RESEARCH ARTICLE   **Open Access**

Construction and analysis of a lncRNA-miRNA-mRNA network based on competitive endogenous RNA reveal functional lncRNAs in oral cancer

Junhao Yin[1,2], Xiaoli Zeng[1,2], Zexin Ai[1,2], Miao Yu[1,2], Yang'ou Wu[1,2] and Shengjiao Li[1,2]*

Abstract

Background: A growing evidence suggests that long non-coding RNAs (lncRNAs) can function as a microRNA (miRNA) sponge in various diseases including oral cancer. However, the pathophysiological function of lncRNAs remains unclear.

Methods: Based on the competitive endogenous RNA (ceRNA) theory, we constructed a lncRNA-miRNA-mRNA network in oral cancer with the human expression profiles GSE74530 from the Gene Expression Omnibus (GEO) database. We used topological analysis to determine the hub lncRNAs in the regulatory ceRNA network. Then, function enrichment analysis was performed using the clusterProfiler R package. Clinical information was downloaded from The Cancer Genome Atlas (TCGA) database and survival analysis was performed with Kaplan-Meier analysis.

Results: A total of 238 potential co-dysregulated competing triples were obtained in the lncRNA-associated ceRNA network in oral cancer, which consisted of 10 lncRNA nodes, 41 miRNA nodes and 122 mRNA nodes. Additionally, we found lncRNA HCG22 exhibiting superior potential as a diagnostic and prognostic marker of oral cancer.

Conclusions: Our findings provide novel insights to understand the ceRNA regulation in oral cancer and identify a novel lncRNA as a potential molecular biomarker.

Keywords: Competitive endogenous RNA, Oral cancer, Long non-coding RNA, HCG22

Background

Oral cancer is a malignant neoplasia with low overall survival rates [1], among which oral squamous cell carcinoma (OSCC) is the most common type [2]. The predilection sites of oral cancer include the buccal mucosa, tongue and lower lip, and the occurrence is higher in people over fifty [3]. In 2018, the worldwide age-standardized rate per 100,000 person-years of oral cancer was 4.0, and 64.2% of the global incidence clustered in South Asia according to the Global Cancer Observatory [4]. It is well established that the major risk factors for oral cancer include tobacco chewing, smoking, alcohol drinking, excessive sunlight exposure and Human Papilloma Virus (HPV). In addition, several genetic factors such as tumor suppressor genes, oncogenes and regulatory genes may also play a crucial role in oral carcinogenesis [5, 6]. Genetic alterations to the genes of TP53, NOTCH1 and PIK3CA affect the epithelial cells and contribute to the microenvironment alterations such

* Correspondence: 07824@tongji.edu.cn
[1]Department of Oral and Maxillofacial Surgery, School and Hospital of Stomatology, Tongji University, Shanghai 200072, China
[2]Shanghai Engineering Research Center of Tooth Restoration and Regeneration, Shanghai 200072, China

as the ROS accumulation, overproduction of cytokines and epithelial to mesenchymal transition, inducing uncontrolled cell proliferation, growth and tumorigenesis [7, 8].

Initially, it was thought that coding RNAs play the most essential roles in cancer, while non-coding RNAs (ncRNAs) are no more than transcriptional noise. However, there is an increasing evidence indicating the important regulatory roles of ncRNAs in the occurrence and progression of various cancer types [9]. NcRNAs include circular RNAs, microRNAs, intronic RNAs and long non-coding RNAs [10]. MicroRNAs (miRNAs) are representative ncRNAs of 18–25 nucleotides in length [11]. They regulate the expression of target genes by inhibiting their translation and accelerating their degradation [12], and have been shown to be involved in many different physiological and pathological processes, including the epithelial-mesenchymal transition, metabolism, survival and more. First confirmed in 2005, long non-coding RNAs (lncRNAs) represent another type of ncRNA and also lack any protein-coding capacity [13]. The nucleotide sequence length of lncRNAs is typically between 200 and 10,0000 nt. In addition, some lncRNAs may be associated with cancer phenotypes [14] and were chosen as new diagnostic and prognostic biomarkers for various cancer types, including nasopharyngeal carcinoma [15], gastric cancer [16] and prostate cancer [17]. However, the role of lncRNAs in the development of OSCC remains to be explored.

The competitive endogenous RNA (ceRNA) hypothesis [18], a novel regulatory mechanism that received attention in 2011, indicated that circular RNAs, lncRNAs and pseudogenes can regulate the abundance of miRNAs as molecular sponges. The theory may lead to essential clues to understand gene regulatory networks in many diseases, including OSCC. For example, one of the initial findings indicate that the pseudogene PTENP1 has the miRNA sponge capacity to regulate the levels of the PTEN gene in cancer [19]. ANRIL is a miRNA sponge for miR-125a-3p that regulates the abundance of FGFR1 to promote the tumorigenesis of head and neck squamous cell carcinoma [20]. Kanagaraj Arun et al. [21] found that both lncRNAs PTENP1-AS and GAS5 could act as tumor suppressive ceRNAs in gastric cancer. These studies have shown lncRNAs to be potential biomarkers for the diagnosis and prognosis of various diseases.

In this study, we performed an analysis of the RNA expression profiles in oral cancer patients from the Gene Expression Omnibus (GEO) database at the National Center for Biotechnology Information (NCBI) to screen the differentially expressed lncRNAs (DELs) and mRNAs (DEMs) that are related to oral cancer. Then, we constructed a lncRNA-associated ceRNA network by combining bioinformatics and correlation analyses to find the hub lncRNAs in OSCC. Meanwhile, sequence data and clinical information were obtained from The Cancer Genome Atlas (TCGA) database. To further investigate the relationship between the expression pattern and clinical information of the oral cancer samples, the Kaplan-Meier survival analysis of hub lncRNAs was carried out.

Methods

GEO data collection

We retrieved the human expression profiles (accession number: GSE74530) of oral cancer from NCBI GEO [22], which were extracted from a study carried out by Oghumu et al. [23]. The expression data included the lncRNA and mRNA expression profiles. The samples were derived from the tumor tissue and adjacent non-tumor tissue of six OSCC patients who were enrolled in a Phase 0 clinical trial study. The microarray platform used to analyze these data was the GPL570 Affymetrix Human Genome U133 Plus 2.0 Array and the direct web link was listed in Additional file 1.

Data quality assessment

The AffyPLM package [24] in the R statistical language was used to analyze the data quality at the probe level. The boxplot representations of residuals and weights from probe level fits were obtained, by which the trend consistency of the expression data can be tested. The degradation of the RNA was assessed using the AffyRNAdeg function of AffyPLM, to assure the consistent trend and RNA integrity of the microarray dataset before further processing.

Data pre-processing and screening of differentially expressed lncRNAs and mRNAs

In order to identify the biological significance of each probe, the comprehensive gene annotation files were obtained from GENCODE in GTF format. The GENCODE annotation was the default gene annotation displayed in the Ensembl Genome Browser. The transcripts with a length of more than 200 nucleotides and a biotype categorized as "non_coding","processed_transcript","lincRNA", "retained_intron","antisense","sense_overlapping", "sense_intronic" and "bidirectional_promoter_lncrna" were labeled as "lncRNAs", while the transcripts with a biotype categorized as "protein_coding" were labeled as "mRNAs". Finally, 1210 expressed lncRNAs and 16,434 expressed mRNAs were annotated.

The Affymetrix probe level data were obtained by reading the CEL files using the ReadAffy function of the Affy R package [25], and then the raw data were preprocessed (background correction, normalization and summary expression computation). We used the limma

Bioconductor package [26] to explore the differentially expressed genes (DEGs) between adjacent normal tissue and tumors groups, including lncRNAs (DELs) and mRNAs (DEMs). The DELs and DEMs were filtered according to the cut-off criteria of adjusted p-value < 0.05 and $|\log_2 (\text{fold change})| > 1$.

Prediction of target miRNAs of DELs and miRNA-target interactions

The lncRNA targets of the miRNAs were collected using the transcriptome-wide mircoRNA target predictions from the miRcode database [27], which contained more than 10,000 lncRNAs. The predictions were made based on the GENCODE transcripts. Interactions between miRNA and mRNA were found in the 3 miRNA databases of miRTarBase 6.0 [28], miRDB [29] and TargetScan 7.0 [30], with the criteria that each target mRNA appears in at least 2 of them. The mRNA targets of the miRNA-mRNA interactions were merged with DEMs for further analysis.

Construction of the lncRNA-associated ceRNA network and topological analysis

A lncRNA-miRNA-mRNA interaction was identified as a potential ceRNA triple based on the following criteria [31]:

(1) The Pearson's correlation coefficients (PCCs) between each DELs-DEMs pair in oral cancer were calculated. The DEL-DEM pairs were regarded as co-dysregulated DEL-DEM pairs if the thresholds of the PCC value ranked in the top 0.05 percentile (PCC > 0.886) with a p-value < 0.05.

(2) Once it was confirmed that both mRNA and lncRNA in a co-dysregulated DEL-DEM pair were targeted by the same miRNA, this lncRNA-miRNA-mRNA interaction would then be identified as a potential co-dysregulated competing triple.

To give an insight into the roles of lncRNAs in the ceRNA network, we assembled all the potential co-dysregulated competing triples to build the lncRNA-miRNA-mRNA network and visualized the regulation network built from these interactions using the Cytoscape 3.7.1 [32] software.

Topological analysis is important to discover information in complex data sets. In order to study the geometric relationships between the data nodes, we computed the node degree and betweenness centrality (BC) of each node, which are both network topological features. Then, the nodes with a high node degree (> 5 connections) and a greater BC value were considered to be the hub nodes in the regulation network, which were more likely to play an important role in oral cancer [33].

Functional enrichment analysis

In order to explore the functions of the obtained lncRNAs, we performed Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of mRNAs in the lncRNA-associated regulation network using the clusterProfiler R package [34]. Simultaneously, the GO interaction network was built using the Biological Networks Gene Ontology tool (BiNGO) in the Cytoscape software [32]. We set p-value < 0.05 and Benjamini-Hochberg corrected p-value < 0.05 as the thresholds of the functional categories.

Construction of the key lncRNA-associated subnetworks

The lncRNAs with high node degree and BC value were chosen as hub lncRNAs, which were used with their related miRNAs and mRNAs in the regulation network to construct the subnetworks using the Cytoscape software. After that, gene functional enrichment analyses were performed for each subnetwork.

TCGA data collection

The RNAseq data and corresponding clinical data of 319 oral cancer samples (including 19 buccal mucosa samples, 54 floor of mouth samples, 116 larynx samples, and 130 tongue samples from the TCGA-HNSC project) were downloaded from The Cancer Genome Atlas (TCGA) database (http://portal.gdc.cancer.gov/; dbGaP Study Accession:phs000178) using the TCGAbiolinks R package [35]. The expression data and clinical information included 319 oral cancer samples and 44 matched normal samples. Meanwhile, 14 samples of non-solid tumors were excluded from the dataset and the accession numbers of all samples were included in Additional file 2. The raw data were preprocessed by TCGAanalyze_Preprocessing and TCGAanalyze_purity function of TCGAbiolinks. A total of 236 oral cancer samples and 44 normal samples were filtered.

Differentially expressed analysis and survival analysis

The TCGAbiolinks R package was used to filter the differentially expressed genes between normal and tumor samples according to the cut-off criteria of adjusted p-value < 0.05 and $|\log2 (\text{fold change})| > 1$. GO function analysis was carried out for up-regulated and down-regulated genes respectively. Furthermore, to determine the prognosis of oral cancer patients in relation to differentially expressed RNA signatures, survival curves of hub lncRNAs were analyzed using the survival package in the R statistical language with a threshold of log-rank $P < 0.05$.

Results

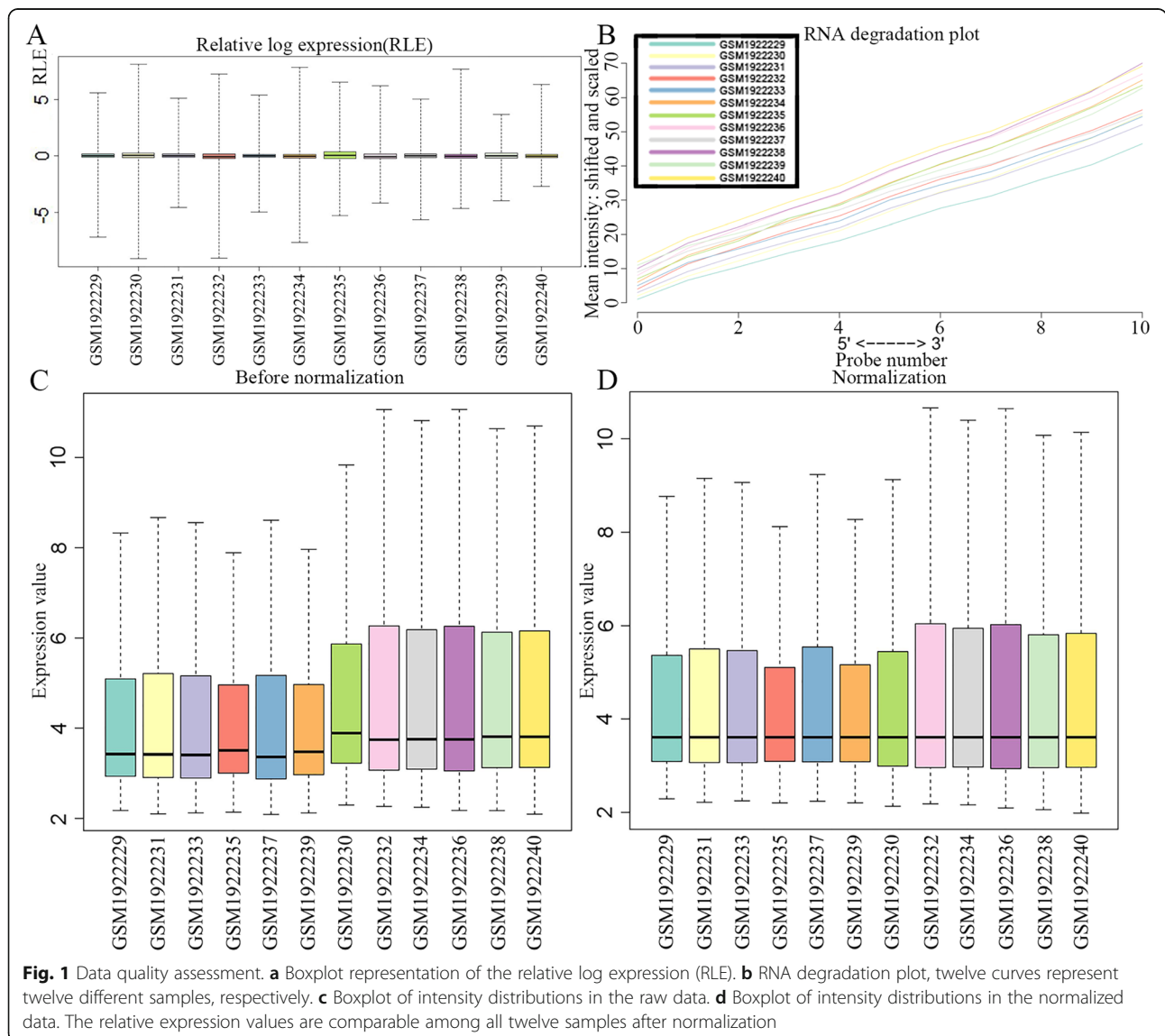### Data quality assessment and preprocessing

Regression analysis of the raw data was performed using the affyPLM R package. The relative log expression (RLE) plot revealed that the gene expression levels in GSE74530 were consistent with the median approaching 0 (Fig. 1a), indicating that the quality of the expression data was reliable. As for the RNA degradation plot, it showed that the RNA integrity has a good quality (Fig. 1b), and all the 12 samples can be used for further analysis. A total of 16,434 mRNAs and 1210 lncRNAs were identified in the microarray data using the human comprehensive gene annotation from GENCODE. As shown in Fig. 1c, there were no obvious outliers in the spread and location in boxplot, and small discrepancies can be sufficiently removed by normalization. The median values of 12 samples were almost at the same level after normalization (Fig. 1d),
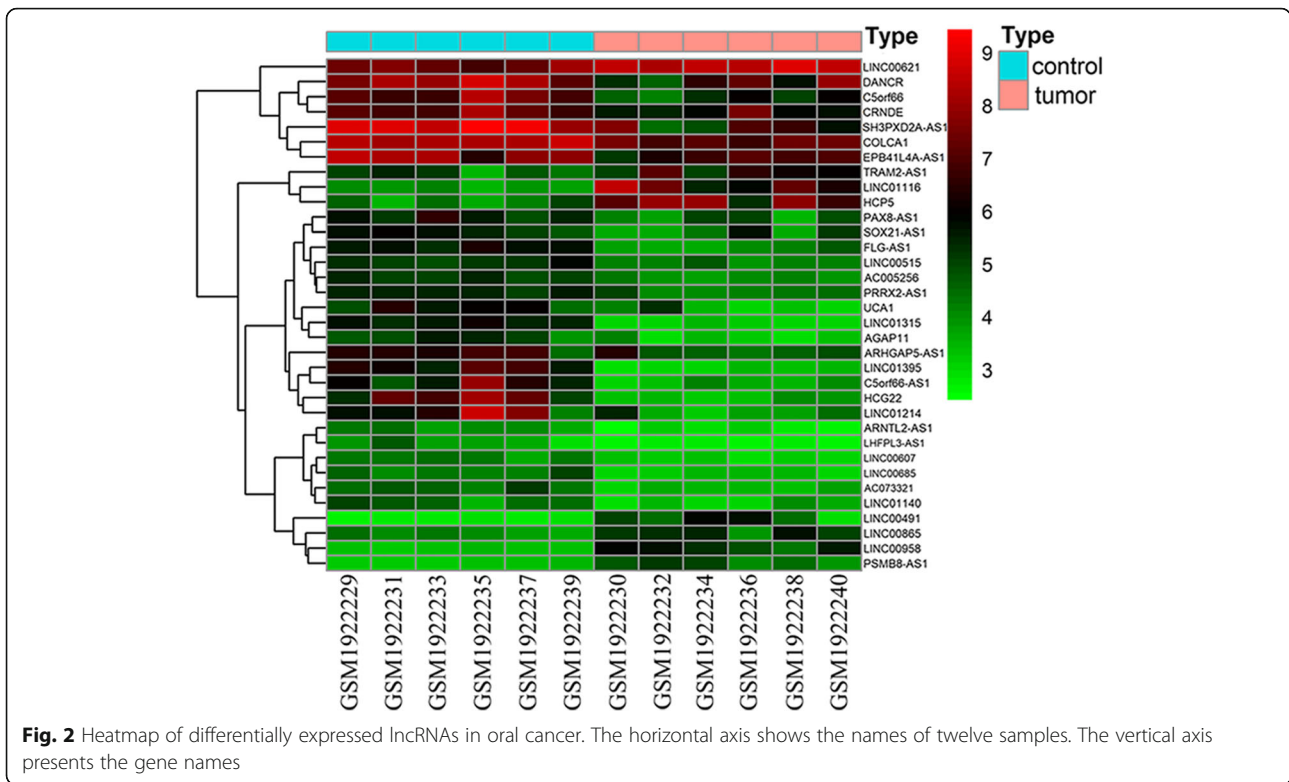
which effectively corrected the systematic differences between the chips.

The screening results of DELs and DEMs

We identified 34 DELs and 1641 DEMs by comparing the tumor groups with the adjacent normal tissue group using the limma package. The expression of these DELs and all the DEGs are visualized in a heatmap (Fig. 2) and a volcano plot (Fig .3), respectively. The miRNA targets of the lncRNAs were predicted using the miRcode database in R. The target mRNAs of these miRNAs were obtained using the three highly reliable microRNA target prediction databases of miRTarBase, TargetScan and miRDB, and the result was intersected with the above-mentioned DEMs.

As a result, we got a total of 2137 reliable miRNA-mRNA pairs and 162 predicted lncRNA-miRNA pairs

Fig. 1 Data quality assessment. **a** Boxplot representation of the relative log expression (RLE). **b** RNA degradation plot, twelve curves represent twelve different samples, respectively. **c** Boxplot of intensity distributions in the raw data. **d** Boxplot of intensity distributions in the normalized data. The relative expression values are comparable among all twelve samples after normalization

Fig. 2 Heatmap of differentially expressed lncRNAs in oral cancer. The horizontal axis shows the names of twelve samples. The vertical axis presents the gene names

(including 11 lncRNAs, 51 miRNAs and 851 mRNAs), which were used for further analysis.

Construction of the ceRNA network in oral cancer

In order to reduce the false positives, we calculated the PCC between the 34 DELs and 1641 DEMs, and the co-expressed pairs with the top 5% PCC value (PCC >

Fig. 3 Volcano plot of all differentially expressed genes in oral cancer. FC are fold-change. Downregulated genes are green and upregulated genes are red

0.886) were defined as the significant co-dysregulated competing pairs. Next, the results were merged with the 11 lncRNAs, 51 miRNAs and 851 mRNAs from the previous screening step. The intersection resulted in the selection of a total of 238 potential co-dysregulated competing triples, the full list is shown in Additional file 3.

To perform a deeper functional study of the lncRNAs that act as a miRNA sponge in oral cancer, we built a ceRNA network and applied the Cytoscape software to perform visualization (Fig. 4a). The network included 10 lncRNA nodes, 41 miRNA nodes, 122 mRNA nodes and 238 edges. In addition, several miRNAs in the network had been identified in oral cancer as listed in Table 1.
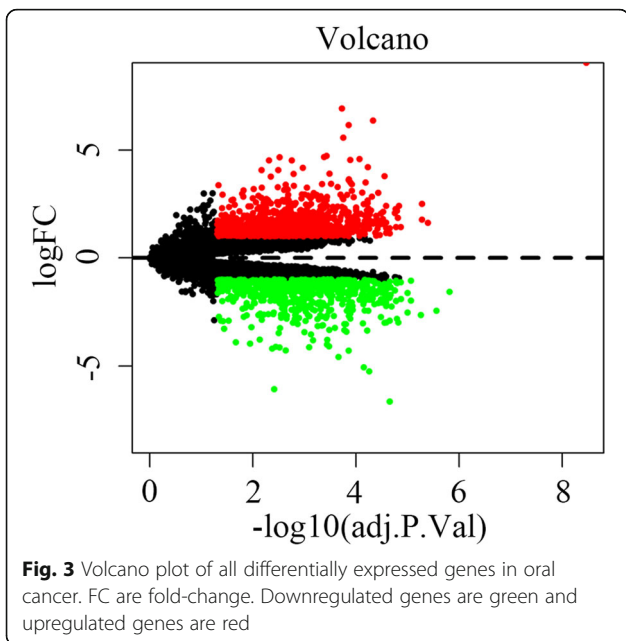
Functional enrichment analysis

Speculations on the possible functions of the lncRNAs were made through the functional enrichment analysis of their linked mRNAs. GO analysis was performed to analyze the functions of the mRNA nodes, and the GO interaction network was constructed using the BiNGO tool (Additional file 4). As a result, 18 GO terms were found to be significantly enriched (Fig. 4b and Additional file 5). Among these terms, the top three enriched ones were collagen binding, extracellular matrix binding and cell adhesion molecule binding, all of which belonged to the molecular function (MF) terms. Interestingly, extracellular matrix (ECM) binding
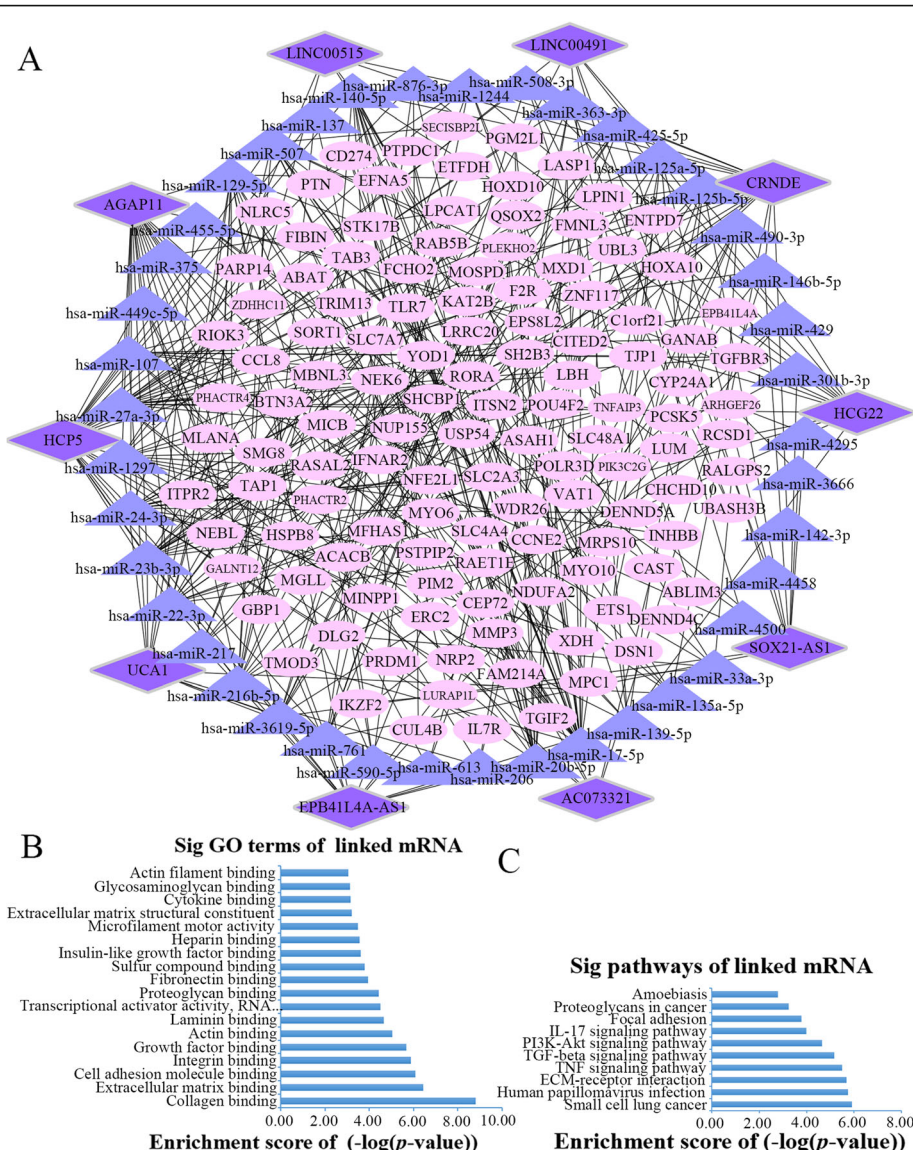
Fig. 4 The lncRNA associated ceRNA network and barplots of function enrichment analyses. **a** The lncRNA-miRNA-mRNA ceRNA network. The parallelograms represent lncRNAs, the ellipses represent mRNAs, and the triangles represent miRNAs. **b** The top 18 most significant Gene ontology terms. **c** The top 10 most significant pathway terms

Table 1 The effect characteristic of miRNAs in oral squamous cell carcinoma (OSCC)

miRNA	Targets	Effect characteristic	Reference
hsa-miR-139-5p	AKT	Inhibits oral cancer cell proliferation and induces cell apoptosis	[36]
hsa-miR-17-5p	Casp-2/−7/− 8, Bcl-2 and DIALO	Inhibits hypoxia-induced apoptosis	[37]
hsa-miR-107	p53	p53 upregulates miR-107 in normal condition	[38]
hsa-miR-375	Sp1 and cyclin D1	Inhibits cell growth and its expression is correlatedwith prognosis of TSCC	[39]
hsa-miR-455-5p	UBE2B	Binds SMAD3-specific promoter regions, leads toUBE2B downregulation, and contributes to oralcancer tumorigenesis	[40]
hsa-miR-137	CDK6	Arrests cell cycle at the G1-S checkpoint	[41]
hsa-miR-140-5p	ADAM10	Inhibits the invasion and migration of TSCC cells	[42]
hsa-miR-125b-5p	BAK1	Boosts the level of BAK1 that is controlling the apoptotic pathway	[38]

Table 2 The list of differentially expressed genes (node degree > 5)

Number	Degree	Name	Type
1	27	hsa-miR-17-5p	miRNA
2	24	HCP5	lncRNA
3	21	AGAP11	lncRNA
4	19	hsa-miR-24-3p	miRNA
5	18	hsa-miR-27a-3p	miRNA
6	17	hsa-miR-1297	miRNA
7	17	hsa-miR-23b-3p	miRNA
8	16	hsa-miR-140-5p	miRNA
9	16	hsa-miR-129-5p	miRNA
10	15	CRNDE	lncRNA
11	15	hsa-miR-22-3p	miRNA
12	15	hsa-miR-20b-5p	miRNA
13	14	HCG22	lncRNA
14	13	EPB41L4A-AS1	lncRNA
15	13	hsa-miR-507	miRNA
16	13	hsa-miR-107	miRNA
17	13	hsa-miR-216b-5p	miRNA
18	12	SOX21-AS1	lncRNA
19	12	hsa-miR-3619-5p	miRNA
20	11	UCA1	lncRNA
21	11	hsa-miR-761	miRNA
22	9	hsa-miR-490-3p	miRNA
23	9	hsa-miR-125b-5p	miRNA
24	9	hsa-miR-1244	miRNA
25	9	hsa-miR-139-5p	miRNA
26	7	YOD1	mRNA
27	7	LINC00491	lncRNA
28	7	hsa-miR-125a-5p	miRNA
29	7	hsa-miR-425-5p	miRNA
30	7	hsa-miR-363-3p	miRNA
31	7	hsa-miR-876-3p	miRNA
32	7	hsa-miR-455-5p	miRNA
33	7	hsa-miR-217	miRNA
34	6	NEK6	mRNA
35	6	MFHAS1	mRNA
36	6	RORA	mRNA
37	6	hsa-miR-33a-3p	miRNA
38	6	LINC00515	lncRNA
39	6	AC073321	lncRNA
40	6	hsa-miR-146b-5p	miRNA
41	6	hsa-miR-508-3p	miRNA
42	6	hsa-miR-135a-5p	miRNA

Table 3 List of the 15 genes with the top betweenness centrality

Number	Type	name	Betweenness Centrality
1	lncRNA	HCP5	0.26
2	lncRNA	AGAP11	0.16
3	miRNA	hsa-miR-17-5p	0.14
4	miRNA	hsa-miR-24-3p	0.13
5	miRNA	hsa-miR-27a-3p	0.10
6	miRNA	hsa-miR-1297	0.10
7	miRNA	hsa-miR-22-3p	0.09
8	lncRNA	CRNDE	0.09
9	miRNA	hsa-miR-140-5p	0.09
10	miRNA	hsa-miR-129-5p	0.08
11	lncRNA	HCG22	0.08
12	miRNA	hsa-miR-216b-5p	0.07
13	miRNA	hsa-miR-507	0.07
14	miRNA	hsa-miR-23b-3p	0.06
15	miRNA	hsa-miR-107	0.06

was related to the proliferation of OSCC cells [43] and played an important role in the growth and survival of oral cancer cells [44]. It was demonstrated that cell adhesion molecules, together with tumor-associated matrix molecules, are functionally involved in the progression of oral cancer [45]. What's more, discoidin domain receptor-1 (DDR1) could be activated by the specific binding with collagens (II,III) [46], and the activation of DDR1 has been reported in oral cancer [47]. The KEGG pathway analysis resulted in 10 enriched pathway terms, shown in Fig. 4c, including the terms of ECM-receptor interaction, small cell lung cancer, PI3K-Akt signaling pathway, focal adhesion, TGF-β signaling pathway and more (Additional file 6). Among these pathways, ECM-receptor interaction [48], PI3K-Akt signaling pathway [49], TNF signaling pathway [50] and TGF-β signaling pathway [51] were OSCC-related pathways.

Topological analysis of the ceRNA network

In order to identify the hub genes in the lncRNA-miRNA-mRNA network that are related to oral cancer, we computed the node degrees. In the study of Han et al. [33], the nodes with a degree greater than 5 were defined as hubs. Based on this research, a total of 42 nodes could be chosen as hubs, including 10 lncRNAs, 28 miRNAs and 4 mRNAs (Table 2 and Additional file 7). In addition, BC was also calculated as a measure to select the hubs [52] (Table 3). Higher BC values of the nodes indicated an increased important of these nodes in the regulatory network [53].

We found that the three lncRNAs of HCP5, AGAP11 and HCG22 had higher node degrees along with greater
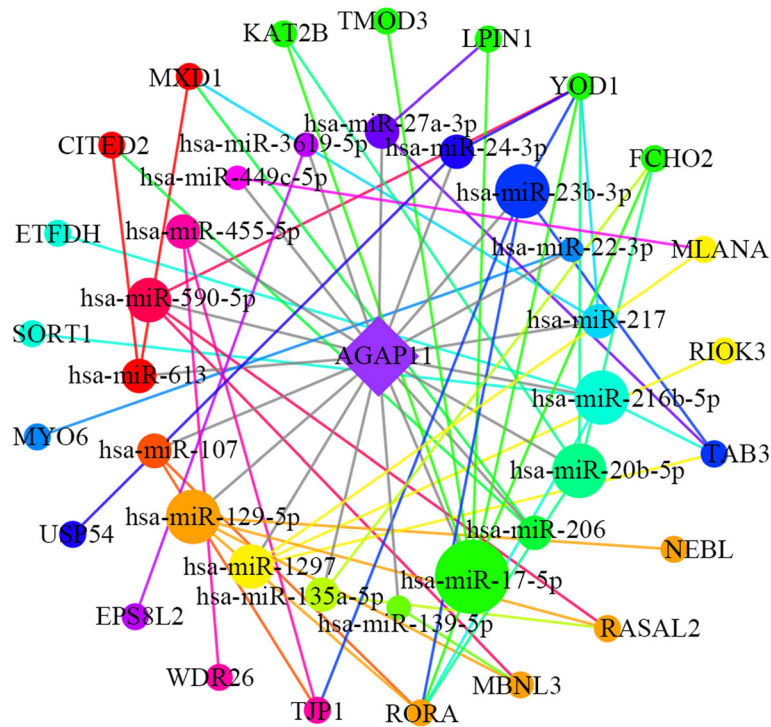
Fig. 5 The sub-network of hub lncRNA AGAP11. The rhombuses represent lncRNAs, the circles on the inner loop represent miRNAs, and the circles on the outer loop represent mRNAs. The bigger size circles have, the more nodes they are connected to
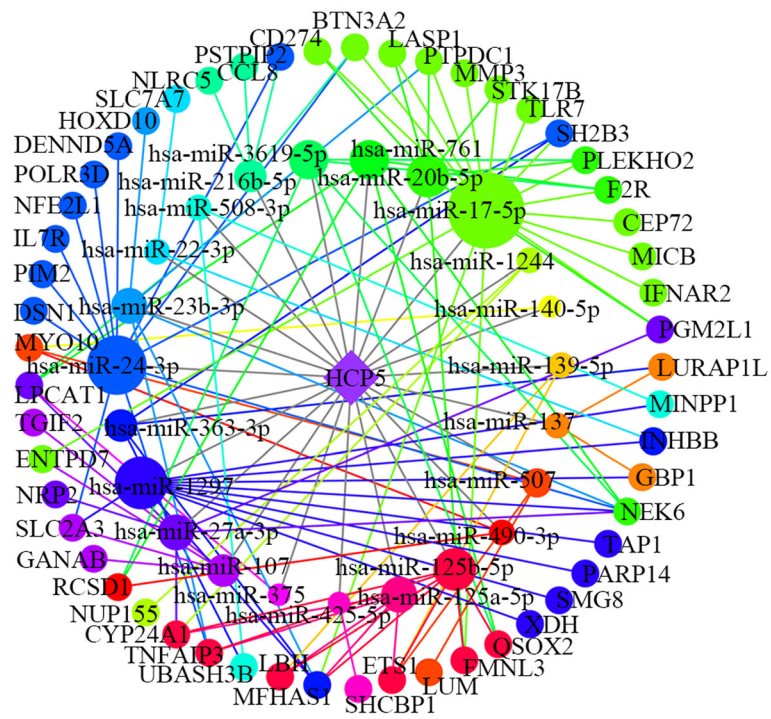

Fig. 6 The sub-network of hub lncRNA HCP5. The rhombuses represent lncRNAs, the circles on the inner loop represent miRNAs, and the circles on the outer loop represent mRNAs. The bigger size circles have, the more nodes they are connected to

BC values, suggesting that they may be potential key regulators controlling the oral cancer related ceRNA network.

Key lncRNA-miRNA-mRNA subnetwork

Based on the above-conducted analysis, we obtained three key lncRNAs: HCP5, AGAP11 and HCG22, which may play a role in oral cancer. These hub lncRNAs were used with their linked miRNAs and mRNAs to construct three more specific functional lncRNA-associated subnetworks. The AGAP11-associated subnetwork included 1 lncRNA, 19 miRNAs, 21 mRNAs and 48 edges (Fig. 5). As shown in Fig. 6, the subnetwork of HCP5 consisted of 1 lncRNA, 23 miRNAs, 53 mRNAs and 110 edges. As for HCG22, it interacted with 14 miRNAs and 34 mRNAs (Fig. 7).

To further understand the biological functions of the three hub lncRNAs, we performed GO functional enrichment analysis and KEGG pathway analysis for each hub-associated subnetwork. The results of the functional enrichment analysis revealed 14 enriched GO terms and 11 enriched pathway terms in the AGAP11-associated subnetwork (Fig. 8a), while there were 15 enriched GO terms and 6 enriched KEGG terms in the HCP5-associated subnetwork (Fig. 8b). Regarding the HCG22-associated subnetwork, there were 10 differentially
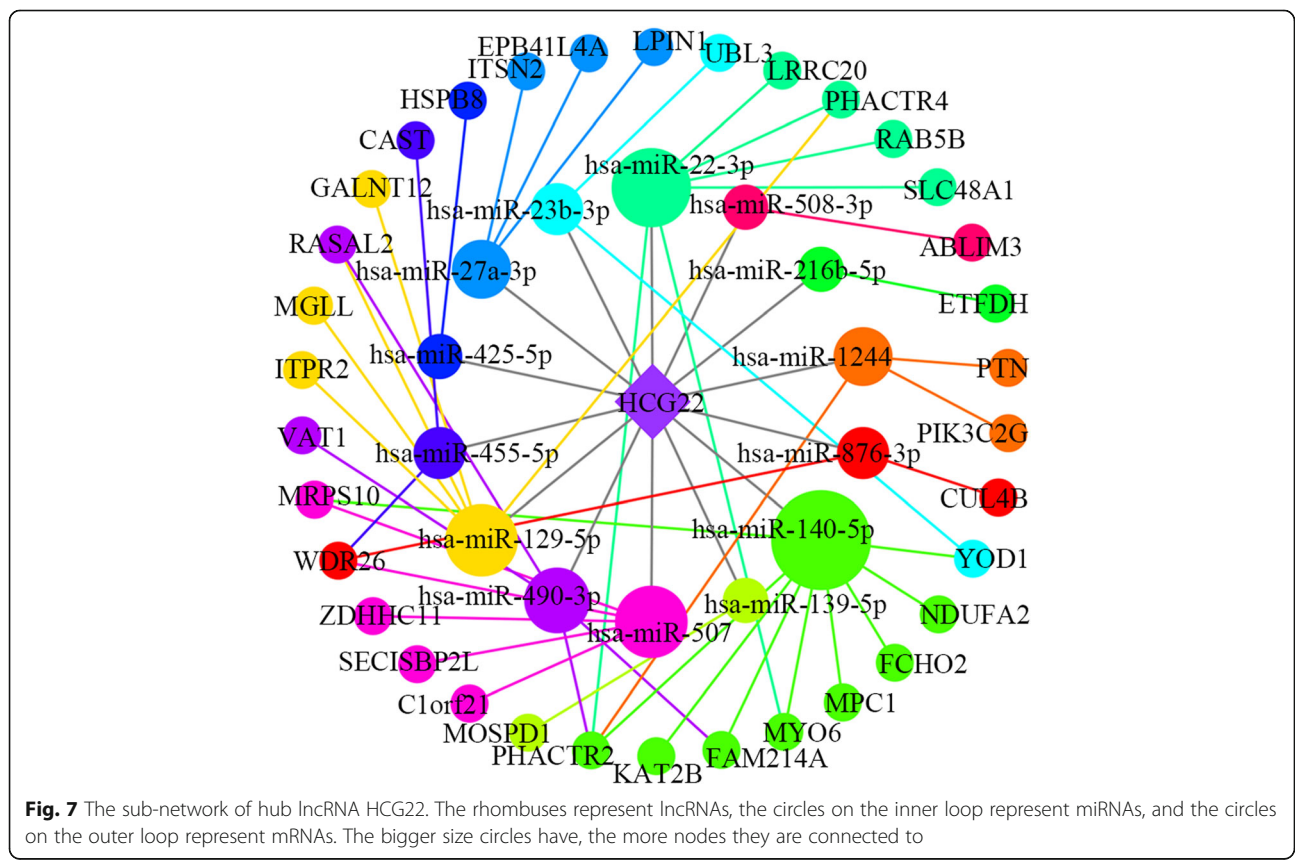
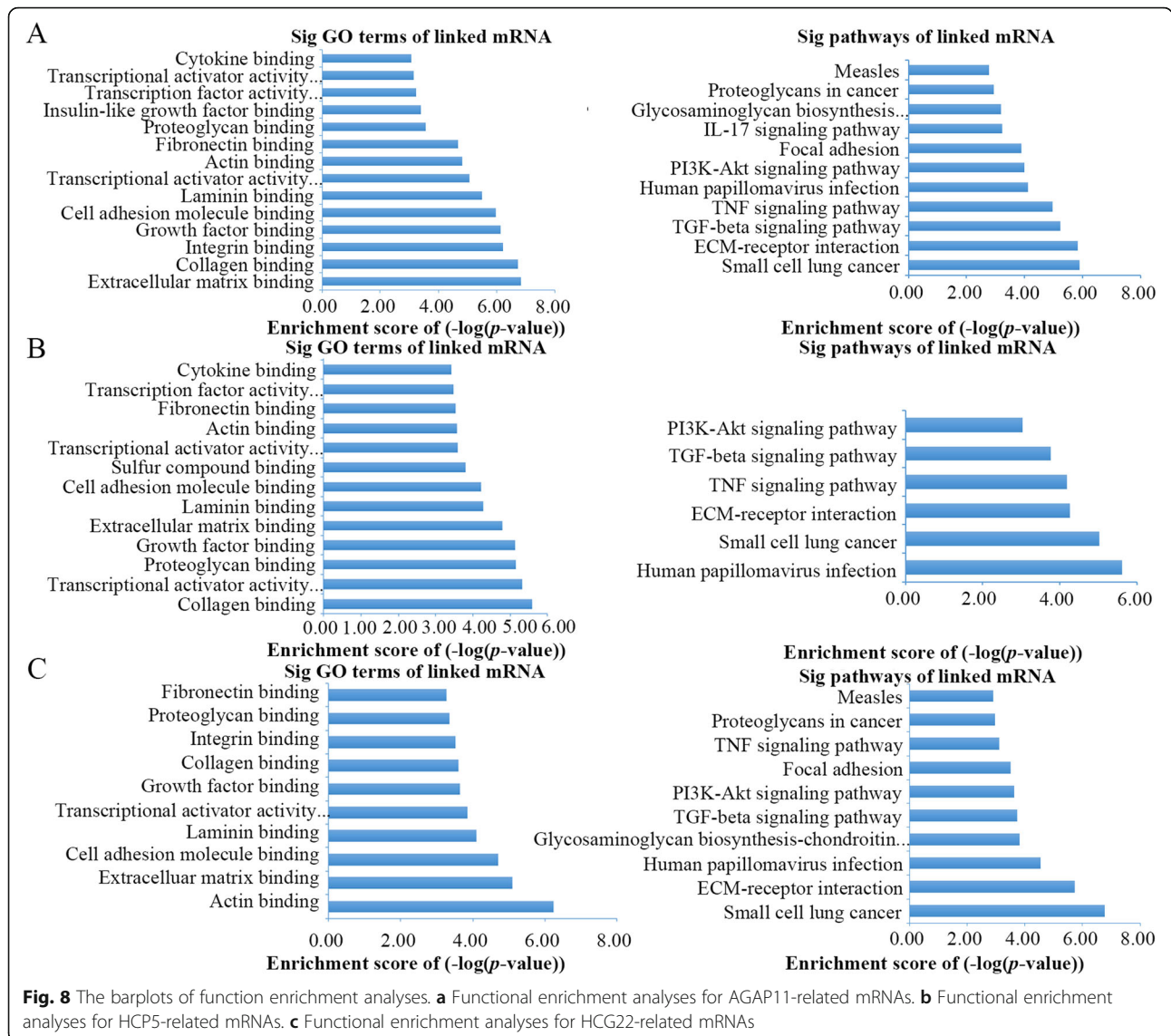enriched GO terms and 10 enriched KEGG pathways (Fig. 8c).

Differentially expressed RNAs in TCGA data

A total of 236 oral cancer samples and 44 normal samples were obtained after data preprocessing. To enhance the data reliability, genes with high expression values (data values were more than the third quartile) were filtered for further analysis. However, lncRNA AGAP11 was excluded. We identified 193 differentially expressed RNAs: 99 genes were downregulated and 94 were upregulated. The full list of differentially expressed RNAs was shown in Additional file 8.

Functional enrichment analysis

GO analysis of the up-regulated genes revealed that 17 enriched clusters were associated with biological processes (BP), 6 with cellular components (CC), and 9 with molecular function (MF) (Additional file 9). Among these terms, the top three enriched biological process were cell-cell adhesion, extracellular region and water transporter activity. Functional enrichment analysis of the down-regulated genes showed that 84 enriched clusters were associated with BP, 5 with CC, and 9 with MF (Additional file 10). Among them, 26 genes were

Fig. 7 The sub-network of hub lncRNA HCG22. The rhombuses represent lncRNAs, the circles on the inner loop represent miRNAs, and the circles on the outer loop represent mRNAs. The bigger size circles have, the more nodes they are connected to

Yin *et al. BMC Medical Genomics*   (2020) 13:84

Page 10 of 14

Fig. 8 The barplots of function enrichment analyses. **a** Functional enrichment analyses for AGAP11-related mRNAs. **b** Functional enrichment analyses for HCP5-related mRNAs. **c** Functional enrichment analyses for HCG22-related mRNAs

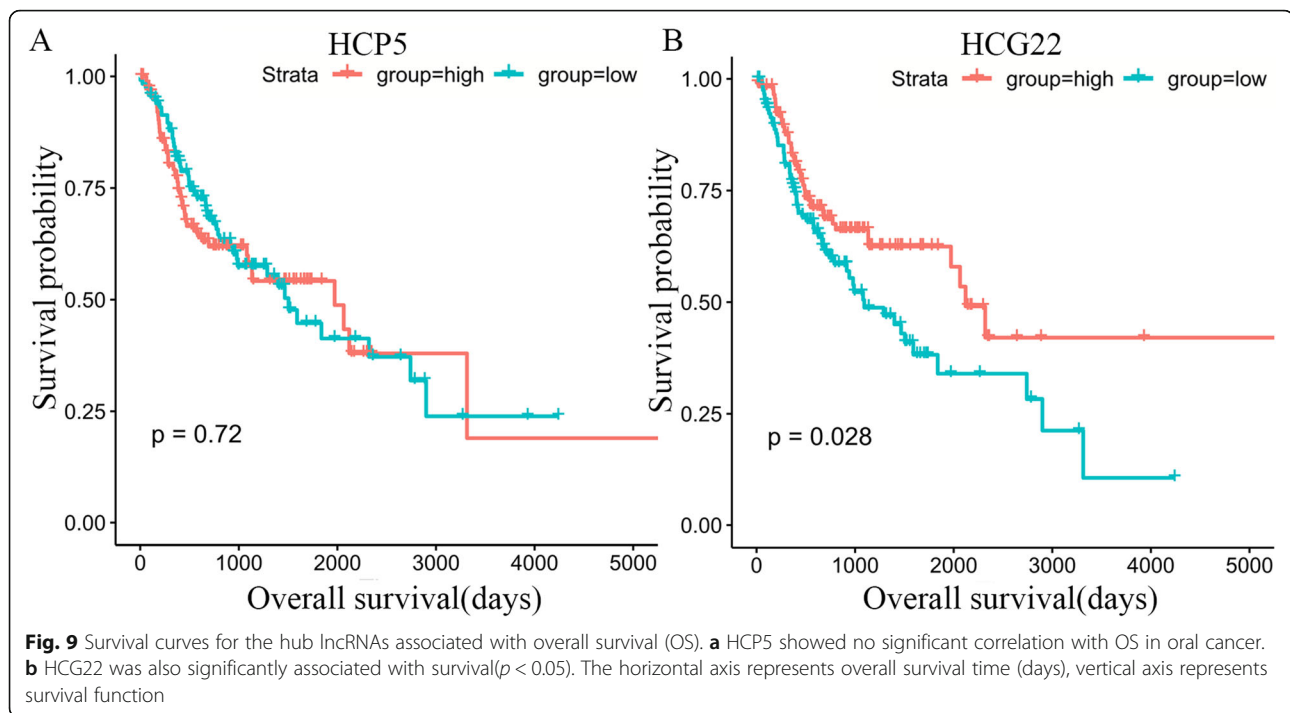enriched in sequence-specific DNA binding and represented the lowest FDR.

Survival analysis

Survival analysis was estimated based on Kaplan-Meier curve analysis. The difference was statistically significant with log-rank $P < 0.05$. As a result, HCG22 was significantly correlated with reduced survival time in patients with oral cancer (Fig. 9b), while HCP5 showed no significant correlation with overall survival in oral cancer (Fig. 9a).

Discussion

In this study, we used the expression profiles from the NCBI GEO to construct a lncRNA-associated network

of oral cancer based on the ceRNA theory. According to the results of the bioinformatics prediction and correlation analyses, the nodes of the network included 10 lncRNAs, 41 miRNAs and 122 mRNAs. Functional enrichment analysis was performed to reveal the biological functions of the mRNAs, the ceRNA counterparts of lncRNA [53]. As a result, three differentially enrichment GO terms (collagen binding [54], extracellular matrix binding [43] and cell adhesion molecule binding [45]) were associated with oral cancer. Moreover, four KEGG pathways (TGF-β signaling pathway [51], ECM-receptor interaction [44], TNF-signaling pathway [50] and PI3K-Akt signaling pathway [49]) showed to be related to the tumorigenesis of OSCC. It was shown that the nodes with a higher degree of connectivity to other nodes are

Fig. 9 Survival curves for the hub lncRNAs associated with overall survival (OS). **a** HCP5 showed no significant correlation with OS in oral cancer. **b** HCG22 was also significantly associated with survival($p < 0.05$). The horizontal axis represents overall survival time (days), vertical axis represents survival function

often more important in the network. By applying the topology analysis to the ceRNA network, we found three lncRNAs (HCP5, AGAP11 and HCG22) with significantly higher degrees and BC values compared with the other nodes, which means that these hubs are essential in the network organization and play a critical role in the ceRNA network. However, the average expression level of AGAP11 in oral cancer samples obtained from TCGA database was low. Subsequently, the prognostic value of each hub lncRNA was evaluated using Kaplan-Meier curve and Log-rank method. We found that HCG22 was positively correlated with overall survival and considered HCG22 as a key lncRNA responsible for the prognosis of oral cancer. Therefore, we supposed that HCG22 may play a more significant role in the pathogenesis and prognosis of oral cancer.

HCG22 (HLA complex group 22) is a long non-coding RNA gene that has been found to be down-regulated in oral cancer recently [55]. Low-expression of HCG22 has been confirmed to be associated with several types of diseases, including esophageal squamous cell carcinoma [56], bladder cancer [57], and steroid-induced ocular hypertension [58]. Consistent with another previous study [59], we identified that the low expression level of lncRNA HCG22 was associated with poor survival in oral cancer. However, there is no experimental evidence to support the contribution of HCG22 to the development of oral cancer. Based on the HCG22-associated subnetwork, we proposed that HCG22 might be an essential regulator in oral cancer by being a sponge for miRNAs.

Several miRNAs competed by HCG22 and mRNAs were associated with oral cancer. For example, miR-139-5p could induce oral cancer cell apoptosis through the Akt signaling pathway [36] and could be used as an effective biomarker to detect the tongue squamous cell carcinoma (TSCC) [60]. Another study has demonstrated that miR-140-5p targeted ADAM10 and inhibited the invasion and migration of the TSCC cells [42]. These studies may support our proposal of the regulatory function of HCG22 in oral cancer. Moreover, functional annotations of the 34 putative target mRNAs in the HCG22-miRNA-mRNA subnetwork revealed the biological functions of HCG22. It has been demonstrated that normal cell migration requires interactions with the extracellular matrix (ECM), which mainly includes collagens, laminins and fibronectin [44], while changes in the composition of ECM may contribute to the development and invasion of the oral cancer cells. In detail, it was observed that more fibrillary collagen type III than thick collagen type I existed in poorly-differentiated SCC compared with well-differentiated one [61]. When oral cancer cells invaded the connective tissue region from the basal membrane, a switch in the ECM's composition from a laminin-enriched environment to a collagen and fibronectin-enriched one would influence the metastatic and invasive behavior of the tumor cells, since the tumor formation was highly sensitive to the micro-environment [62]. Another GO term, regulation of transcription by RNA polymerase II, also had a connection with oral cancer. Xu et al. provided evidence that Histone acetylation and RNA polymerase II recruitment on the

integrin β6 promoter are involved in the TGF-β1-induced integrin β6 expression in OSCC cells, which would then promote tumorigenesis and metastasis [63]. Additionally, ten KEGG pathways, including the human papillomavirus infection, ECM-receptor interaction, small cell lung cancer, TGF-β signaling pathway, PI3K-Akt signaling pathway and TNF signaling pathway were determined. In accordance with the results from a previous report, the ECM-receptor interaction pathway was one of the most significantly altered pathways in the OSCC samples [64]. Tang et al. [50] indicated that TNF-α enhances the invasion and metastasis ability of the OSCC cells via the NF-Kb signaling pathway. The TGF-β/Smad pathway contributed to oral cancer tumorigenesis [51], and the PI3K-Akt signaling pathway was also considered to be important in the development of OSCC [49]. In summary, HCG22 may have the potential to become a novel biomarker for the detection and diagnosis of oral cancer.

Importantly, these results provided us with important information regarding the diagnostic and prognostic role of lncRNAs in oral cancer and pointed out lncRNA HCG22 as a candidate prognosis biomarker or potential therapeutic target. However, since no verification experiments were included in our study, the functional role of HCG22 still needs further investigation.

Conclusion
Overall, we constructed a lncRNA–miRNA–mRNA network based on the ceRNA theory, which enabled us to screen and analyze the lncRNAs that play functional roles in the progression of OSCC as miRNA sponges. Furthermore, we identified a hub lncRNA HCG22 in the complex ceRNA network. This study offered a unique insight into the ceRNA regulation network in OSCC and laid the foundation for further experimental and clinical research.

Supplementary information
Supplementary information accompanies this paper at https://doi.org/10.1186/s12920-020-00741-w.

Additional file 1. Data banks/repositories corresponding to all datasets analyzed in this study.

Additional file 2. TCGA datasets of all oral cancer samples obtained from the TCGA. The data included 305 oral cancer samples (14 samples of non-solid tumors were excluded) and 44 matched normal samples.

Additional file 3. Potential co-expression competing triples (lncRNA-miRNA-mRNA triples)

Additional file 4. Gene ontology (GO) terms interaction network. Yellow nodes mean nodes with *P*-value < 0.05 and Benjamini corrected *P*-value < 0.05.

Additional file 5. The enriched Gene ontology (GO) terms of linked mRNA in the ceRNA network.

Additional file 6. The enriched pathway terms of linked mRNA in the ceRNA network.

Additional file 7. All node degree analysis reveals the distribution of the points with different node degrees in ceRNA network.

Additional file 8. All differentially expressed genes identified between oral cancer tissues and matched normal tissues.

Additional file 9. The enriched Gene ontology (GO) terms of up-regulated genes in oral cancer.

Additional file 10. The enriched Gene ontology (GO) terms of down-regulated genes in oral cancer.

Abbreviations
BC: Betweenness centrality; BiNGO: Biological networks gene ontology tool; ceRNA: Competitive endogenous RNA; DEG: Differentially expressed gene; DEL: Differentially expressed lncRNA; DEM: Differentially expressed mRNA; ECM: Extracellular matrix; GEO: Gene expression omnibus database; GO: Gene ontology; KEGG: Kyoto encyclopedia of genes and genomes; RLE: Relative log expression; lncRNA: Long non-coding RNA; miRNA: microRNA; mRNA: Messenger RNAs; NCBI: National center for biotechnology information; ncRNA: Non-coding RNA; OSCC: Oral squamous cell carcinoma; PCC: Pearson's correlation coefficient; TCGA: The cancer genome atlas

Authors' contributions
JY: Drafted the manuscript, XZ: Bioinformatics analysis and data mining, ZA: Statistical analysis, MY: Literature review, YW: Literature review, SL: Edited the manuscript, corresponding author. All the authors read and approved the final manuscript.

Availability of data and materials
The datasets supporting the conclusions of this article were retrieved from the GEO repository (https://www.ncbi.nlm.nih.gov/geo/) with the accession number GSE74530 and the series matrix files were available at (ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSE74nnn/GSE74530/matrix/). The raw data for data quality assessment was available at (https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE74530&format=file). The full data table of microarray platform GPL570 was available at (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL570). The human comprehensive gene annotation we used to identify lncRNAs was available at (ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_34/gencode.v34.annotation.gtf.gz). The case IDs and direct web links of all oral cancer samples obtained from the TCGA data portal (http://portal.gdc.cancer.gov/; dbGaP Study Accession: phs000178) by Bioconductor package TCGAbiolinks were shown in Additional file 2. Data banks/repositories corresponding to all datasets analyzed in this study were listed in Additional file 1.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no potential conflicts of interest.

References
1. Wang B, Zhang S, Yue K, Wang XD. The recurrence and survival of oral squamous cell carcinoma: a report of 275 cases. Chin J Cancer. 2013;32(11):614–8.
2. Rivera C. Essentials of oral cancer. Int J Clin Exp Pathol. 2015;8(9):11884–94.

3. D'souza S, Addepalli V. Preventive measures in oral cancer: an overview. Biomed Pharmacother. 2018;107(3):72–80.

4. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.

5. Murugan AK, Munirajan AK, Tsuchida N. Ras oncogenes in oral cancer: the past 20 years. Oral Oncol. 2012;48(5):383–92.

6. Murugan AK, Hong NT, Cuc TTK, Hung NC, Munirajan AK, Ikeda M-A, Tsuchida N. Detection of two novel mutations and relatively high incidence of H-RAS mutations in Vietnamese oral cancer. Oral Oncol. 2009;45(10): e161–6.

7. Arunkumar G, Anand S, Raksha P, Dhamodharan S, Rao HPS, Subbiah S, Murugan AK, Munirajan AK. LncRNA OIP5-AS1 is overexpressed in undifferentiated oral tumors and integrated analysis identifies AS a downstream effector of stemness-associated transcription factors. Sci Rep. 2018;8(1):7018.

8. Curry JM, Sprandio J, Cognetti D, Luginbuhl A, Bar-ad V, Pribitkin E, Tuluc M. Tumor microenvironment in head and neck squamous cell carcinoma. Semin Oncol. 2014;41(2):217–34.

9. Momen-Heravi F, Bala S. Emerging role of non-coding RNA in oral cancer. Cell Signal. 2018;42:134–43.

10. Morris KV, Mattick JS. The rise of regulatory RNA. Nat Rev Genet. 2014;15(6): 423–37.

11. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009;136(2):215–33.

12. Lin S, Gregory RI. MicroRNA biogenesis pathways in cancer. Nat Rev Cancer. 2015;15(6):321–33.

13. ChengJ KP, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G, Sementchenko V, Piccolboni A, Bekiranov S, Bailey DK, Ganesh M, Ghosh S, Bell I, Gerhard DS, Gingeras TR. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science. 2005; 308(5725):1149–54.

14. Schmitt AM, Chang HY. Long noncoding RNAs in Cancer pathways. Cancer Cell. 2016;29(4):452–63.

15. Ma DD, Yuan LL, Lin LQ. Lncrna hotair contributes to the tumorigenesis of nasopharyngeal carcinoma via up-regulating FASN. Eur Rev Med Pharmacol Sci. 2017;21(22):5143–52.

16. Gao JF, Cao RM, Mu HL. Long non-coding RNA UCA1 may be a novel diagnostic and predictive biomarker in plasma for early gastric cancer. Int J Clin Exp Pathol. 2015;8(10):12936–42.

17. Misawa A, Takayama KI, Inoue S. Long non-coding RNAs and prostate cancer. Cancer Sci. 2017;108(11):2107–14.

18. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the rosetta stone of a hidden RNA language? Cell. 2011;146(3):353–8.

19. Poliseno L, Salmena L, Zhang JW, Carver B, Haveman WJ, Pandolfi PP, et al. Nature. 465(7301):1033–8.

20. Zhang LM, Ju HY, Wu YT, Guo W, Mao L, Ma HL, Xia WY, Hu JZ, Ren GX. Long non-coding RNA ANRIL promotes tumorgenesis through regulation of FGFR1 expression by sponging miR-125a-3p in head and neck squamous cell carcinoma. Am J Cancer Res. 2018;8(11):2296–310.

21. Arun K, Arunkumar G, Bennet D, Chandramohan SM, Murugan AK, Munirajan AK. Comprehensive analysis of aberrantly expressed lncRNAs and construction of ceRNA network in gastric cancer. Oncotarget. 2018;9(26): 18386–99.

22. Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002; 30(1):207–10.

23. Oghumu S, Knobloch TJ, Terrazas C, Varikuti S, Jarvis JA, Bollinger CE, Iwenofu H, Weghorst CM, Satoskar AR. Deletion of macrophage migration inhibitory factor inhibits murine oral carcinogenesis: potential role for chronic proinflammatory immune mediators. Int J Cancer. 2016;139(6):1379–90.

24. Heber S, Sick B. Quality assessment of Affymetrix GeneChip data. Omi A J Integr Biol. 2006;10(3):358–68.

25. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci. 2001;98(9):5116–21.

26. Ritchie ME, Phipson B, Wu D, Hu YF, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.

27. Jeggari A, Marks DS, Larsson E. MiRcode: a map of putative microrna target sites in the long non-coding transcriptome. Bioinformatics. 2012;28(15): 2062–3.

28. Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, Chan WL, Tsai WT, Chen GZ, Lee CJ, Chiu CM, Chien CH, Wu MC, Huang CY, Tsou AP, Huang HD. MiRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res. 2011;39(SUPPL.1):D163–9.

29. Wong N, Wang XW. MiRDB: an online resource for microRNA target prediction and functional annotations. Nucleic Acids Res. 2015;43(D1):D146–52.

30. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. eLife. 2015;4(AUGUST2015):e05005.

31. Liao Q, Liu CN, Yuan XY, Kang SL, Miao RY, Xiao H, Zhao GG, Luo HT, Bu DC, Zhao HT, Skogerbo G. Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. Nucleic Acids Res. 2011;39(9):3864–78.

32. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11): 2498–504.

33. Han JDJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, AJM W, Cusick ME, Roth FP, Vidal M. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature. 2004; 430(6995):88–93.

34. Yu G, Wang LG, Han YY, He QY. ClusterProfiler: an R package for comparing biological themes among gene clusters. Omi A J Integr Biol. 2012;16(5):284–7.

35. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G, Noushmehr H. TCGAbiolinks: an R/bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 2016;44(8):e71.

36. Ren YZ, Zhu HG, Chi CY, Yang FH, Xu X. MiRNA-139 regulates oral cancer Tca8113 cells apoptosis through Akt signaling pathway. Int J Clin Exp Pathol. 2015;8(5):4588–94.

37. Coutinho-Camillo CM, Lourenço SV, de Araújo LL, Kowalski LP, Soares FA. Expression of apoptosis-regulating miRNAs and target mRNAs in oral squamous cell carcinoma. Cancer Genet. 2015;208(7–8):382–9.

38. Mitra S, Mukherjee N, Das S, Das P, Panda CK, Chakrabarti J. Anomalous altered expressions of downstream gene-targets in TP53-miRNA pathways in head and neck cancer. Sci Rep. 2014;4(1):6280.

39. Jia LF, Huang YP, Zheng YF, Lyu MY, Zhang CN, Meng Z, Gan YH, Yu GY. MiR-375 inhibits cell growth and correlates with clinical outcomes in tongue squamous cell carcinoma. Oncol Rep. 2015;33(4):2061–71.

40. Cheng CM, Shiah SG, Huang CC, Hsiao JR, Chang JY. Up-regulation of miR-455-5p by the TGF-β–SMAD signalling axis promotes the proliferation of oral squamous cancer cells by targeting UBE2B. J Pathol. 2016;240(1):38–49.

41. Kozaki K, Imoto I, Mogi S, Omura K, Inazawa J. Exploration of tumor-suppressive MicroRNAs silenced by DNA Hypermethylation in Oral Cancer. Cancer Res. 2008;68(7):2094–105.

42. Yang K, Wang P, Wu L, Hao JB, Bian Z. Reciprocal effects between microRNA-140-5p and ADAM10 suppress migration and invasion of human tongue cancer cells. Biochem Biophys Res Commun. 2014;448(3):308–14.

43. Tsuneki M, Maruyama S, Yamazaki M, Cheng J, Saku T. 8520 POSTER podoplanin regulates the proliferation of oral squamous cell carcinoma cells via its binding to extracellular matrix. Eur J Cancer. 2011;47:S550.

44. Ziober AF, Falls EM, Ziober BL. The extracellular matrix in oral squamous cell carcinoma: friend or foe? Head Neck. 2006;28(8):740–9.

45. Lyons AJ, Jones J. Cell adhesion molecules, the extracellular matrix and oral squamous carcinoma. Int J Oral Maxillofac Surg. 2007;36(8):671–9.

46. Xu HF, Raynal N, Stathopoulos S, Myllyharju J, Farndale RW, Leitinger B. Collagen binding specificity of the discoidin domain receptors: binding sites on collagens II and III and molecular determinants for collagen IV recognition by DDR1. Matrix Biol. 2011;30(1):16–26.

47. Hidalgo-Carcedo C, Hooper S, Chaudhry SI, Williamson P, Harrington K, Leitinger B, Sahai E. Collective cell migration requires suppression of actomyosin at cell–cell contacts mediated by DDR1 and the cell polarity regulators Par3 and Par6. Nat Cell Biol. 2011;13(1):49–59.

48. Lakshminarayana S, Augustine D, Rao RS, Patil S, Awan KH, Venkatesiah SS, Haragannavar VC, Nambiar S, Prasad K. Molecular pathways of oral cancer that predict prognosis and survival: a systematic review. J Carcinog. 2018;17(1):7.

49. Smolensky D, Rathore K, Bourn J, Cekanova M. Inhibition of the PI3K/AKT pathway sensitizes oral squamous cell carcinoma cells to anthracycline-based chemotherapy in vitro. J Cell Biochem. 2017;118(9):2615–24.

50. Tang DF, Tao DT, Fang Y, Deng C, Xu Q, Zhou JP. TNF-alpha promotes invasion and metastasis via NF-kappa B pathway in oral squamous cell carcinoma. Med Sci Monit Basic Res. 2017;23:141–9.

51. Wu FL, Weigel KJ, Zhou HM, Wang XJ. Paradoxical roles of TGF-β signaling in suppressing and promoting squamous cell carcinoma. Acta Biochim Biophys Sin. 2018;50(1):98–105.

52. Jiang H, Ma R, Zou SB, Wang YZ, Li ZQ, Li WP. Reconstruction and analysis of the lncRNA-miRNA-mRNA network based on competitive endogenous RNA reveal functional lncRNAs in rheumatoid arthritis. Mol BioSyst. 2017; 13(6):1182–92.

53. Song C, Zhang J, Liu Y, Pan H, Qi HP, Cao YG, Zhao JM, Li S, Guo J, Sun HL, Li CQ. Construction and analysis of cardiac hypertrophy-associated lncRNA-mRNA network based on competitive endogenous RNA reveal functional lncRNAs in cardiac hypertrophy. Oncotarget. 2016;7(10):10827–40.

54. Moilanen JM, Löffek S, Kokkonen N, Salo S, Väyrynen JP, Hurskainen T, Manninen A, Riihilä P, Heljasvaara R, Franzke CW, Kähäri VM, Salo T, Mäkinen MJ, Tasanen K. Significant role of collagen XVII and integrin β4 in migration and invasion of the less aggressive squamous cell carcinoma cells. Sci Rep. 2017;7(1):45057.

55. Feng L, Houck JR, Lohavanichbutr P, Chen C. Transcriptome analysis reveals differentially expressed lncRNAs between oral squamous cell carcinoma and healthy oral mucosa. Oncotarget. 2017;8(19):31521–31.

56. Li XZ, Xiao XX, Chang RM, Zhang CF. Comprehensive bioinformatics analysis identifies lncRNA HCG22 as a migration inhibitor in esophageal squamous cell carcinoma. J Cell Biochem. 2020;121(1):468–81.

57. Jiang D, Zhang YY, Yang LW, Lu WZ, Mai L, Guo HX, Liu XL. Long noncoding RNA HCG22 suppresses proliferation and metastasis of bladder cancer cells by regulation of PTBP1. J Cell Physiol. 2020;235(2):1711–22.

58. Jeong S, Patel N, Edlund CK, Hartiala J, Hazelett DJ, Itakura T, Wu PC, Avery RL, Davis JL, Flynn HW, Lalwani G, Puliafito CA, Wafapoor H, Hijikata M, Keicho N, Gao XY, Argüeso P, Allayee H, Coetzee GA, Pletcher MT, Conti DV, Schwartz SG, Eaton AM, Fini ME. Identification of a novel mucin gene HCG22 associated with steroid-induced ocular hypertension. Investig Ophthalmol Vis Sci. 2015;56(4):2737–48.

59. Nohata N, Abba MC, Gutkind JS. Unraveling the oral cancer lncRNAome: identification of novel lncRNAs associated with malignant progression and HPV infection. Oral Oncol. 2016;59:58–66.

60. Duz MB, Karatas OF, Guzel E, Turgut NF, Yilmaz M, Creighton CJ, Ozen M. Identification of miR-139-5p as a saliva biomarker for tongue squamous cell carcinoma: a pilot study. Cell Oncol. 2016;39(2):187–93.

61. Kumari K, Ghosh S, Patil S, Augustine D, Venkatesiah SS, Rao RS. Expression of type III collagen correlates with poor prognosis in oral squamous cell carcinoma. J Investig Clin Dent. 2017;8(4):e12253.

62. De Oliveira RG, Bernardi L, Lauxen I, Filho M, Horwitz A, Lamers M. Fibronectin modulates cell adhesion and signaling to promote single cell migration of highly invasive oral squamous cell carcinoma. PLoS One. 2016; 11(3):e0151338.

63. Xu MY, Yin LQ, Cai YH, Hu QW, Huang J, Ji Q, Hu YP, Huang WX, Liu F, Shi SL. Epigenetic regulation of integrin β6 transcription induced by TGF-β1 in human oral squamous cell carcinoma cells. J Cell Biochem. 2018;119(5): 4193–204.

64. Li SM, Chen XJ, Liu XQ, Yu Y, Pan HY, Haak R, Schmidt J, Ziebolz D, Schmalz G. Complex integrated analysis of lncRNAs-miRNAs-mRNAs in oral squamous cell carcinoma. Oral Oncol. 2017;73:1–9.

Publisher's Note