Identification of Candidate Biomarkers and Analysis of Prognostic Values in Oral Squamous Cell Carcinoma

Guang-zhao Huang, Qing-qing Wu, Ze-nan Zheng, Ting-ru Shao and Xiao-Zhi Lv*

Department of Oral and Maxillofacial Surgery, NanFang Hospital, Southern Medical University, Guangzhou, China

Objectives: Oral squamous cell carcinoma (OSCC) is the most common oral cancer with a poor prognosis owing to limited understanding of the disease mechanisms. The aim of this study was to explore and identify the potential biomarkers in OSCC by integrated bioinformatics analysis.

Materials and Methods: Expression profiles of long non-coding RNAs (lncRNAs), microRNAs (miRNAs), and messenger RNAs (mRNAs) were downloaded from The Cancer Genome Atlas (TCGA) and differentially expressed RNAs (DERNAs) were subsequently identified in OSCC by bioinformatics analysis. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were used to analyze DERNAs. Then, the competing endogenous RNA (ceRNA) network was constructed in Cytoscape and the protein-protein interaction (PPI) network was established in the STRING database. We established a risk model to predict the overall survival of OSCC on the basis of DElncRNAs with Kaplan–Meier analysis and combined with logrank p test. Furthermore, we identified potential biomarkers by combining univariate Cox regression with overall survival rate, which were then validated in Gene Expression Omnibus (GEO), OSCC cell lines and OSCC specimens.

Results: A total of 1,919 DEMRNAs, 286 DElncRNAs and 111 DEMiRNAs were found to be dysregulated in OSCC. A ceRNA network included 46 DElncRNAs, 7 DEMiRNAs and 10 DEMRNAs, and the PPI network included 712 DEMRNAs including 31 hub genes. Moreover, a 7 lncRNAs risk model was established and four genes (CMA1, GNA14, HCG22, HOTTIP) were identified as biomarkers on overall survival in patients with OSCC.

Conclusions: This study successfully constructed a ceRNA network and a PPI network which play a crucial role in OSCC. A risk model was established to predict the prognosis, and four DERNAs are revealed with overall survival in patients with OSCC, suggesting that they may be potential biomarkers in tumor diagnosis and treatment.

Keywords: competing endogenous RNA, protein-protein interaction, long non-coding RNA, biomarker, oral squamous cell carcinoma
INTRODUCTION

Oral squamous cell carcinoma (OSCC) is one of the most common oral cancers worldwide (1) and has the most severe impact on the quality of life of patients. The most significant risk factors in OSCC are cigarettes, alcohol and betel nut consumption, which seem to have a synergistic effect. One recent study showed that it may be associated with the infection of human papillomavirus (HPV) (2). Although medical equipment and treatment methods are more advanced in recently years, the 5-year overall survival rate of OSCC remains only 40–50% (3) on account of relatively low responsiveness to treatment, severe drug resistance (4), diagnosis at terminal stage, etc., whereas the 5-year overall survival rate can rise markedly to more than 85% in patients diagnosed with early-staged tumors (5). Therefore, early diagnosis is very important for improving the prognosis of patients with OSCC.

As is known, more than 70% of the human genome can be transcribed to functional products (6). Among them, long non-coding RNAs (lncRNAs), RNA transcripts longer than 200 nucleotides and those with limited protein-coding potential are in the majority. lncRNAs affect various aspects of cellular homeostasis in OSCC, including proliferation, survival, migration, and genomic stability (7). Moreover, cancer-specific lncRNA expression patterns appear more tissue- and stage-specific than those of protein-coding genes, indicating the potential development of lncRNAs as powerful alternative biomarkers and therapeutic targets (8). Meanwhile, recent research demonstrated that lncRNA combined with mRNA biomarkers may improve diagnosis (9). However, there are fewer effective lncRNAs and mRNAs biomarkers used for diagnosis in OSCC.

In 2011, Salmena et al. put forward a competing endogenous RNA (ceRNA) hypothesis (10). MiRNA as a regulatory molecule regulates mRNA expression by targeting the 3'-UTR (11), which typically inhibits the translation and the stability of mRNAs (12). However, lncRNAs can compete with the miRNA binding to the mRNA to regulate gene expression. Recently, increasing evidence indicated that this hypothesis was closely related to the development and initiation of OSCC. For example, IncRNA TUG1 through sponging miR-524-5p to mediate DLX1 expression promotes proliferative and migratory potentials in OSCC (13). In addition, protein-protein interaction (PPI) also plays a crucial role in various cancers. PPI is composed of proteins interacting with each other to participate in various biological processes such as biological signal transmission, regulation of gene expression, energy metabolism, and cell cycle regulation. For example, c-Myc, MMP-2, and GSK3β regulated the expression of MMP9 to accelerate OSCC progression and invasion (14). However, fewer studies have been reported on the ceRNA and PPI networks in OSCC.

In this study, we aim to analyze the differentially expressed profile of non-coding RNAs (ncRNAs) in the OSCC and establish a Cox regression model to predict the overall survival of patients with OSCC. Further, we analyzed and predicted the functions of the ceRNA and PPI networks. This study will contribute to understanding the molecular mechanism and provide new therapeutic targets for OSCC.

MATERIALS AND METHODS

Data Preparation

All primitive data of OSCC (oral cavity, floor of mouth, palate, buccal mucosa, the anterior 2/3 of the tongue, gingiva, and so on) from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/) were downloaded through GDC Data Transfer Tool, including the RNA-Seq and miRNA-Seq of Transcriptome Profiling and Clinical data. Then, we excluded three samples because of their low-quality clinical data. Finally, 316 OSCC samples and 32 normal control samples were collected in our study. Gene expression profiles of OSCC were obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/), including GSE9844 (26 OSCC samples and 12 normal controls), GSE30784 (167 OSCC samples and 45 normal controls), and GSE74530 (6 OSCC samples and 6 normal controls).

Collection of OSCC Specimens

A total of 49 pairs of OSCC specimens and normal adjacent tissues (about more than 1.5 cm from the edge of the tumor) were collected at Nanfang Hospital, Southern Medical University (Guangzhou, China), and written informed consent was obtained from all patients. The tissue specimens were stored in RNA Wait and then frozen at −80°C. All tumor tissues and normal adjacent tissues were pathologically confirmed as squamous cell carcinoma and normal tissues, respectively.

Differentially Expressed Gene Analysis

EdgeR (Version 3.8) package in R software was used to analyze and identify differentially expressed RNAs (DERNAs) and differentially expressed miRNAs (DEmiRNAs) with the thresholds of \log2 (fold change [FC])\geq2.0 and FDR (adjusted \textit{P}-value) <0.01. Then, we used the annotation file in GTF format (Homo_sapiens.GRCh38.95.chr.gtf) to identify and annotate differentially expressed long non-coding RNA (DElncRNAs) with the thresholds of \log2[FC]\geq2.0 and FDR<0.01. The heatmap and volcano were constructed by the gplots package in R software.

Abbreviations: OSCC, oral squamous cell carcinoma; TCGA, The Cancer Genome Atlas; DERNAs, differentially expressed RNAs; DEMiRNAs, differentially expressed miRNAs; DElncRNAs, differentially expressed long non-coding RNAs; GEO, Gene Expression Omnibus; ceRNA, competing endogenous RNA; PPI, protein-protein interaction; DMEM, Dulbecco’s modified Eagle’s medium; FBS, fetal bovine serum; MREs, miRNA response elements; ROC, receiver operating characteristic; GO, Gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; ncRNAs, non-coding RNAs. 3'-UTR, 3'-Untranslated region.
Functional Enrichment Analysis
We employed DAVID (https://david.ncifcrf.gov/) to obtain information for Gene Ontology (GO) including biological processes, the cellular component and molecular function. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was used to annotate the potential functions. A significance level of \( P < 0.05 \) was set as the cutoff criteria and the plots were constructed by the gplots package in R software.

Protein-Protein Interaction Analysis
The DEMRNAs were enrolled in a protein-protein interaction (PPI) network through the STRING database (https://string-db.org/) with a confidence score \( \geq 0.9 \), and the PPI network was visualized in Cytoscape (Version 3.7.1) software. Moreover, genes with degree \( \geq 25 \) were selected as hub genes. Subsequently, module analysis (16) of the PPI network was performed using the Molecular Detection (MCODE) tool of Cytoscape software, and GO and KEGG analysis of the modules was carried out using the DAVID database.

Construction of the ceRNA Network
According to the hypothesis of ceRNA, a lncRNA-miRNA-mRNA network was constructed. Relevant miRNA-target data were obtained from the miCode database (http://www.mircode.org/) (17). Then, the DElncRNA-DEmiRNA interactions were predicted according to the miCode database. Furthermore, target DEMRNAs were predicted for DEMiRNAs using miRDB (http://www.mirdb.org/) (18), miRTarBase(http://mirtarbase.mbc.nctu.edu.tw/) (19) and TargetScan database (http://www.targetscan.org/) (20), and only the miRNA-mRNA interactions that existed in all the three databases were enrolled in the ceRNA network. Eventually, Cytoscape (Version 3.7.1) was employed to establish the lncRNA-miRNA-mRNA network.

Cox Risk Regression Establishment and Validation
The lncRNAs raw data were transformed and normalized in a log2\((x+1)\) manner (21). OSCC samples were randomly divided into a training set and a validation set. Univariate Cox regression was used to select prognosis-associated genes \( p < 0.05 \). Subsequently, we performed Cox regression analysis combined with LASSO to establish a prognostic risk score model, and the penalty regularization parameter lambda \( (\lambda) \) was chosen through cross-validation with an \( n \)-fold equal to 10 by using the R package glmnet (21). Lambda.min was identified to pick out the variables. According to these variables, a stepwise regression was performed to establish the Cox model. Finally, a validation set and Kaplan–Meier survival curves along with a logrank \( p \) test were applied to validate its accuracy. In addition, receiver operating characteristic (ROC) analysis was used to estimate the predictive power of this signature.

Cell Culture
The human OSCC cell lines SCC9, SCC15, SCC25, CAL27, and KB and the normal oral epithelial cell line HOK were obtained from the Institute of Antibody Engineering, Southern Medical University (Guangzhou, China). Cells HOK, SCC9, SCC15, and SCC25 were cultured in Dulbecco's modified Eagle's medium F12 (DMEM/F12) (Invitrogen, Carlsbad, CA, USA), CAL27 in \( \alpha \)-MEM (Invitrogen, Carlsbad, CA, USA) and KB in RPMI 1640 (Invitrogen, Carlsbad, CA, USA), supplemented with 10% fetal bovine serum (FBS) at 37 \(^\circ\)C with 5% CO2.

RNA Extraction and qRT-PCR
Total RNA was extracted from tissues and cells in a TRIzol reagent (Takara) manner. RNA reverse transcription to cDNA was performed with a Reverse Transcription Kit (Vazyme). Quantitative real-time Polymerase Chain Reaction (qRT-PCR) analyses used SYBR Green I (Vazyme) in triplicate. The results were normalized to the expression of GAPDH. The primer sequences are as follows. HCG22 forward primer (5'-3'): CTCTCTGTGCTCCGTCCCT; reverse primer (5'-3'): ACTCCACTCTTCAGTGCTACC. HOTTIP forward primer (5'-3'): CCTAAAAGCCTCAGCTTCTGG; reverse primer (5'-3'): TGCAGGCTGAGATCCTACT. GNA14 forward primer (5'-3'): CCCA ACAAGATGTGCTTCGC; reverse primer (5'-3'): TCCGTCTTTCCGATCTGG. CMA1 forward primer (5'-3'): TCAAGCTGTGTGTGCGGCAATC; reverse primer (5'-3'): CTTTGCAATCCG ACCGTCCAT. GAPDH forward primer (5'-3'): CGCTGATGATCTTGAGGCTGTTG; GAPDH reverse primer: (5'-3'): GCTGATGATCTTGAGGCTGTTG.

Western Blot Analysis
Cells and OSCC tissues were lysed in RIPA lysis buffer. According to the procedure, proteins were separated by electrophoresis, transferred to membranes and then sealed with 5% skim milk. The primary antibodies including CMA1 (dilution 1:1,000), GNA14 (dilution 1:1,000) and \( \alpha \)-tubulin (dilution 1:1,000) were incubated in 4\(^\circ\)C for one night. Subsequently, goat anti-mouse and goat anti-rabbit secondary antibodies were incubated for 1h at room temperature, and finally immunoreactive bands were visualized with a chemiluminescence system.

Biomarkers Screening and Validation
The status and survival time of OSCC patients were extracted. Subsequently, the mRNA was enrolled in the PPI and ceRNA networks, and lncRNA and miRNA identified in ceRNA were selected for screening biomarkers. We used univariate Cox regression to screen prognostic factors \( (p < 0.05) \), as well as significant factors whose expression levels were significantly relevant to patients’ overall survival \( (p < 0.01) \) were selected as primitive biomarkers. Ultimately, the gene expression profile from the GEO and OSCC cell lines and tissues were used to verify these biomarkers. In addition, a combination of two or more biomarkers was performed to predict OSCC overall survival according to the gene expression in TCGA.
Statistical Analysis
Statistical analyses were performed using SPSS23.0 software (IBM). The differences between groups were tested using a two-tailed Student’s t-test. The survival analysis between two groups was conducted by logrank test. P-values < 0.05 were considered statistically significant. Differences were considered significant if *p < 0.05; **p < 0.01; ***p < 0.001; or ****p < 0.0001.

RESULTS
Identification of DEmRNA, DEmiRNAs, and DEIncRNAs
RNA expression profiles and corresponding clinical data of 316 OSCC patients and 32 normal controls were downloaded from TCGA database. With the cut-off criteria unified, |log2FoldChange| > 2 and FDR < 0.01, 1919 DEmRNAs (673 upregulated and 1246 downregulated, Figure 1A), 286 DEIncRNAs (192 upregulated and 94 downregulated, Figure 1B) and 111 DEmiRNAs (61 upregulated and 50 downregulated, Figure 1C) were sorted out.

Functional Analysis of DERNAs
GO and KEGG enrichment analysis were used to explore the potential function of DERNAs. The results indicated that these genes were mainly enriched in tissue development, cell differentiation and calcium binding (Figures 2A, B). Moreover, the majority of genes were located in the extracellular region. In addition,
KEGG pathway analyses demonstrate that the most significant pathways were the calcium signaling, protein digestion and absorption and focal adhesion pathways (Figures 2C,D).

**Protein-Protein Interaction (PPI) Network Analysis**

A total of 712 proteins and 3,181 edges were selected in the PPI network. The confidence score is >0.9. A total of 31 hub genes...
were selected from PPI network with degree \( \geq 25 \) (Figure 3A). Among them, 6 mRNAs including GNA14, GRM5, KRT3, KRT26, TACR1, and HTR2C were related to overall survival. In addition, according to module analysis, three modules were identified in the PPI network (Figures 3B–E). Surprisingly, most hub genes, including all of these associated with overall survival (Supplementary Figure 1), were enriched in module 2 indicating that module 2 plays a vital role in the PPI network. According to the GO terms analysis, three modules related to cell adhesion, tissue development, and protein binding played a crucial role in cancer. With respect to KEGG enrichment analysis, modules 1 and 2 regulated metabolic pathways such as protein digestion and absorption (Tables 1, 2). Moreover, modules 1 and 3 were significantly relevant to the PI3K-Akt signaling and calcium signaling pathways, which were associated with the occurrence and development of tumors (Table 3).

Construction of ceRNA Network in OSCC

A total of 46 lncRNAs, 10 mRNAs and 7 miRNAs were enrolled in the ceRNA network (Figure 4A). We employed miRcode to predict the potential miRNA target by DElncRNAs. As a result, 46 of 286 DElncRNAs and 7 of 111 DEMiRNAs formed 119 lncRNA-miRNA pairs (Table 4). Then, miRDB, miranda and TargetScan were used to predict the miRNA-mRNA pairs. Only the miRNA-mRNA interactions that exist in all three databases were brought into the ceRNA network (Table 5). Finally, there were 10 DEMiRNAs could target to the 7 miRNAs (Figure 4B). Therefore, according to Table 6, the ceRNA network including 46 lncRNAs, 10 mRNAs and 7 miRNAs was completely constructed, and the lncRNA-miRNA-mRNA network was visualized in Cytoscape (Version 3.7.1).

Establishment and Validation of Cox Regression Model

A total of 316 OSCC samples were randomly divided into a training set and validation set. Subsequently, combined univariate Cox regression with a LASSO Cox regression model along with 10-fold cross-validation, 11 variables (AC073130.1, AFAP1-AS1, AQP4-AS1, C11orf97, HOTTIP, HOXA11-AS, LINC00460, LINC01234, LINC01391, SLC8A1-AS1, and SRGAP3-AS2) were identified (Figures 5A,B). Furthermore, a stepwise regression was performed

FIGURE 3 | Protein-protein network (PPI) based on the DEmRNAs with a combined score was \( > 0.90 \). (A) 31 hug genes were listed because the degree was \( > 25 \) in the PPI network. (B) Module analysis of DEmRNA enrolled in PPI network with the criterion cut. Height = 0.8, min size = 10. Same color means it belongs to the same module. And 3 modules were visualized in Cytoscape (C–E). The connection between the nodes indicates the potential interaction between different mRNA, and red represents the hub gene in the PPI network. Meanwhile, GO and KEGG analysis of 3 modules was performed in DAVID (Tables 1–3).
### TABLE 1 | Module 1 GO and KEGG pathways analysis.

| ID        | GO-term                                         | Count | p-value     |
|-----------|------------------------------------------------|-------|-------------|
| Biological process | GO:0030199 Collagen fibril organization | 10    | 5.07E-22    |
|           | GO:0030198 Extracellular matrix organization      | 13    | 7.24E-22    |
|           | GO:0043933 Protein-containing complex subunit organization | 11    | 2.22E-08    |
|           | GO:0071230 Cellular response to amino acid stimulus | 5     | 3.58E-08    |
|           | GO:0043588 Skin development                      | 7     | 1.08E-07    |
| Molecular function | GO:0048407 Platelet-derived growth factor binding | 5     | 1.33E-11    |
|           | GO:0048332 SMAD binding                          | 3     | 0.0003      |
|           | GO:0005201 Extracellular matrix structural constituent | 3     | 0.0003      |
|           | GO:0004222 Metalloendopeptidase activity          | 3     | 0.00059     |
|           | GO:0002020 Protease binding                      | 3     | 0.00088     |
| Cellular component | GO:0005581 Collagen trimer                      | 9     | 5.36E-17    |
|           | GO:0062023 Collagen-containing extracellular matrix | 9     | 1.85E-15    |
|           | GO:0031012 Extracellular matrix                   | 10    | 2.75E-15    |
|           | GO:0005583 Fibrillar collagen trimer              | 6     | 4.20E-15    |
|           | GO:0044420 Extracellular matrix component         | 7     | 5.33E-14    |
| KEGG pathways | hsa04974 Protein digestion and absorption        | 9     | 9.42E-18    |
|           | hsa04512 ECM-receptor interaction                 | 4     | 1.03E-06    |
|           | hsa04510 Focal adhesion                          | 4     | 2.15E-05    |
|           | hsa05165 Human papillomavirus infection           | 4     | 7.32E-05    |
|           | hsa05146 Amoebiasis                              | 3     | 7.32E-05    |

### TABLE 2 | Module 2 GO and KEGG pathways analysis.

| ID        | GO-term                                         | Count | p-value     |
|-----------|------------------------------------------------|-------|-------------|
| Biological process | GO:0050896 Response to stimulus                | 214   | 1.08E-23    |
|           | GO:0065008 Regulation of biological quality    | 127   | 2.60E-18    |
|           | GO:0010033 Response to organic substance        | 110   | 5.34E-18    |
|           | GO:0042221 Response to chemical                 | 137   | 1.32E-17    |
|           | GO:0043062 Extracellular structure organization | 38    | 1.58E-17    |
| Molecular function | GO:0005488 Binding                              | 242   | 1.86E-10    |
|           | GO:005102 Signaling receptor binding            | 64    | 1.86E-10    |
|           | GO:0030545 Receptor regulator activity          | 33    | 8.42E-10    |
|           | GO:0004252 Serine-type endopeptidase activity   | 21    | 8.42E-10    |
|           | GO:0048018 Receptor ligand activity             | 31    | 3.00E-09    |
| Cellular component | GO:0044421 Extracellular region part          | 92    | 1.22E-30    |
|           | GO:005576 Extracellular region                   | 123   | 1.22E-30    |
|           | GO:005615 Extracellular space                    | 76    | 5.72E-25    |
|           | GO:0031012 Extracellular matrix                 | 30    | 1.14E-13    |
|           | GO:0044420 Extracellular matrix component       | 14    | 4.21E-10    |
| KEGG pathways | hsa04080 Neuroactive ligand-receptor interaction | 23    | 3.49E-08    |
|           | hsa04657 IL-17 signaling pathway                | 12    | 6.60E-06    |
|           | hsa04512 ECM-receptor interaction               | 11    | 9.63E-06    |
|           | hsa00830 Retinol metabolism                     | 10    | 9.63E-06    |
|           | hsa04610 Complement and coagulation cascades     | 10    | 3.33E-05    |

according to the 11 lncRNAs. Consequently, 7 lncRNAs including AFAP1-AS1, AQP4-AS1, C11orf97, HOTTIP, LINC00460, LINC01234, and SLC8A1-AS1 were harvested in Cox regression. Risk score = (0.06844*AFAP1-AS1)–(1.7559* C11orf97) +(0.20486*LINC00460)+ (0.09473*LINC01234)–(0.16046*SLC8A1-AS1)–
### TABLE 3 | Module 3 GO and KEGG pathways analysis.

| ID          | GO-term                                      | Count | p-value     |
|-------------|----------------------------------------------|-------|-------------|
|             | **Biological process**                       |       |             |
| GO:0003012  | Muscle system process                        | 41    | 1.28E-49    |
| GO:0006936  | Muscle contraction                           | 39    | 4.87E-49    |
| GO:0030049  | Muscle filament sliding                       | 24    | 1.43E-40    |
| GO:0070252  | Actin-mediated cell contraction              | 26    | 6.47E-38    |
| GO:0030029  | Actin filament-based process                 | 31    | 1.35E-26    |
|             | **Molecular function**                       |       |             |
| GO:0008092  | Cytoskeletal protein binding                 | 33    | 1.23E-21    |
| GO:0033779  | Actin binding                                | 25    | 1.28E-20    |
| GO:0003779  | Actin binding                                | 25    | 1.28E-20    |
| GO:0008307  | Structural constituent of muscle             | 13    | 4.89E-18    |
| GO:0051015  | Actin filament binding                        | 14    | 3.53E-13    |
| GO:0005198  | Structural molecule activity                 | 19    | 9.42E-10    |
|             | **Cellular component**                       |       |             |
| GO:0044449  | Contractile fiber part                       | 34    | 1.75E-42    |
| GO:0030017  | Sarcomere                                    | 33    | 4.42E-42    |
| GO:009512   | Supramolecular fiber                         | 35    | 5.34E-25    |
| GO:0031674  | I band                                       | 19    | 3.53E-22    |
| GO:0015629  | Actin cytoskeleton                           | 26    | 3.53E-22    |
|             | **Pathways**                                 |       |             |
|             | **KEGG pathways**                            |       |             |
| hsa05414    | Dilated cardiomyopathy (DCM)                 | 14    | 2.96E-16    |
| hsa05410    | Hypertrophic cardiomyopathy (HCM)            | 13    | 1.66E-15    |
| hsa04261    | Adrenergic signaling in cardiomyocytes       | 15    | 1.66E-15    |
| hsa04260    | Cardiac muscle contraction                   | 11    | 1.03E-12    |
| hsa04020    | Calcium signaling pathway                    | 12    | 2.48E-10    |

**FIGURE 4 | CeRNA network was visualized in Cytoscape (A).** The yellow represents miRNA upregulation and purple downregulation. The red means lncRNA upregulation and green downregulation. The mazarine indicates mRNA upregulation and blue downregulation. Gray edges indicate lncRNA-miRNA-mRNA interactions. (B) Venn diagram of DEmiRNAs might target DEMRNAs.

(0.20076* AQP4-AS1) (Figure 5C). Afterwards, the OSCC patients were divided into high risk and low risk group based on median value of Cox regression model. The distribution of the risk score along with the corresponding survival data and the 7 lncRNAs expression demonstrated that the high-risk OSCC patients tended to experience shorter survival time, and low-risk
TABLE 4 | The miRcode database revealed interactions between DElncRNAs and DEmiRNAs.

| miRNA     | IncRNA          |
|-----------|-----------------|
| hsa-mir-204 | PART1 LINC00314 | LINC00221 MUC19 LINC00487 |
|           | SFTA1P HOTAIR   | LINC00452 LINC00443 ERVME61-1 |
|           | LINC00486 IDI2-AS1 ZFY-AS1 LINC00472 LINC00330 |
|           | TM4SF19-AS1 ADAMTS9-AS2 HOTTIP LINC00461 LINC00491 |
|           | NIK2-1-AS1 RMST LINC00520 LINC00524 LINC00261 |
| hsa-mir-210 | ERVME61-1 LINC00461 ALDH1L1-AS2 TSPEAR-AS1 LINC00443 |
|           | PART1 LINC00314 | LINC00221 MUC19 LINC00487 |
|           | SFTA1P HOTAIR   | LINC00452 |
| hsa-mir-211 | LINC00486 IDI2-AS1 ZFY-AS1 LINC00472 LINC00330 |
|           | TM4SF19-AS1 ADAMTS9-AS2 HOTTIP LINC00461 LINC00491 |
|           | NIK2-1-AS1 RMST LINC00520 LINC00524 LINC00261 |
|           | LINC00486 IDI2-AS1 ZFY-AS1 LINC00472 LINC00330 |
|           | TM4SF19-AS1 ADAMTS9-AS2 HOTTIP LINC00461 LINC00491 |
|           | NIK2-1-AS1 RMST LINC00520 LINC00524 LINC00261 |
|           | LINC00486 IDI2-AS1 ZFY-AS1 LINC00472 LINC00330 |
|           | TM4SF19-AS1 ADAMTS9-AS2 HOTTIP LINC00461 LINC00491 |
|           | NIK2-1-AS1 RMST LINC00520 LINC00524 LINC00261 |

The miRDB, miRanda, and TargetScan database revealed interactions between DEmiRNAs and DEmRNAs.

| miRNA     | mRNA          |
|-----------|---------------|
| hsa-mir-204 | COL5A3 PTPRT KHL40 |
|           | SAMD5 CHRD1 HOXC8 IL11 |
| hsa-mir-210 | GPD1L SERTM1 |
| hsa-mir-211 | PTPRT IL11 KHL40 |
|           | HOXC8 CHRD1 SAMD5 GREM2 |

patients were opposite (Figures 5D,E). Correspondingly, the protective genes in high risk group expression level were lower than low group. On the contrary, risky genes were higher in high-risk group (Figure 5F).

Kaplan–Meier curves along with logrank p test were used to evaluate its accuracy in the training and validation set. According to the risk formula, obvious differences in survival analysis were determined in high- and low-risk groups (Supplementary Figures 2A,B). Meanwhile, stratified analysis in disparate grades was further carried out and indicated that risk level was relevant to prognosis (Supplementary Figures 2C,D). Subsequently, the ROC was plotted, and its area Under the Curve (AUC) is 0.776 (Supplementary Figure 2E).

TABLE 5 | The miRDB, miRanda, and TargetScan database revealed interactions between DEmiRNAs and DEmRNAs.

| miRNA     | mRNA          |
|-----------|---------------|
| hsa-mir-204 | COL5A3 PTPRT KHL40 |
|           | SAMD5 CHRD1 HOXC8 IL11 |
| hsa-mir-210 | GPD1L SERTM1 |
| hsa-mir-211 | PTPRT IL11 KHL40 |
|           | HOXC8 CHRD1 SAMD5 GREM2 |

Screening Biomarkers
A total of 6 genes (GNA14, CMA1, DKK1, HOXC6, HCG22, and HOTPTTIP) were identified as primitive biomarkers. The results of
univariate Cox regression indicated that there were 36 mRNAs (Table 7-3) and 6 lncRNAs (Table 7-1) regarded as prognostic factors \( (p < 0.05) \). However, none of the miRNAs were related to prognosis (Table 7-2). Meanwhile, following the combined Kaplan–Meier curves with logrank \( p \) test, the genes were clearly associated with overall survival \( (p < 0.01) \) and selected as primitive biomarkers. Finally, 4 mRNAs (Figures 6A–D) and 2 lncRNAs (Figures 6E,F) were identified. Subsequently, GEO expression profiles were used to verify the 6 biomarkers. As expected, most of the biomarker expression levels in GEO were consistent with TCGA. However, lncRNA HOTTIP upregulated in tumor tissues compared with adjacent normal tissues. Meanwhile, Kaplan–Meier analysis suggested that low expression of GNA14, CMA1, and HCG22 were related to poor survival (Figures 7A–C). High HOTTIP expression showed obviously poorer overall survival than those with low HOTTIP expression (Figure 7D). Unfortunately, there was no difference for DKK1 and HOXC6 in OSCC tissues, indicating that the 2 mRNAs might not be biomarkers in OSCC (Supplementary Figure 3A). Meanwhile, the protein levels of CMA1 and GNA14 were detected in OSCC cell lines and tissues (Supplementary Figure 3B). Finally, a combination of these biomarkers can predict overall survival better (Supplementary Figures 3A,B).

**DISCUSSION**

In the past decades, the 5-year survival rate of OSCC has remained low (3) in spite of advances in surgical treatment and radiotherapy. Hence, it is vital to find and identify biomarkers for early diagnosis and prognosis of OSCC.
In this study, a total of 1,919 DEMRNAs, 286 DElncRNAs and 111 DEMiRNAs were identified. GO analysis revealed that the function of DEMRNAs is mainly associated with tissue development, cell differentiation and calcium binding, which play a vital role in tumorigenesis. In addition, KEGG pathways analysis showed that DEMRNAs mainly enriched in the protein digestion and absorption, calcium signaling and focal adhesion pathways. Among these pathways, the calcium signaling and focal adhesion pathways are significantly associated with cancers. For example, abnormal Ca2+ level is related to glioblastoma and gastric adenocarcinoma (22, 23). In addition, research shows that focal adhesion is relevant to therapy resistance and plays a vital role in carcinogenesis, tumor progression and metastasis (24).

In the present study, 712 mRNAs were enrolled in the PPI network, and module analysis was performed. The majority of these genes were relevant to the PI3K-Akt signaling and calcium signaling pathways, which are associated with occurrence and development of tumors (25, 26). Meanwhile, the PI3K-Akt signaling pathway also played a significant role in OSCC (27). Six hub genes associated with overall survival—GNA14, GRM5, KRT3, KRT26, TACR1, and HTR2C—were enriched in module 2, indicating that module 2 plays a vital role in the PPI network and OSCC prognosis. To our surprise, GNA14 was identified as a potential biomarker by PCR assay and as a hub gene by bioinformatics analysis, which indicated that GNA14 may play a crucial role in OSCC carcinogenesis.

In our ceRNA network, 2 lncRNAs (HOTIP, HCG22) were identified as prognostic biomarkers. Recently, increasing studies showed that aberrant expression of HOTTIP is associated with various cancers. For instance, knockdown of HOTTIP suppresses growth and invasion and induces apoptosis of oral tongue squamous carcinoma cells (28). However, HOTTIP expression has no difference in GEO, which may be associated with racial difference. In addition, bioinformatics analysis confirmed that HCG22 is differentially expressed in various cancers (29), and the mechanism in OSCC should be researched.

Because of poor prognosis and high recurrence, we constructed a risk score formula by comprehensive analysis of lncRNA to predict overall survival in OSCC. Finally, 7 lncRNA were enrolled in Cox regression and it can predict overall survival accurately. In oral cancer, overexpression of the lncRNA AFAP1-AS1 is associated with the proliferation, invasion and survival of tongue squamous cell carcinoma via

| TABLE 7-1 | lncRNA univariate Cox regression results. |
|------------|------------------------------------------|
| lncRNA     | HR           | p-value     |
| LINCO1234  | 1.132608781  | 0.000684542 |
| AC073130.1 | 1.283657314  | 0.007536904 |
| HCG22      | 0.76621593   | 0.020101083 |
| HOTTIP     | 1.190591243  | 0.02547972  |
| AFAP1-AS1  | 1.07265706   | 0.037909467 |
| LINCO1322  | 1.090081006  | 0.040038896 |

| TABLE 7-2 | miRNA univariate Cox regression results. |
|------------|------------------------------------------|
| miRNA      | HR           | p-value     |
| hsa-mir-503 | 0.912139652  | 0.187845207 |
| hsa-mir-301b| 1.054389006  | 0.446577624 |
| hsa-mir-31  | 1.039652476  | 0.334137026 |
| hsa-mir-204 | 1.077656651  | 0.08420409  |
| hsa-mir-211 | 1.034251589  | 0.67807554  |
| hsa-mir-375 | 0.948826626  | 0.186073347 |
| hsa-mir-210 | 1.00407662   | 0.941092685 |

| TABLE 7-3 | mRNA univariate Cox regression results. |
|------------|------------------------------------------|
| mRNA       | HR           | p-value     |
| CMA1       | 0.84609833   | 0.000233691 |
| RSPO1      | 0.841607027  | 0.00031998  |
| DK1        | 1.113114821  | 0.001466491 |
| GRIA3      | 0.875364383  | 0.002180697 |
| IL17A      | 0.836798749  | 0.002950362 |
| IL17F      | 0.770775988  | 0.003080682 |
| ALB        | 1.097351002  | 0.001978569 |
| PLA1       | 1.227694927  | 0.010331743 |
| HAO2       | 0.835004357  | 0.010788042 |
| GNA14      | 0.82424831   | 0.011610204 |
| EGF        | 1.104743198  | 0.01450506  |
| RSPH4A     | 0.83254099   | 0.018772383 |
| ENPP3      | 0.874588953  | 0.023424344 |
| GRM5       | 0.83886418   | 0.023657185 |
| AGTR2      | 1.186356377  | 0.024809836 |
| MLXIPL     | 1.082210876  | 0.028255755 |
| PLIN1      | 1.104355635  | 0.030146844 |
| IL13RA2    | 0.916682105  | 0.030234187 |

| mRNA       | HR           | p-value     |
|------------|---------------|--------------|
| SLC1O1B1   | 1.106636606   | 0.031730512  |
| GLP1R      | 0.873802152   | 0.03185888   |
| TMOD1      | 1.083680927   | 0.032298388  |
| ADCY2      | 1.074843204   | 0.034551472  |
| HOXC6      | 1.152748928   | 0.035502816  |
| CDKN2A     | 0.947072266   | 0.036677858  |
| HIST1H2AG  | 1.160899705   | 0.036710592  |
| PKLR       | 0.908396028   | 0.03764026   |
| NPSR1      | 1.087738487   | 0.038197924  |
| ACAN       | 0.887171624   | 0.038226114  |
| ENO2       | 1.137767875   | 0.038883131  |
| HBA2       | 0.914606949   | 0.0437109   |
| NRG3       | 0.868606968   | 0.043744345  |
| PCK1       | 1.100965409   | 0.044003417  |
| DMP1       | 1.128478651   | 0.044165512  |
| SPTB       | 1.080475996   | 0.04595825   |
| AKR1C3     | 1.073842523   | 0.047306782  |
| OLFM4      | 1.053069233   | 0.048107359  |
FIGURE 6 | Combining Kaplan–Meier survival analysis with univariate Cox regression to screen biomarkers in OSCC patients. Kaplan–Meier survival curves and GEO gene expression profiles for the three protective genes, GNA14 (A), CMA1 (B) and HCG22 (C), and risky genes DKK1 (D), HOXC6 (E), and HOTTIP (F) were associated with overall survival in OSCC. GNA14: G protein subunit alpha 14; CMA1: chymase 1; DKK1: dickkopf WNT signaling pathway inhibitor 1; HOXC6: homeobox C6; HOTTIP: HOXA distal transcript antisense RNA; HCG22: HLA complex group 22.
FIGURE 7 | The expression levels of 2 mRNAs, GNA14 (A) and CMA1 (B), and 2 lncRNAs HCG22 (C), and HOTTIP (D) in 5 OSCC cell lines and 49 pairs adjacent tissues and tumor tissues. HOK is used as control. Then, Kaplan–Meier analysis along with logrank p was used to compare the survival of the low expression group and the high expression group. However, mRNA levels of DKK1 (E) and HOXC6 (F) showed no difference in OSCC tissues. *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001.
the Wnt/β-catenin signaling pathway. LINC00460/EZH2/ KLF2 and LINC00460/miR-149-5p/CUL4A crosstalk serve as critical effectors in CRC tumorigenesis and progression (30). Chen et al. showed that LINC01234 expression was significantly upregulated in gastric cancer tissues and functioned as a ceRNA to regulate CBFB expression by sponging miR-204-5p (31). However, there is no report on lncRNA C11orf97, SLC8A1-AS1, and AQP4-AS1. Whether these lncRNAs play important roles in the development and progression of OSCC remains to be further investigated. Meanwhile, univariate Cox regression and multivariate Cox regression analysis suggested that HOTITP was related to prognosis, which indicated that HOTITP was regarded as an independent prognostic biomarker.

Eventually, four biomarkers including GNA14, CMA1, HOTITP, and HCG22 were selected as biomarkers in OSCC by combining comprehensive analysis with PCR validation. Neuhaus J et al. indicated that aberrant CMA1 expression was found in Prostate Cancer (32). In addition, GNA14 was identified as biomarkers and a hub gene, suggesting that GNA14 was obviously relevant to OSCC. Our analysis shows that lncRNA HOTITP and HCG22 are also biomarkers of OSCC. HOTITP and HCG22 can interact with hsa-miR-21 in the ceRNA network, which promote oral cancer invasion via the Wnt/β-catenin pathway (33). Accordingly, exploring their mechanism in OSCC may provide new therapeutic targets. Furthermore, DKK1 and HOXC6 were excluded because there was no difference in mRNA level in OSCC tissues. The reasons may be racial difference or limited sample size.

We successfully constructed the ceRNA network, which reveals the potential mechanisms between lncRNA and mRNA. It may provide a new vision on therapeutic targets in OSCC by exploring the underlying mechanisms. At the same time, we also constructed a 7 lncRNA risk score model which is positively associated with overall survival in OSCC. We can put forward reasonable therapy at an appropriate time according to the risk model. Meanwhile, determining revisiting patients and follow-up improve the overall survival in OSCC based on risk level. Eventually, based on the 4 biomarkers, it may be beneficial to realize early-stage diagnosis and provide new therapeutic targets in OSCC.

Though the study might have crucial clinical importance, we still need to consider several limitations. First, in terms of sample numbers, both the TCGA database and clinical specimens which were collected at Nanfang Hospital are far from inadequate. We need to collect more information to continue verifying its accuracy. Second, the information from TCGA may show bias. Although we have validated it in cell lines and clinical specimens, we should continue to do further research. Third, the function and mechanism of biomarkers in OSCC need to be further studied in vivo and in vitro.

CONCLUSION

In summary, our study identifies that 2 mRNAs and 2 lncRNAs might be novel important prognostic factors and potential treatment targets for OSCC. Furthermore, we established a disordered lncRNA-miRNA-mRNA ceRNA network which is beneficial to understanding the relationship between lncRNA and mRNA and provides efficient strategies for subsequent functional studies of lncRNAs. Meanwhile, construction of the PPI network provides a new vision for OSCC treatment, and the risk score model is helpful for improving the overall survival in OSCC.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov/, https://www.ncbi.nlm.nih.gov/geo/, https://string-db.org/, http://www.mircode.org/, http://www.mirdb.org/, http://mirtarbase.mbc.nctu.edu.tw/, http://www.targetscan.org.

ETHICS STATEMENT

The study protocol was approved by the Ethics Committees of Nanfang Hospital of Guangdong Province (NO: NFEC-2018-027).

AUTHOR CONTRIBUTIONS

GH: design and initiation of the study, quality control of data and algorithms, data analysis and interpretation, manuscript preparation and editing. QW: data acquisition. ZZ: statistical analysis. TS: patient recruitment and clinical support and oversight. X-ZL: study concept, design and initiation of the study. All authors are revision and approval the final version of the paper.

FUNDING

This study was supported by the National Natural Science Foundation of China (Grant Number: 81472536), the Science and Technology Planning Project of Guangdong Province (Grant Number: 2017A020215181), the Scientific Research Launch Project of Southern Medical University (Grant Numbers: CX2018N016, PY2018N031), the Project of Guangdong Provincial Department of Education (Grant Number: 2018KTSCX026) and the NanFang Hospital President Funding (Grant Number: 2014027).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01054/full#supplementary-material

Supplementary Figure 1 | 6 hub genes associated with overall survival. (A) GNA14, (B) GRM5, (C) KRT3, (D) KRT26, (E) TRAC1, (F) HTR2C.

Supplementary Figure 2 | Kaplan–Meier analysis along with logrank p was used to compare the survival of the low-risk group and high-risk group. (A) Survival analysis of training set. (B) Survival analysis of validation set. (C) OSCC patients with grade I and II. (D) OSCC patients with grade IV and III. (E) ROC based on risk score model.

Supplementary Figure 3 | The mRNA expression level of DKK1 and HOXC6 in OSCC (A). The protein level of CMA1 and GNA14 in OSCC cell lines and tissues (B). Combination of two and more biomarkers predicted OSCC patients overall survival.
REFERENCES

1. Leemans CR, Braakhuis BJ, Brakenhoff RH. The molecular biology of head and neck cancer. Nat Rev Cancer. (2011) 11:9–22. doi: 10.1038/nrc2982

2. Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tan PF, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med. (2010)363:24–35. doi: 10.1056/NEJMoa0912217

3. Zini A, Czerninski R, Sgan-Cohen HD. Oral cancer over four decades: epidemiology, trends, histology, and survival by anatomical sites. J Oral Pathol Med. (2010) 39:299–305. doi: 10.1111/j.1600-0714.2009.00845.x

4. Vermorken JB, Herbst RS, Leon X, Amellal N, Baselga J. Overview of the efficency of cetuximab in recurrent and/or metastatic squamous cell carcinoma of the head and neck in patients who previously failed platinum-based therapies. Cancer-Am Cancer Soc. (2008) 112:2710–9. doi: 10.1002/cncr.23442

5. Takahashi H, Yamanoto S, Yamada S, Umeda M, Shigeta T, Minamikawa T, et al. Effects of post-operative chemotherapy and radiotherapy on patients with squamous cell carcinoma of the oral cavity and multiple regional lymph node metastases. Int J Oral Maxillofac Surg. (2014) 43:680–5. doi: 10.1016/j.iojm.2013.11.013

6. Hangauer MJ, Vaughn IW, McManus MT. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. PLoS Genet. (2013) 9:e1003569. doi: 10.1371/journal.pgen.1003569

7. Zhang L, Meng X, Zhu WX, Yang DC, Chen R, Jiang Y, Xu T. Long non-coding RNAs in Oral squamous cell carcinoma: biologic function, mechanisms and clinical implications. Mol Cancer. (2019) 18:102. doi: 10.1186/s12943-019-1021-3

8. Wang J, Zhang X, Chen W, Hu X, Li J, Liu C. Regulatory roles of long noncoding RNAs implicated in cancer hallmarks. Int J Cancer. (2019). doi: 10.1002/ijc.32277. [Epub ahead of print].

9. Shao T, Huang J, Zheng Z, Wu Q, Liu T, Li X. SCCA, TSGF, and the long non-coding RNA AC007271.3 are effective biomarkers for diagnosing oral squamous cell carcinoma. Cell Physiol Biochem. (2018) 47:26–38. doi: 10.1159/000489741

10. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfo P. A cRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell. (2011) 146:353–8. doi: 10.1016/j.cell.2011.07.014

11. El-Sakka H, Kujan O, Farah CS. Assessing miRNAs profile expression as a risk stratification biomarker in oral potentially malignant disorders: a systematic review. Oral Oncol. (2018)77:57–82. doi: 10.1016/jIELD016/0106/02.

12. Cao M, Zheng L, Liu J, Doblieuman T, Hu S, Guo V, et al. MicroRNAs as effective surrogate biomarkers for early diagnosis of oral cancer. Clin Onc Investig. (2018) 22:571–81. doi: 10.1007/s00785-017-3137-6

13. Liu S, Liu LH, Hu WW, Wang M. Long non-coding RNA TUG1 regulates the development of oral squamous cell carcinoma through sponging miR-524-5p to mediate DLX1 expression as a competitive endogenous RNA. J Cell Physiol. (2019) 234:20260–16. doi: 10.1002/jcp.28620

14. Pramanik KK, Nagini S, Singh AK, Mishra P, Kashyap T, Nath N, et al. Glycogen synthase kinase-3β regulated metabolism of matrix metalloproteinase-9 and its involvement in oral squamous cell carcinoma progression and invasion. Cell Oncol. (2018)41:47–60. doi: 10.1007/s13402-017-0538-0

15. Lun AT, Chen Y, Smyth GK. It's DE-licious: a recipe for differential expression analyses of RNA-seq experiments using quasi-likelihood methods in edgeR. Methods Mol Biol. (2016) 1418:391–416. doi: 10.1007/978-1-4939-3571-9_19

16. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinform. (2003) 4:2. doi: 10.1186/1471-2105-4-2

17. Jeggari A, Marks DS, Larsson E. miRodeo: a map of putative microRNA target sites in the long non-coding transcriptome. Bioinformatics. (2012) 28:2062–3. doi: 10.1093/bioinformatics/bts344

18. Wang X. Improving microRNA target prediction by modeling with unambiguously identified microRNA-target pairs from CLIP-ligation experiments. Bioinformatics. (2016) 32:1316–22. doi: 10.1093/bioinformatics/btw02

19. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Res. (2018) 46:D296–302. doi: 10.1093/nar/gkx1063

20. Fromm B, Bilipp T, Peck LE, Johansen M, Tarver JE, King BL, et al. A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microRNAome. Annu Rev Genet. (2015) 49:213–42. doi: 10.1146/annurev-genet-120213-092023

21. Mao Y, Fu Z, Dong L, Zheng Y, Dong J, Li X. Identification of a 26-lncRNAs risk model for predicting overall survival of cervical squamous cell carcinoma based on integrated bioinformatics analysis. DNA Cell Biol. (2019) 38:322–32. doi: 10.1089/dna.2018.4533

22. Waugh MG. Chromosomal instability and phosphoinositide pathway gene signatures in glioblastoma multiforme. Mol Neurobiol. (2016) 53:621–30. doi: 10.1007/s12035-014-9034-9

23. Dai L, Zhuang L, Zhang B, Wang F, Chen X, Xia C, et al. DAG1/PKDdelta and IP3/Ca(2+)/CaMK II operate in parallel to each other in PLC(γ)- driven cell proliferation and migration of human gastric adenocarcinoma cells, through Akt/mTOR/S6 pathway. Int J Mol Sci. (2015) 16:28510–22. doi: 10.3390/ijms16122616

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Huang, Wu, Zheng, Shao and Lv. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.