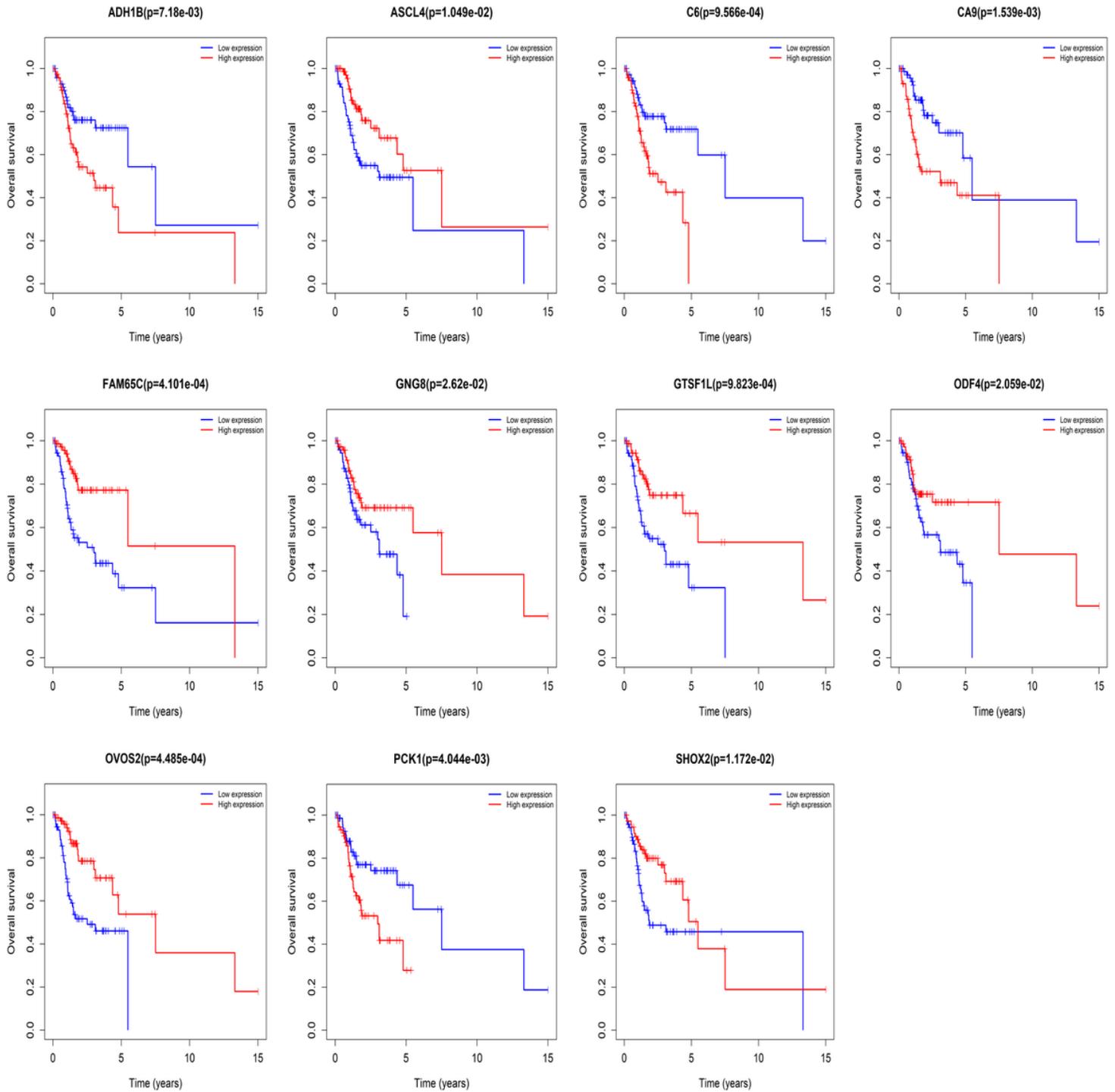


Figure 4

The significant prognostic value of DE mRNAs, including ADH1B, ASCL4, C6, CA9, FAM65C, GNG8, GTSF1L, ODF4, OVOS2, PCK1, and SHOX2.

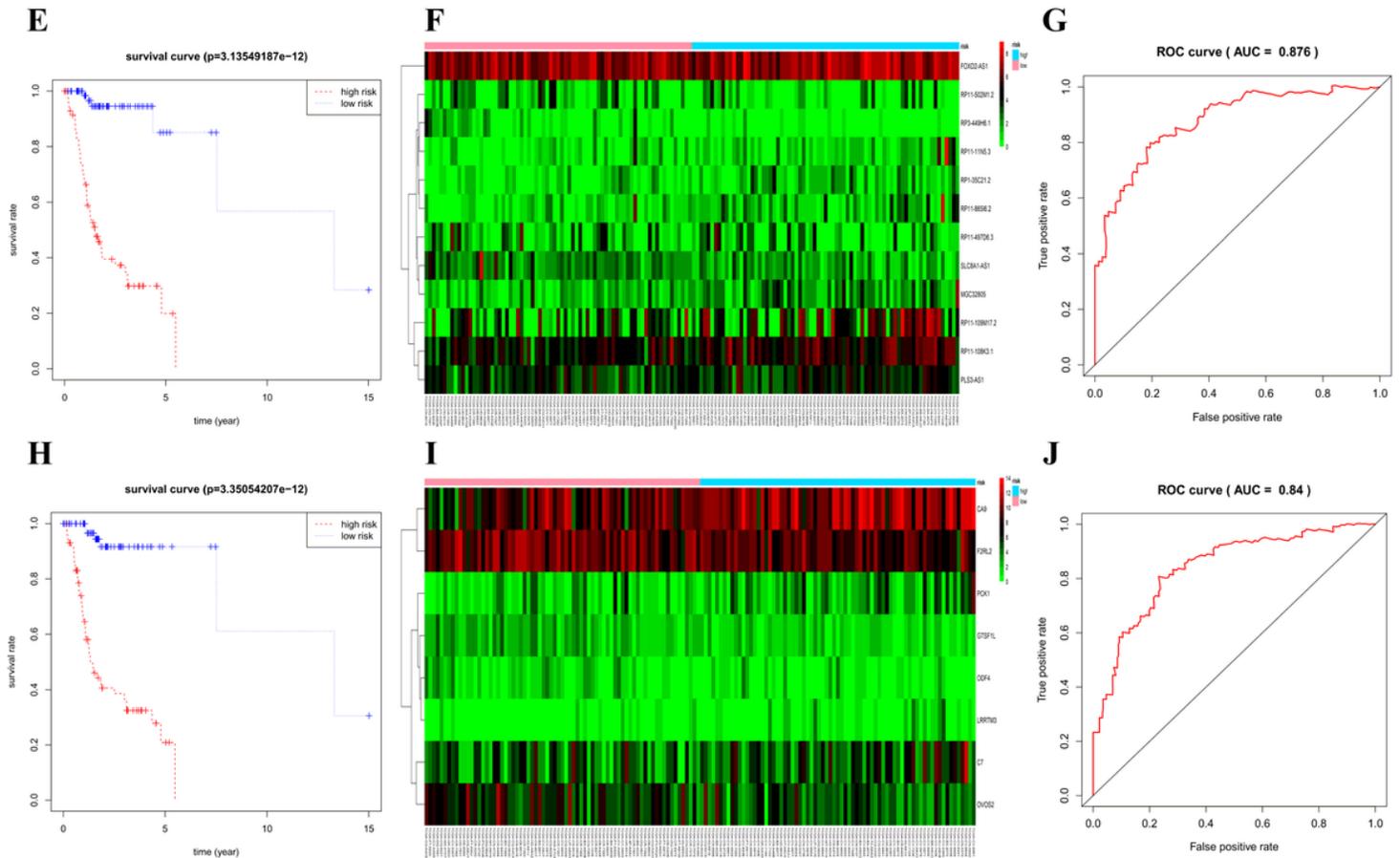


Figure 5

The survival rate of the low-risk patients was significantly higher than that of the high-risk patients (E). The heatmap shows the relationships between the expression of 12 candidate DElncRNAs and the risk score of each patient (F). The 5-year survival correlation of the 12 DElncRNAs was assessed by ROC analysis, and AUCs were computed to assess the discriminatory capacity of the lncRNA signature (G). Similar to that for DElncRNAs, risk score analysis (H) of eight candidate DEmRNAs was performed for each patient and a heatmap was constructed (I). Subsequently, the ROC curve was plotted, and the AUC was 0.84 (J).

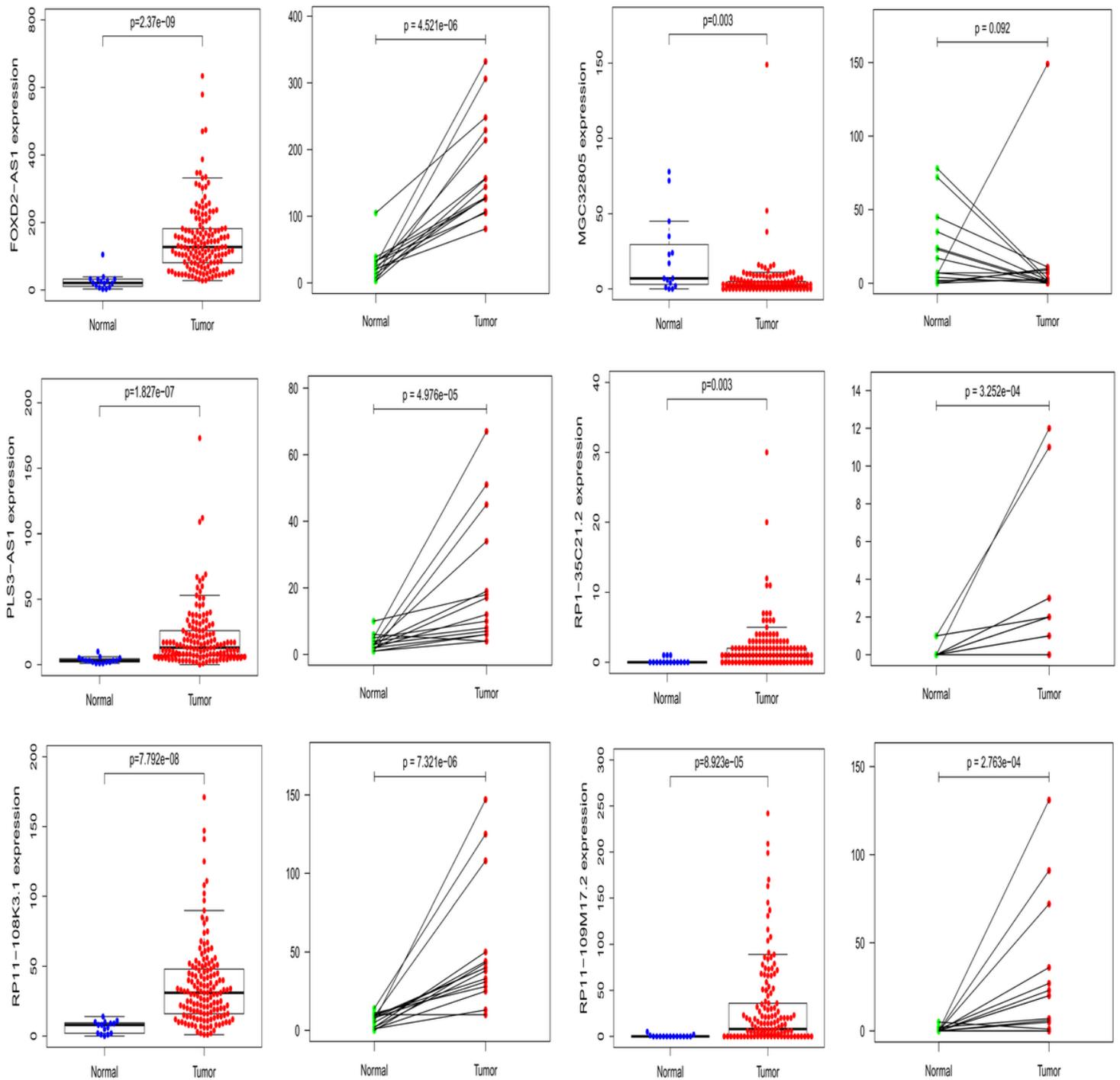


Figure 6

The expression of FOXD2-AS1, MGC32805, PLS3-AS1, RP1-35C21.2, RP11-108K3.1 and RP11-109M17.2 was analyzed in TSCC and normal tissues and in the paired groups.

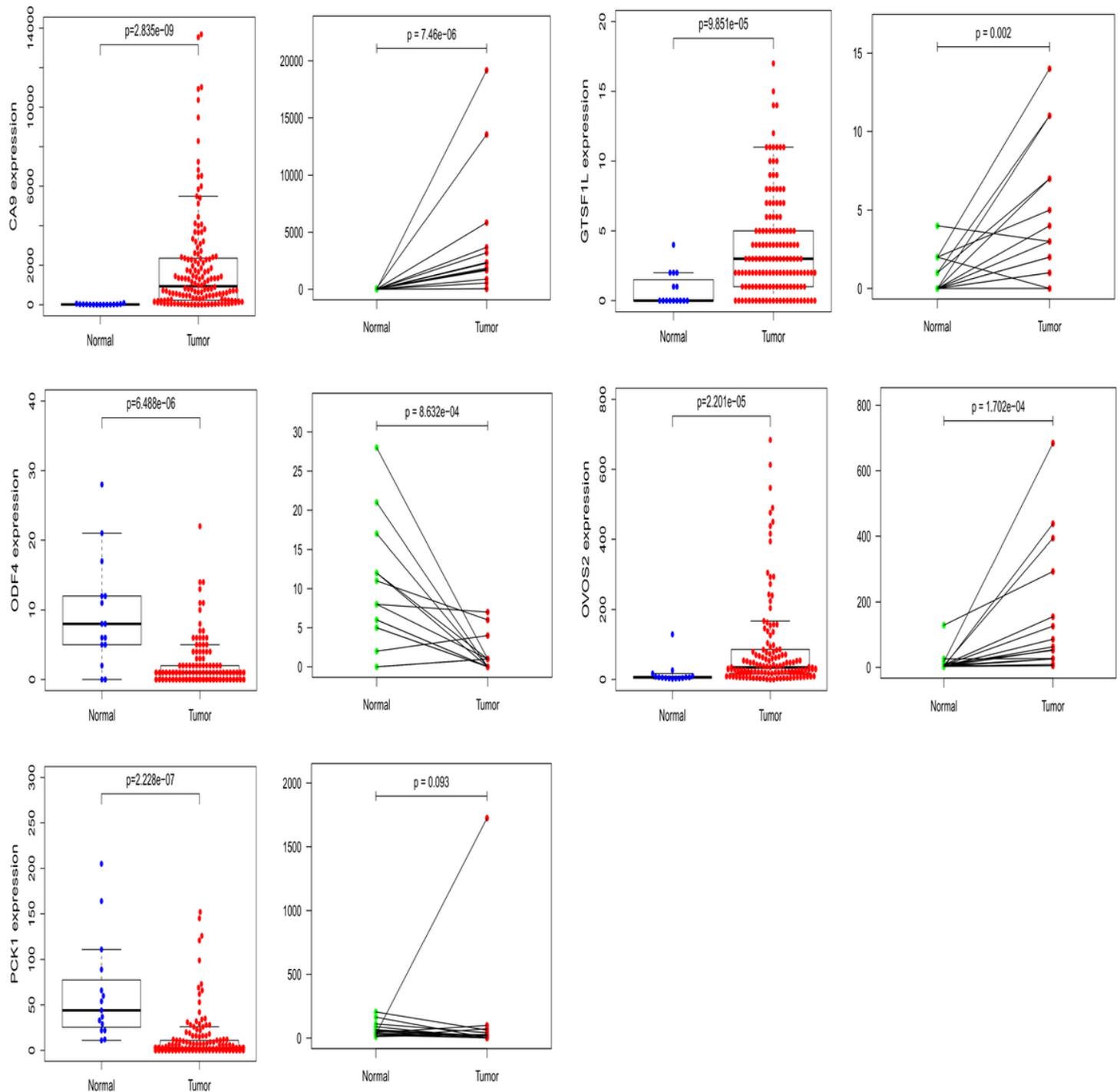


Figure 7

The expression of CA9, GTSF1L, ODF4, OVOS2, and PCK1 was analyzed in TSCC and normal tissues and in the paired groups.

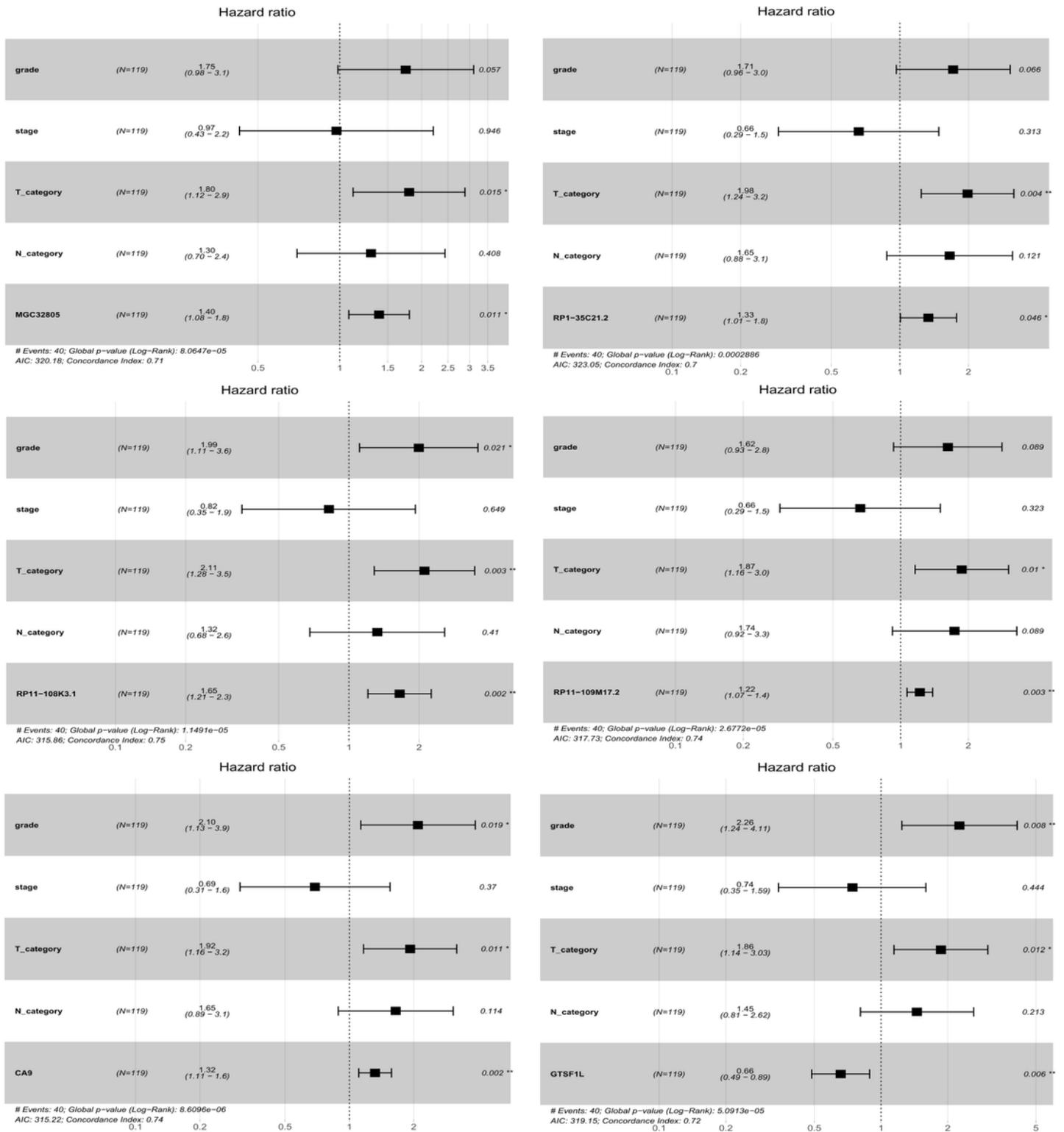


Figure 8

Forest map of MGC32805, RP1-35C21.2, RP11-108K3.1, RP11-109M17.2, CA9, and GTSF1L based on the risk score model.

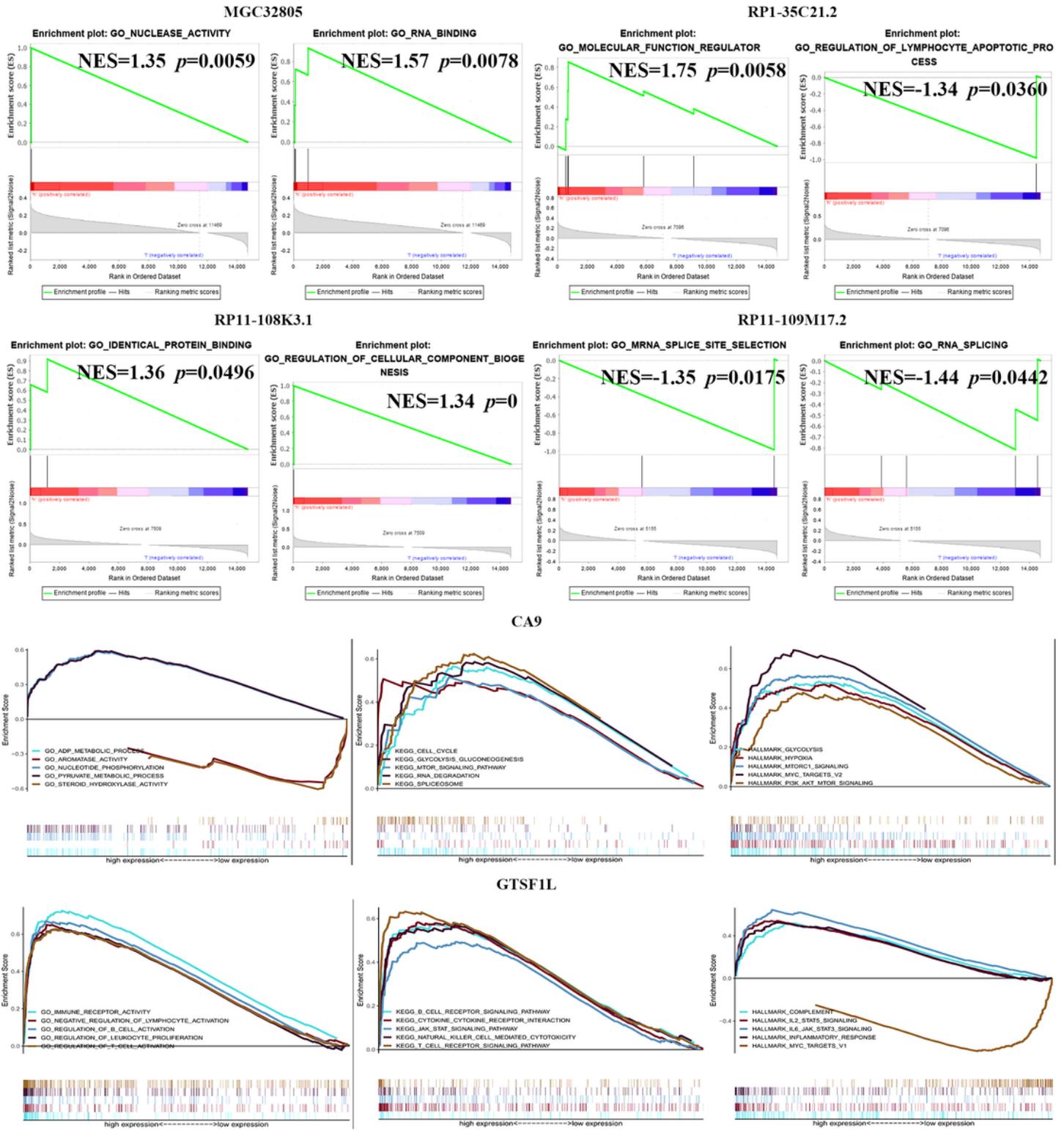


Figure 9

The GSEA results of MGC32805, RP1-35C21.2, RP11-108K3.1, RP11-109M17.2, CA9, and GTSF1L.

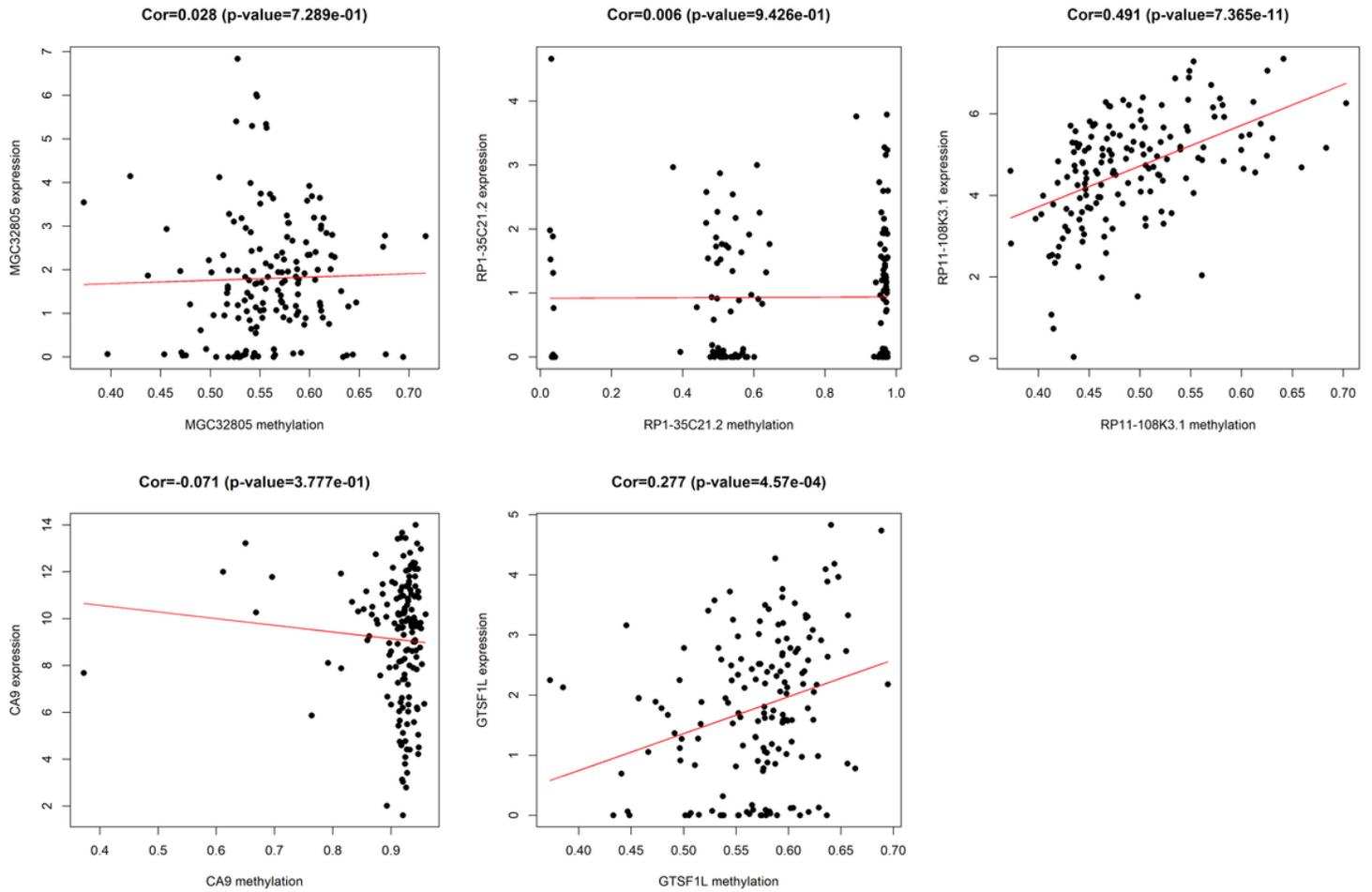


Figure 10

The correlations between the expression of MGC32805, RP1-35C21.2, RP11-109M17.2, CA9, and GTSF1L and their methylation levels.

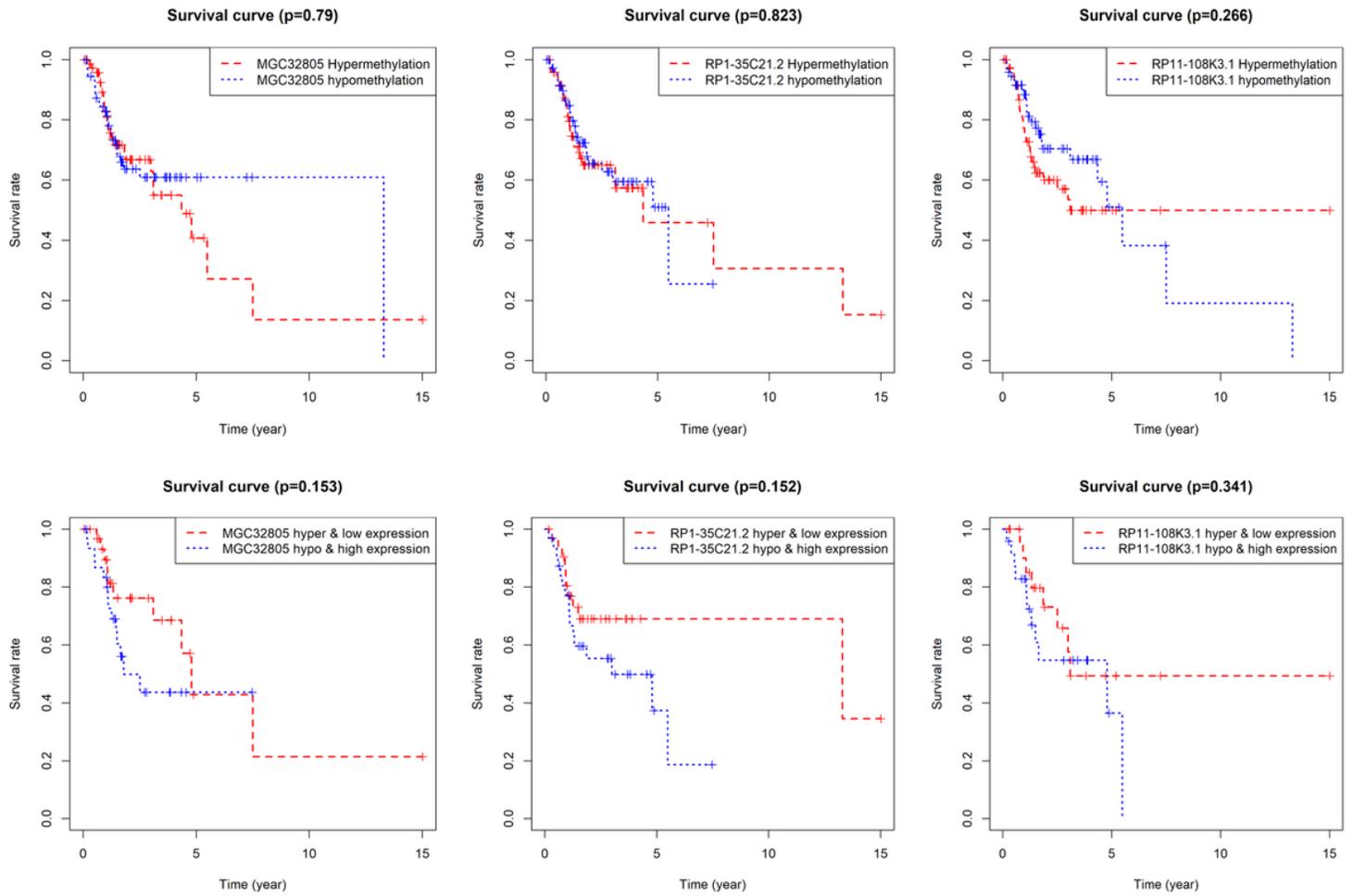


Figure 11

The association of MGC32805, RP1-35C21.2, and RP11-109M17.2 with OS by combining the potential biomarker's expression and methylation.

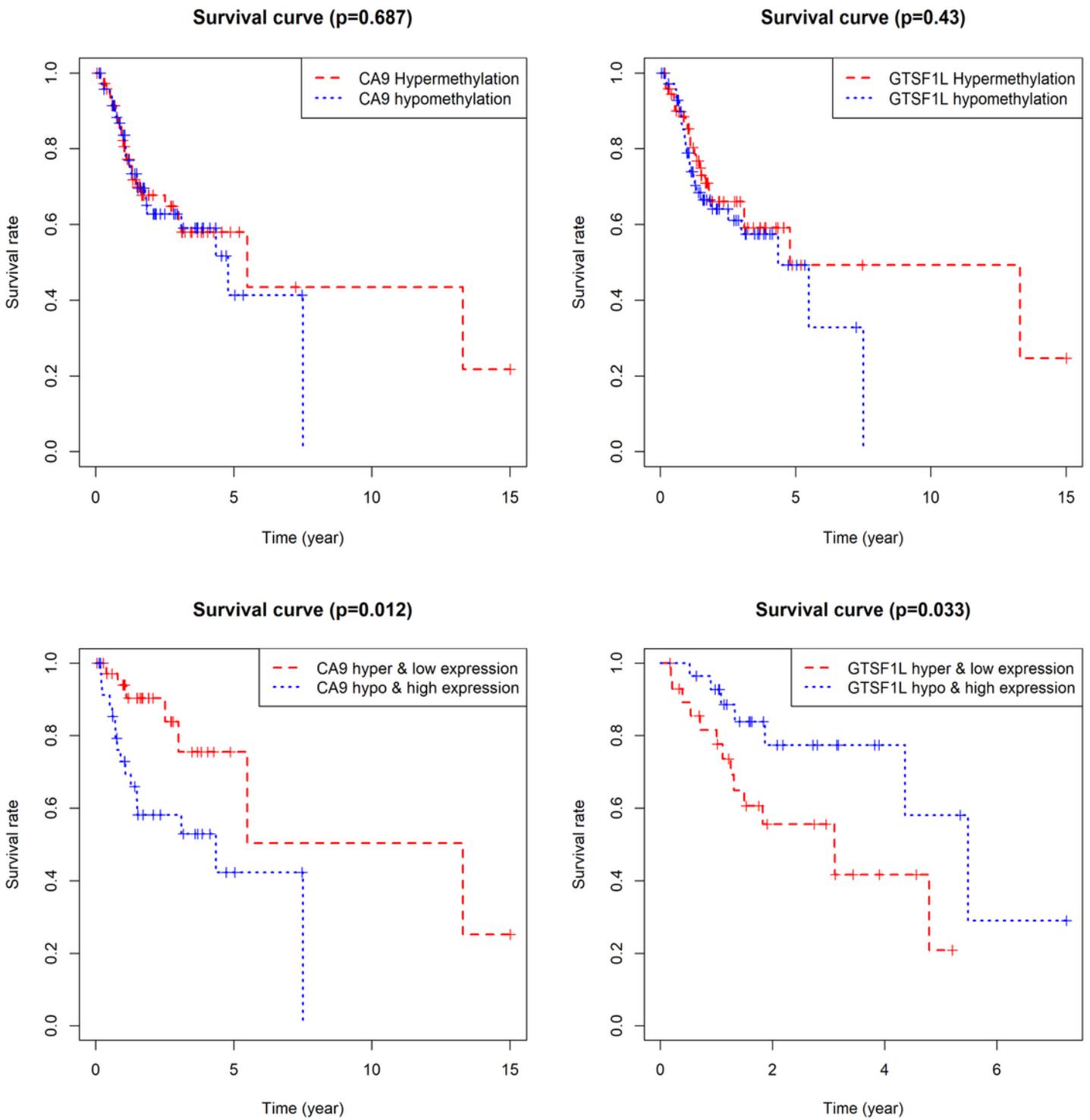
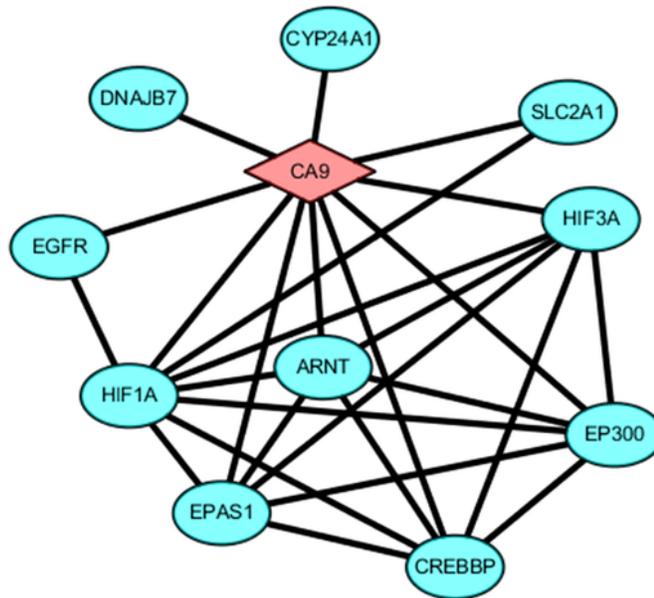


Figure 12

The association of CA9 and GTSF1L with OS by combining the potential biomarker's expression and methylation.

K



L

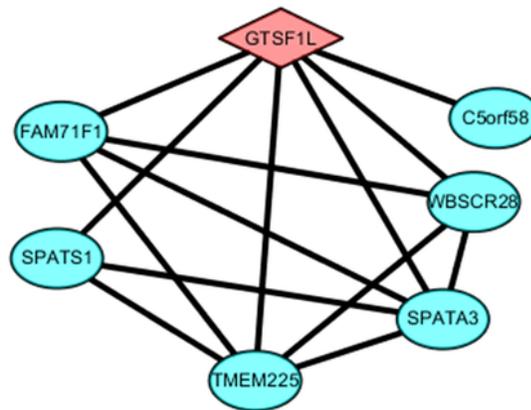
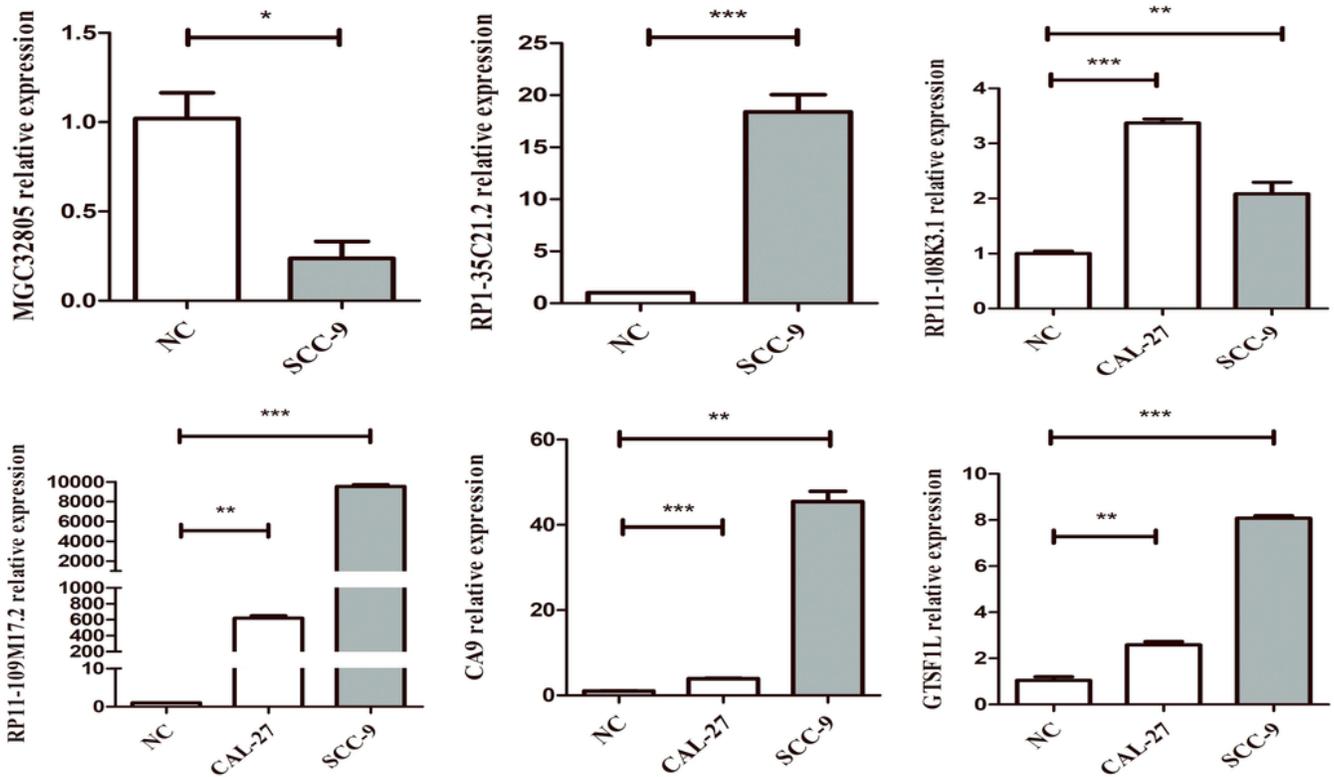


Figure 13

With the thresholds of confidence score > 0.7, the protein network interacting with CA9 (K) and GTSF1L (L).

M



N

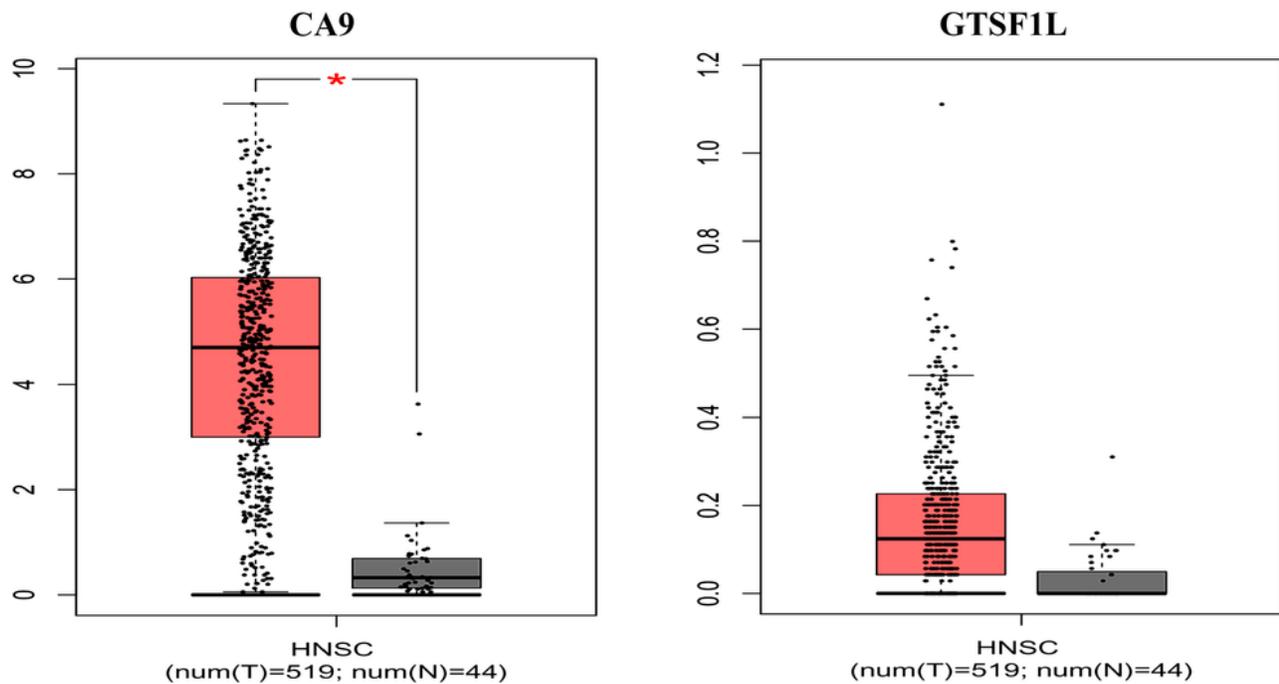


Figure 14

qRT-PCR validation of six biomarkers from sequencing (M). The expression of CA9 and GTSF1L in GEPIA (N) (HNSC, head and neck squamous cell carcinoma. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigS1.tif](#)
- [FigS2.tif](#)
- [FigS3.tif](#)
- [SupplementaryMaterials.2020.10.22.docx](#)