# Comprehensive analysis of an mRNA co-expression network and a ceRNA network reveals potential prognostic biomarkers in oral squamous cell carcinoma

Liming He[1†], Zhisheng Jiang[2†], Yijun Gao[3], Yiyu Zeng[3], Wenhui Ge[3], Yi Yu[3] and Xiaoyan Xie[3*]

## Abstract

**Background**  Oral squamous cell carcinoma (OSCC) is a prevalent and aggressive oral cancer with a poor prognosis. Its polygenic risk is likely influenced by complex transcriptional disorders involving networks of co-expressed and functionally related genes, though such investigations are limited in OSCC.

**Methods**  We analyzed the GSE37991 dataset, comprising 40 OSCC and 40 normal oral tissue samples from the Gene Expression Omnibus. Tumor-specific modules were identified using weighted correlation network analysis (WGCNA), leading to the selection of hub mRNAs and lncRNAs. These lncRNAs were used to construct lncRNA–mRNA and competing endogenous RNA networks. We further examined the expression profiles and survival data of these genes from the Cancer Genome Atlas. Prognostic markers were identified and validated through 5-year survival analysis and Cox proportional hazards modeling. RT-qPCR was used to validate the expression levels in clinical OSCC tissues.

**Results**  We identified 1847 differentially expressed genes in OSCC tissues. WGCNA revealed four OSCC-specific modules, screening 120 hub mRNAs and five hub lncRNAs. Two prognostic markers (AQP5, IL-26) from hub mRNAs and three (FRMD5, INHBB, GUCY1A3) from the lncRNA–mRNA network were associated with survival. Validation showed lower expression of AQP5 and GUCY1A3, and higher expression of FRMD5 and INHBB in OSCC compared to normal tissues.

**Conclusion**  This study enhances our understanding of transcriptional dysregulation in OSCC and may highlights AQP5, IL-26, FRMD5, INHBB, and GUCY1A3 as promising prognostic biomarkers.

**Keywords**  Hub gene, Hub lncRNAs, Oral squamous cell carcinoma, Survival, WGCNA

[†]Liming He and Zhisheng Jiang have contribute equally to this work.

*Correspondence:
Xiaoyan Xie
xyxie@csu.edu.cn
Full list of author information is available at the end of the article

He *et al. Egyptian Journal of Medical Human Genetics*     (2024) 25:100

Page 2 of 13

## Introduction

Oral squamous cell carcinoma (OSCC) is one of the most aggressive malignancies, characterized by a high incidence rate, locoregional metastases, and resistance to existing treatments. These factors contribute to a poor prognosis for patients with OSCC [1–3]. OSCC is highly prevalent and primarily linked to tobacco and excessive alcohol use and infection with the human papillomavirus (HPV) [4–8]. The American cancer society estimates for oral cavity and oropharyngeal cancers in the United States for 2024 are about 58,450 new cases and 12,230 deaths. Squamous cell carcinoma (SCC) accounts for 90% of head and neck cancers, originating from the epithelial lining of the oral cavity, pharynx, and larynx. OSCC is more prevalent in men and adults over 50, with the highest incidence in South and Southeast Asia due to area nut consumption. The global incidence of OSCC is rising, especially among younger populations, with a projected 30% annual increase by 2030, driven by lifestyle changes and the growing prevalence of HPV-related oropharyngeal cancer. Various factors, from unhealthy behaviors to viral infections, contribute to the complexity and heterogeneity of OSCC pathogenesis. While, large-scale genomics [9], transcriptomics [10], and epigenomics [11] have identified potential molecular targets, further investigation is warranted to explore the networks of co-expressed genes associated with OSCC prognosis, given its nature as a multi-step chronic disease.

In the era of big data, high-throughput data analysis and prioritized information network screening are emerging trends in cancer research. Unlike previous studies focusing on individual molecules and pathways, high-throughput data analysis emphasizes networks of co-expressed and functionally related genes. Various tools and algorithms, such as weighted correlation network analysis (WGCNA), have been developed to identify key targets. WGCNA, a systemic biological algorithm, constructs co-expression gene networks using high-throughput gene expression profiles to identify highly synergistic gene sets [12]. This tool effectively compares differentially expressed genes (DEGs) and identifies specific genes in cancer tissues, facilitating the exploration of gene interactions in functional networks throughout cancer progression. WGCNA has been widely applied to various cancers [13–16], including OSCC [17].

With the vast amounts of RNA data uncovered through contemporary biological research, new terms and concepts, such as competing endogenous RNAs (ceRNAs), continually emerge. CeRNAs represent a regulatory mechanism describing the interactions between RNAs, including pseudogene transcripts, long noncoding RNAs (lncRNAs), circular RNAs (circRNAs), and mRNAs. They regulate gene expression by competing with microRNA response elements (MREs) [18]. Dysregulation of the ceRNA network is intricately linked to the initiation and progression of various cancers [19–21]. Investigating ceRNA interactions offers novel insights into cancer pathogenesis and potential therapeutic approaches.

In this study, we used a comprehensive and stepwise approach to uncover the potential molecular mechanisms of OSCC progression. Initially, we utilized GSE37991 microarray data to screen DEGs between tumor tissues and normal controls, identifying their biological functions and related pathways through gene ontology (GO) and Kyoto Encyclopedia of genes and genomes (KEGG) analyses. We then applied WGCNA to identify tumor-specific modules and their core genes. Within these modules, we focused on lncRNAs to construct a competing ceRNA co-expression network. We meticulously analyzed the interactions between key lncRNAs and mRNAs, shedding light on the roles of lncRNAs in OSCC oncogenesis. To evaluate, clinical relevance, we conducted survival analysis and univariate and multivariate Cox proportional hazard regression analyses on the screened mRNAs and lncRNAs. Finally, we validated the expression levels of hub genes in clinical OSCC patient tissues. The workflow of this study is shown in Figure S1. This study may led a light on the five hub genes in the pathogenesis of OSCC, and might be beneficial to further molecular mechanisms investigation and development of targeting therapeutics for OSCC.

## Materials and methods

### Data acquisition and preprocessing

The expression profile GSE37991 was downloaded from the GEO database (www.ncbi.nlm.nih.gov/geo/), and was measured using Affymetrix Human Genome U133 Plus 2.0 Array (Affymetrix, Santa Clara, CA, USA). GSE37991 contains 40 OSCC samples and 40 normal oral tissues samples. Background adjustment, quantile normalization, and summarization of the raw data were conducted using the R software Affy package [22]. The R software limma package [23] was used to identify DEGs using the empirical Bayes method. Genes with $p < 0.05$ and |log (fold change, FC)|>1 were considered statistically significant DEGs.

### Functional enrichment analysis of DEGs

To understand the biological significance of the up-regulated and down-regulated genes, we used the database for annotation, visualization and integrated discovery (DAVID) for gene ontology (GO) analyses, covering biological process (BP), cellular component (CC), and molecular function (MF), as well as KEGG enrichment analyses [24]. Enrichment was considered significant when $p < 0.05$.

## WGCNA module construction and visualization

WGCNA is a systems biology algorithm that identifies correlation patterns among genes in high-throughput data, effectively revealing modules and key hub genes [12]. In this study, we selected genes with $p < 0.05$ and FC > 1.5 or FC < 0.667 for WGCNA to reduce computational complexity and enhance data signals without losing important information [25]. After screening for outliers in cancerous and normal tissues using WGCNA, we constructed co-expression networks using soft thresholds (scale-free network structure $R^2 = 0.85$) for both cancerous and normal samples. Cancer set-specific modules were selected for further analysis. The top 30 genes with the highest intramodular connectivity within these modules were considered hub genes, and weighted gene co-expression networks was constructed by Cytoscape [26].

## Construction of lncRNA–mRNA co-expression network

Human lncRNA data were sourced from the HUGO Gene Nomenclature Committee (HGNC). LncRNAs within each module were filtered by Entrez ID. Co-expression relationships between lncRNAs and mRNAs were used to construct lncRNA–mRNA co-expression network, which was visualized with Cytoscape.

## Construction of ceRNA network

To explore the role of co-expression relationships in the ceRNA network, we constructed a ceRNA network based on the lncRNA–mRNA co-expression network. LncRNA–microRNA interactions were predicted using LncBase Predicted v.2 (http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?r=lncbasev2%2Findex-predicted) with DIANA software (http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php) (only scores > 0.7 were considered) [27]. MicroRNA–mRNA interactions were predicted using starBase v2.0 (http://starbase.sysu.edu.cn/starbase2/index.php) [28]. The resulting networks were visualized with Cytoscape.

## Survival analysis and univariate analysis

Fragments per kilobase million (FPKM) data for head and neck tumors were obtained from TCGA (https://portal.gdc.cancer.gov/) database. A 5-year survival analysis was performed after log2 conversion including 120 hub genes (top 30 genes from each tumor-specific module) and genes from the lncRNA–mRNA co-expression networks. Cases with overall survival (OS) > 1 month were analyzed to avoid including patients with advanced tumors at diagnosis, which may skew results. Patients were classified into high and low expression groups based on median gene expression levels, and genes detected in < 50% of

samples were excluded. Univariate Cox regression analysis, conducted with the survival package [29], identified genes with prognostic value.

## Cox proportion hazard model multivariate analysis

To refine the prognostic gene signature, Cox Proportional Hazards Model Multivariate Analysis was performed using the survival R package. Both hub genes and genes from the lncRNA–mRNA co-expression networks were analyzed separately. The final model, containing only significant prognostic genes ($p < 0.05$), was visualized using the survminer R package [30]. The analysis started with all included genes and sequentially eliminated those not statistically significant.

## Quantitative real-time PCR (qPCR)

Tumor blocks are surgically derived from the primary OSCC site in diagnosed patients. Normal tissues are obtained and histologically confirmed from the patient's negative surgical margin, ensuring the absence of cancer cells. After homogenizing the OSCC tissue with a homogenizer, total RNA was extracted using Trizol reagent (Invitrogen). Reverse transcription was performed with the PrimeScript RT kit with gDNA Eraser (Takara Bio, Shiga, Japan). Gene expression was measured using a fast two-step amplification program with 2 × SYBR Green Fast qPCR master mix (Biotool). The primer sequences used are listed (5′-3′): AQP5-F, TACGGTGTGGCACCGCTCAATG, R, AGTCAGTGGAGGCGAAGATGCA; IL-26-F, GGAAGACGTTTTTGGTCAACTGC, R, CTCTCTAGCTGATGAAGCACAGG; FRMD5-F, CATCGGATGCTGCCTTGTTAGC, R, GGAACTTGGAGCTGTAGCCTTC; INHBB-F, GAAATCATCAGCTTCGCCGAGAC, R, GGCAGGAGTTTCAGGTAAAGCC; GUCY1A3-F, GCTCTTCTCAGACATCGTTGGG, R, ATAGGCATCGCCAATGGTCTCC; GAPDH-F, GTCTCCTCTGACTTCAACAGCG, R, ACCACCCTGTTGCTGTAGCCAA.

## Results

### Identification of DEGs and gene functional enrichment analysis

A total of 1847 DEGs were identified between OSCC and normal tissue samples, with 697 up-regulated and 1150 down-regulated (Figure S2). GO and KEGG enrichment analyses of these genes were performed using DAVID, with the top ten terms ($p < 0.05$) were shown in Fig. 1. GO analyses revealed that up-regulated genes were mainly associated with processes crucial to development, such as collagen and extracellular matrix (ECM) tissue decomposition, cell adhesion, and the inflammation (Fig. 1A–C). Down-regulated genes were primarily related to muscle contraction and exogenous metabolic processes
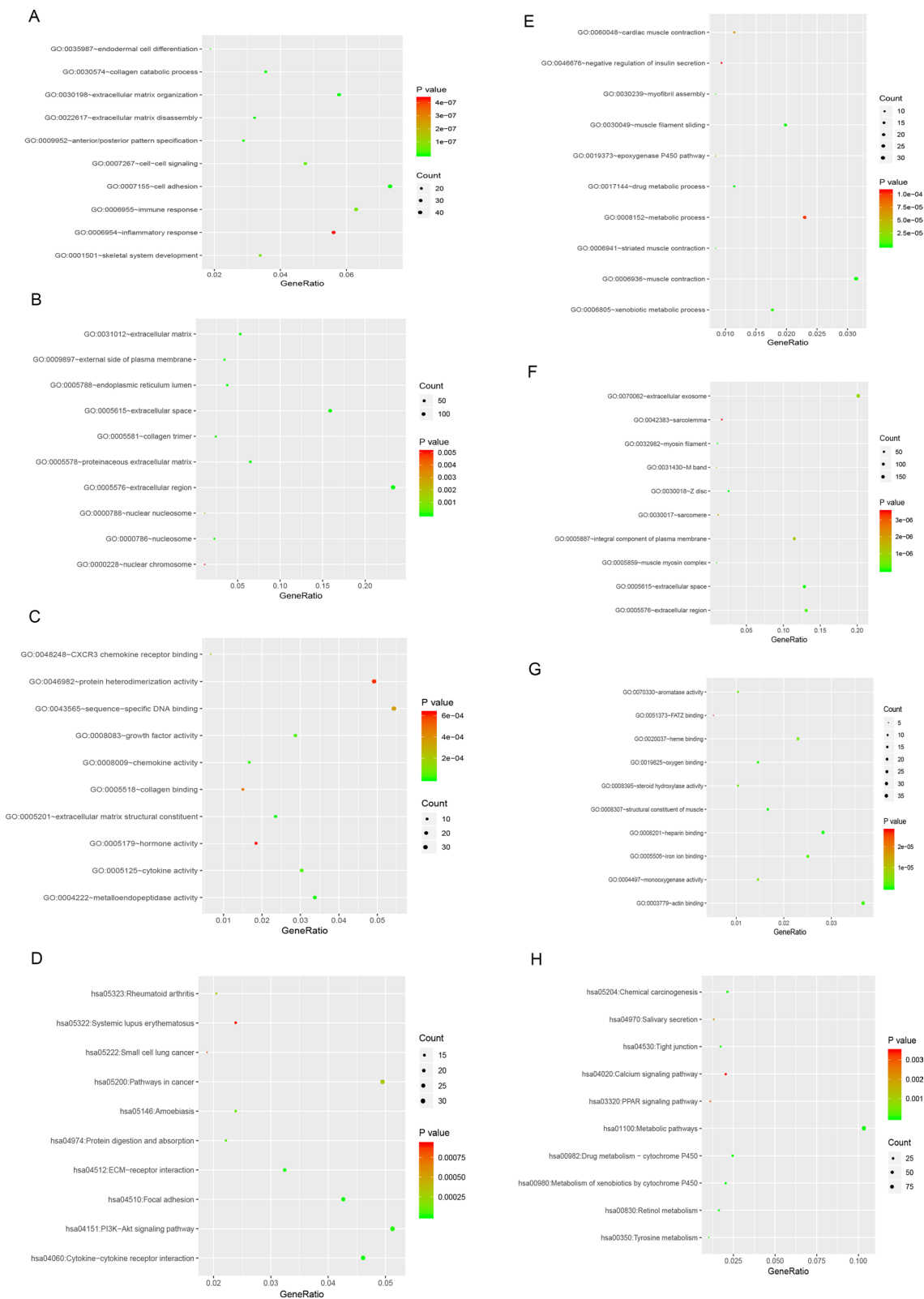
He *et al. Egyptian Journal of Medical Human Genetics*       (2024) 25:100

Page 4 of 13



**Fig. 1** Gene enrichment analysis of up-regulated and down-regulated DEGs. GO enrichment analysis of up-regulated DEGs: **A** Biological Process (BP); **B** Cellular Component (CC); **C** Molecular Function (MF). KEGG analysis of up-regulated DEGs: **D**. GO enrichment analysis of down-regulated DEGs: **E** BP; **F** CC; **G** MF. KEGG analysis of down-regulated DEGs: **H**. The *x*-axis shows the GeneRatio of each term and the *y*-axis shows the GO and KEGG pathway terms

(Fig. 1E–G). KEGG pathway analysis showed that up-regulated genes were enriched in pathways related to ECM receptor interaction, PI3K-AKT signaling, cytokines-receptor interactions, amoebiasis, rheumatoid arthritis and systemic lupus erythematosus (Fig. 1D). Down-regulated genes were enriched in pathways such as salivary secretion, tight junctions, calcium signaling, PPAR signaling, metabolic pathways, and cytochrome P450 enzymes drug metabolism, retinol and tyrosine metabolism (Fig. 1H).

### Construction and analysis of WGCNA cluster tree
WGCNA analysis was performed using genes with $|\log(FC)| > 0.585$ and $p < 0.05$. For the scale-free network structure, soft thresholds were chosen when the initial $R^2$ value exceeded 0.9 (Fig. 2A for normal and Fig. 2B for tumor). Co-expression modules were calculated, with different colors representing gene modules and gray indicating unmerged genes. Seventeen gene modules were identified in normal tissue (Fig. 2A) using a soft threshold of 11 and a minimum module size of 30. Twenty gene modules were identified in OSCC tissue (Fig. 2B) using a soft threshold of six and a minimum module size of 30.

Additionally, we conducted a repeated analysis within each module of the two cluster trees (Fig. 3A). The OSCC modules in light cyan, green, cyan, and salmon did not overlap with any normal modules, suggesting that these modules are specific to OSCC. Using intramodular connectivity to identify key genes, we selected the top 30 as hub genes and visualized them with Cytoscape (Fig. 3B,C).

### Selection of lncRNAs from tumor-specific modules
We selected lncRNAs from the hub sub-network to construct a lncRNA–mRNA co-expression network (Fig. 4A–D). The following lncRNAs were chosen for further analysis: LOC441426 (cyan module), TTTY2 (light cyan module), PART1 (salmon module), MEG3 and C10orf85 (green module). These lncRNAs were used to build the lncRNA–mRNA co-expression network.

### Construction of a ceRNA network
CeRNAs regulate microRNA-induced gene silencing by binding to microRNAs through microRNA response elements (MREs), thus serving as a communication network among RNA transcripts. We predicted interactions between lncRNAs and microRNAs, as well as between microRNAs and mRNAs. Overlapping predictions within these interactions and the lncRNA–mRNA co-expression network were identified, and an lncRNA–microRNA–mRNA ceRNA network was constructed (Fig. 5).

The ceRNA network helps determine whether hub lncRNAs influence hub genes and their pathways. For instance: PART1 in the salmon module acts as a ceRNA for miR-495-3p, miR-544a, miR-433-3p, miR-224-5p, miR-335-5p, miR-590-3p, miR-448, and miR-128-3p, targeting PANK1, and miR-149-5p, targeting SLC38A3. C10orf85 in the green module acts as a ceRNA for miR-27b-3p and miR-147a, targeting ITGA11. MEG3 serves as a ceRNA for 19 miRNAs, including miR-4500, miR-27b-3p, miR-122-5p, miR-372-3p, and miR-202-3p, also targeting ITGA11. These interactions provide new insights into the tumorigenesis of OSCC.

### Survival analysis
First, we selected 120 hub genes (the top 30 genes from each tumor-specific module) and all the genes in the lncRNA–mRNA co-expression networks, excluding those not detected in over 50% of samples. Survival analysis was performed to identify genes with prognostic significance. The analysis included hub genes (AQP5, CYP4X1, RSPO1, IL-26, CYP4Z1, PCP2, BCMO1, GUCY1A3, and CLK4) and lncRNA–mRNA network genes (RSPO1, FAP, FRMD5, TTTY14, SPP1, ZNF566, CEECAM1, INHBB, CLK4, OPN3, GUCY1A3, BCO1, TNFAIP6, TMEM56, FN1, and GFPT2). AQP5, IL-26, FRMD5, INHBB, and GUCY1A3 were found to be statistically significant prognostic genes (Fig. 6A).

### Expression of five hub genes in clinical OSCC tissues
We analyzed the expression levels of five hub genes in three normal and four OSCC tissue samples. The results are shown in Fig. 6B. AQP5 and GUCY1A3 were found to be expressed at lower levels in OSCC tissues as compared to the normal samples, while FRMD5 and INHBB showed higher expression levels in OSCC samples. IL-26 did not exhibit a significant difference between OSCC and normal tissues. These findings suggest potential new marker genes for OSCC tumorigenesis.

## Discussion
OSCC is a prevalent and aggressive malignancy with a 5-year survival rate of 50% [31–34]. Despite advancements in surgical methods, radiotherapy, and chemotherapy improving patient quality of life, the prognosis for OSCC remains poor [2]. Thus, exploring its molecular mechanisms is crucial for developing effective prevention and treatment strategies. In this study, WGCNA was employed to identify OSCC-specific modules, revealing 120 hub mRNAs and five hub lncRNAs. We constructed lncRNA–mRNA and ceRNA co-expression networks using these hub lncRNAs. By analyzing FPKM data from the TCGA database and applying survival analysis with the Cox proportional hazard model, we validated these hub mRNAs and lncRNA–mRNA network genes. Notably, AQP5 and IL-26 (from hub mRNAs), and FRMD5,
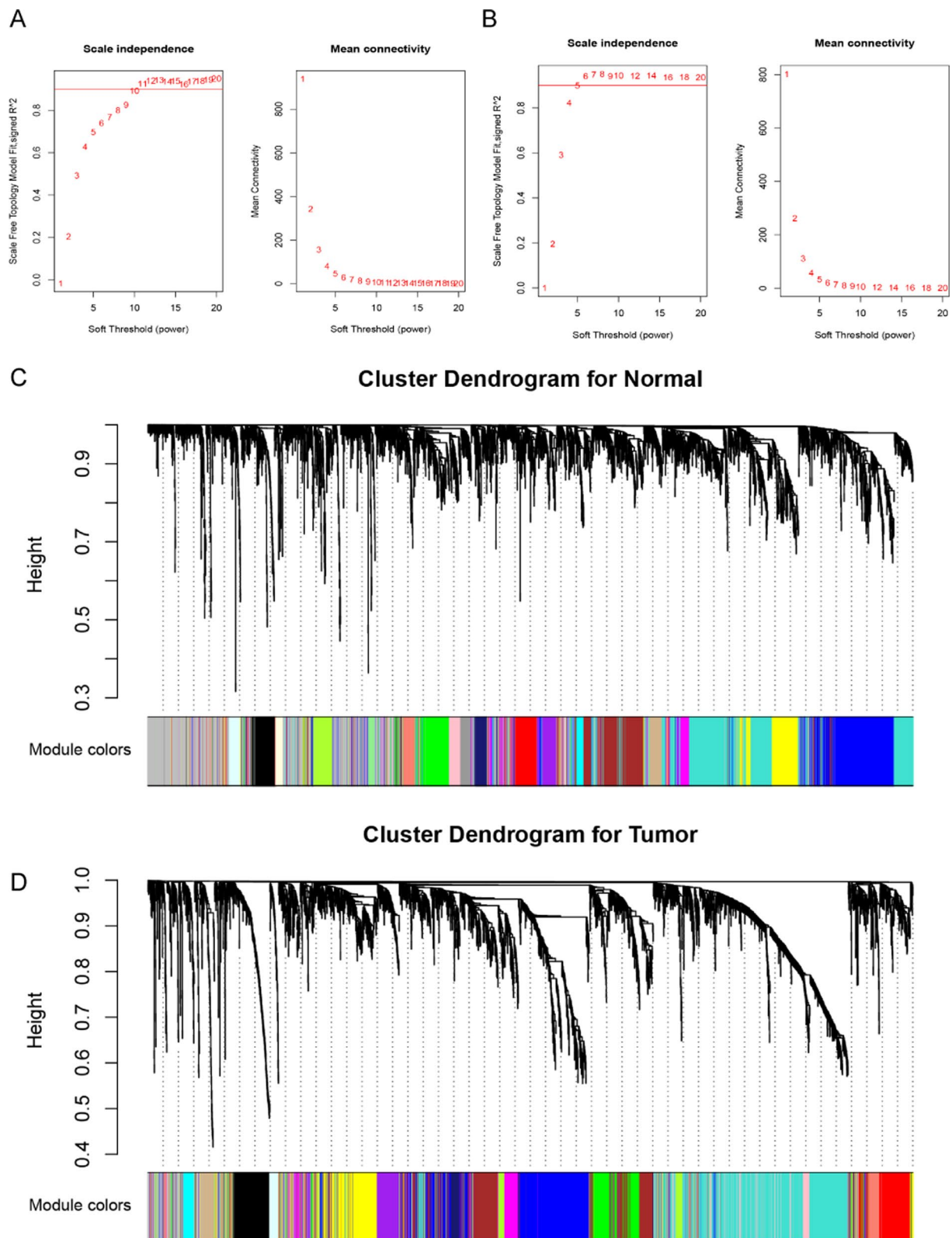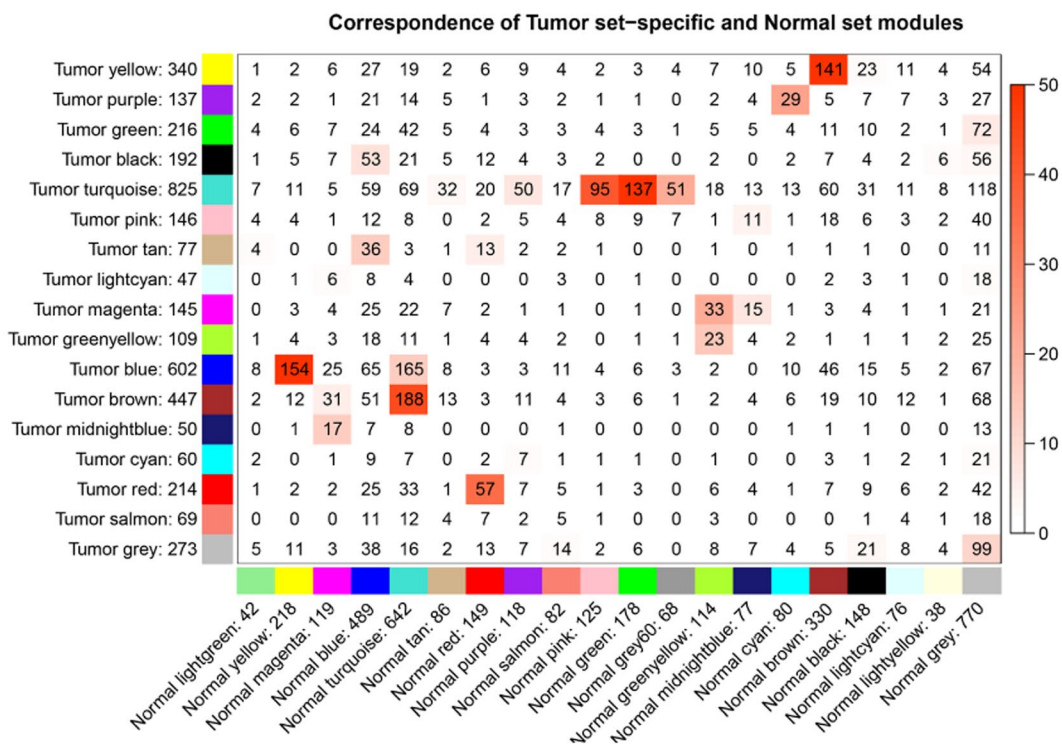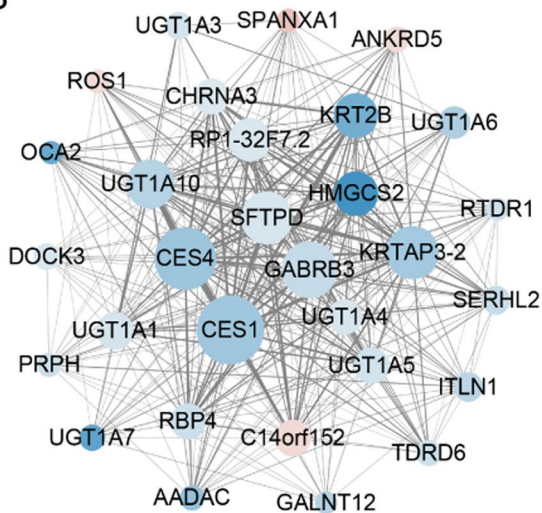
**Fig. 2** WGCNA analysis. Analysis of the scale-free fit index for various soft-thresholding powers for **A** Normal and **B** Tumor. **C**, **D** Clustering dendrograms of genes for tumor and normal. Different colors are used to represent gene modules, where gray represents genes that cannot be merged

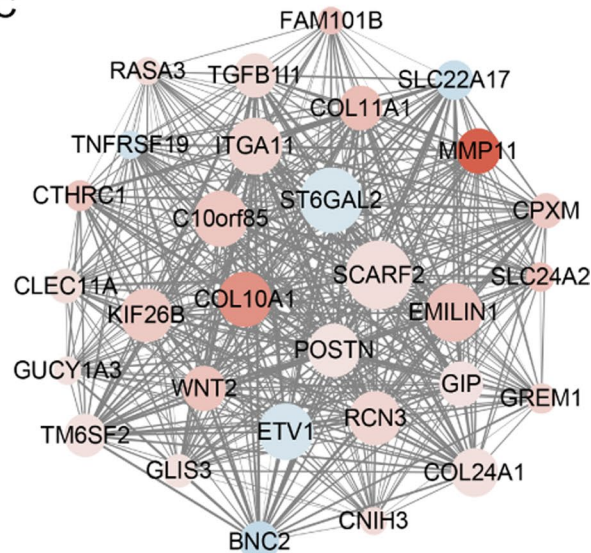He *et al. Egyptian Journal of Medical Human Genetics*     (2024) 25:100

Page 7 of 13



**Fig. 3** Identification of hub gens. Module comparisons: **A** Comparison of tumor and normal modules. The vertical axis represents tumor modules, while the horizontal axis represents normal tissue modules. The depth of red indicates the degree of overlap between modules, and the numbers in each rectangle show the number of overlapping genes. Top 30 hub genes: **B**, **C** The top 30 hub genes in tumor-specific modules (Cyan, Green, Lightcyan, and Salmon) are shown. Nodes represent hub genes, with node size reflecting intramodular connectivity, edge thickness indicating weight, and node color representing log(Fold Change). The top five hub genes are centrally located in the network
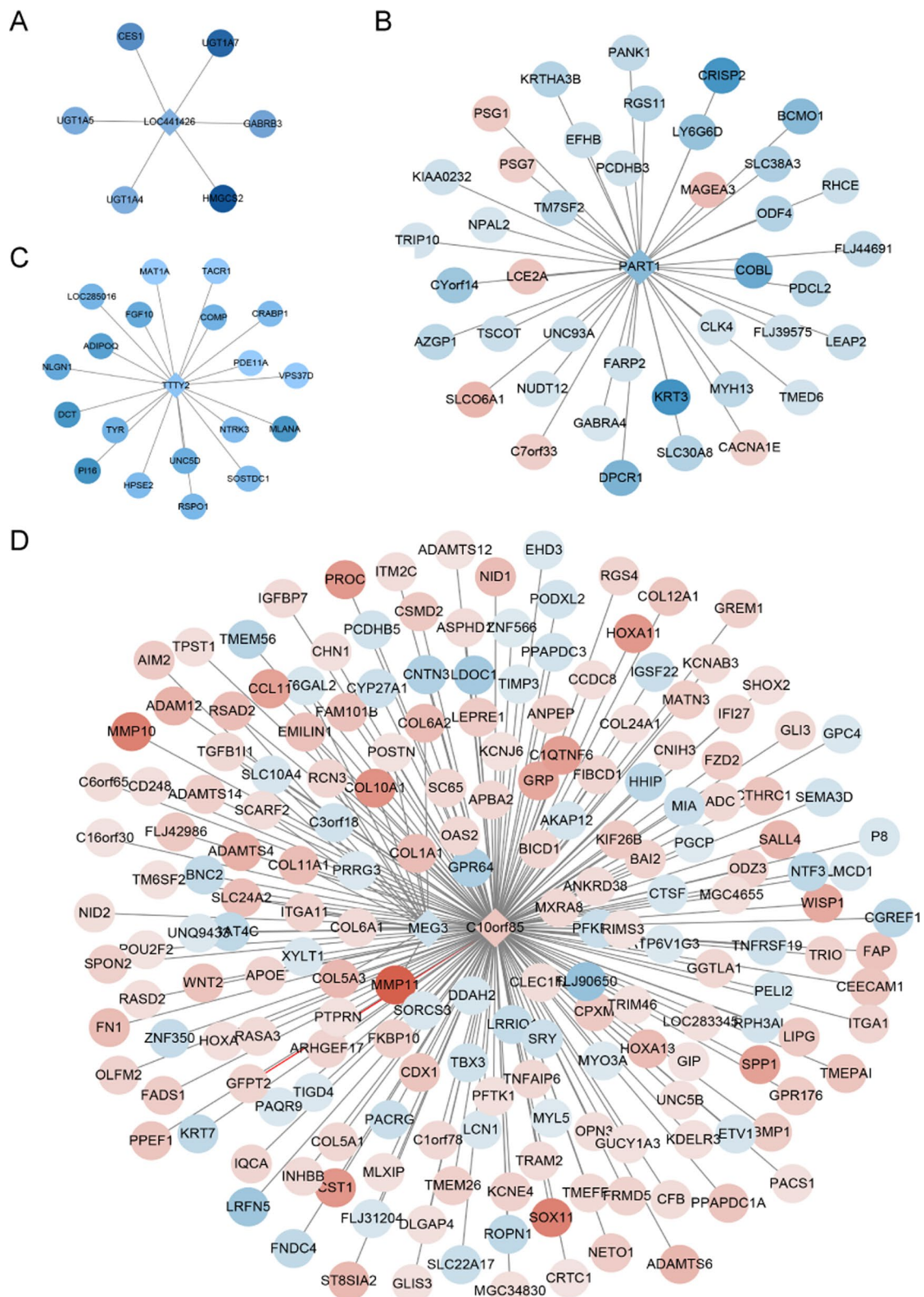
**Fig. 4** Construction of lncRNA–mRNA co-expression network. **A** LncRNA LOC441426; **B** LncRNA PART1; **C** LncRNA TTTY2; **D** LncRNA MEG3 and C10orf85; Round nodes represent mRNA, diamond nodes represent lncRNA, and node colors represent log (Fold Change)
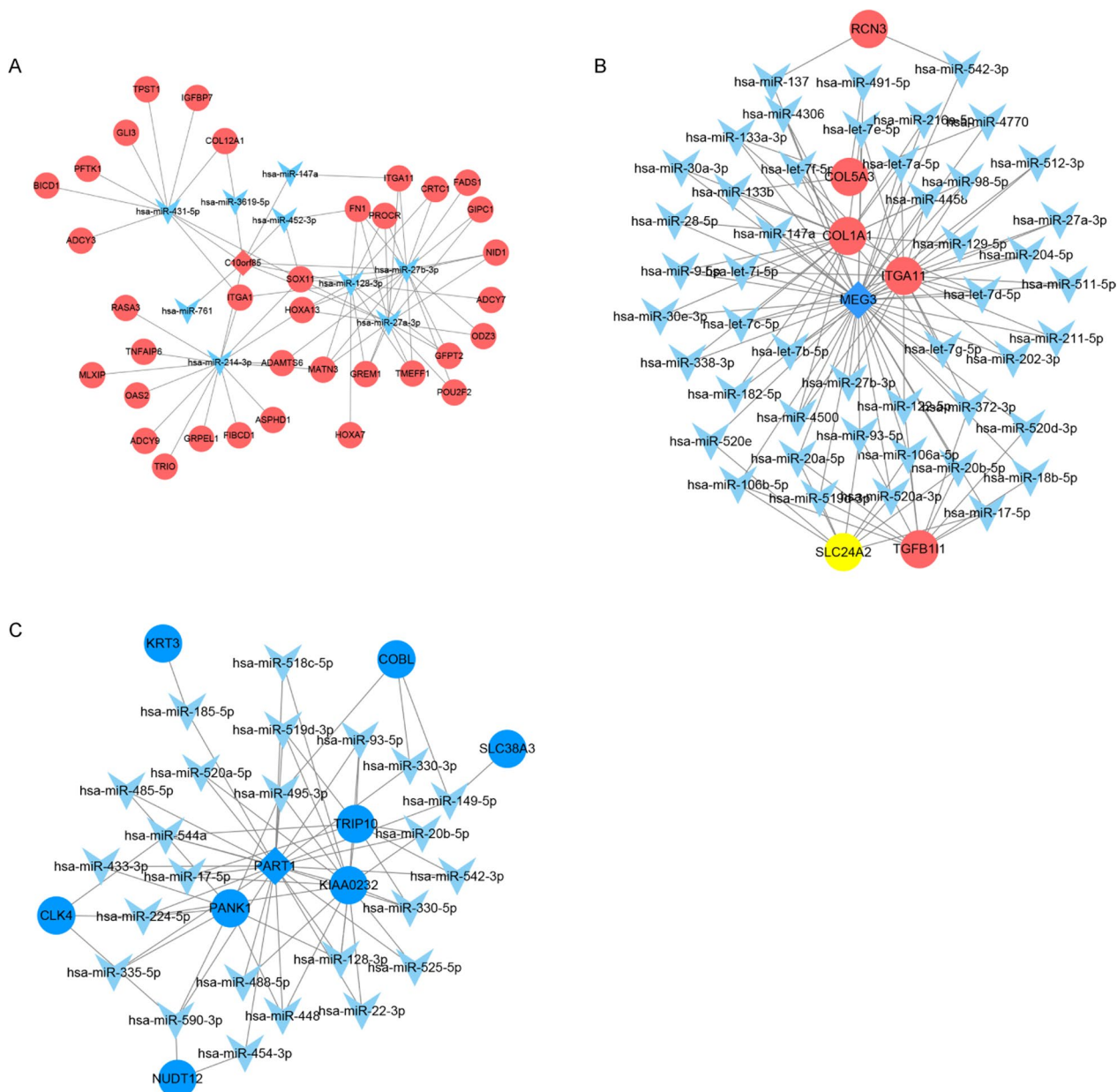
**Fig. 5** Construction of ceRNA co-expression network. **A** LncRNA C10orf85; **B** LncRNA MEG3; **C** LncRNA PART1; The rhombus represents lncRNA, the round represents RNA, and the V-shaped represents microRNAs. Dark blue indicates down-regulation and red indicates up-regulation. The color of the microRNA node is meaningless

INHBB, and GUCY1A3 (from the lncRNA–mRNA network) were significantly associated with patient prognosis, suggesting their potential as prognostic biomarkers for OSCC. This study enhances our understanding of OSCC's molecular mechanisms and could lead to more precise diagnostic methods and targeted therapies.

AQP5, a member of the aquaporin (AQP) family of membrane proteins, facilitates water and glycerol transport and is primarily expressed in secretory glands and certain epithelial tissues [35–37]. Its overexpression is associated with increased cancer cell proliferation, metastasis, and migration in various cancers [38–41]. In OSCC, AQP5 is down-regulated compared to normal tissues and exhibits high intramodular connectivity in the OSCC regulatory network. Cox proportional hazard model analysis identifies AQP5 as an independent prognostic factor for OSCC, suggesting it plays a significant role in OSCC progression. Furthermore, AQP5 may
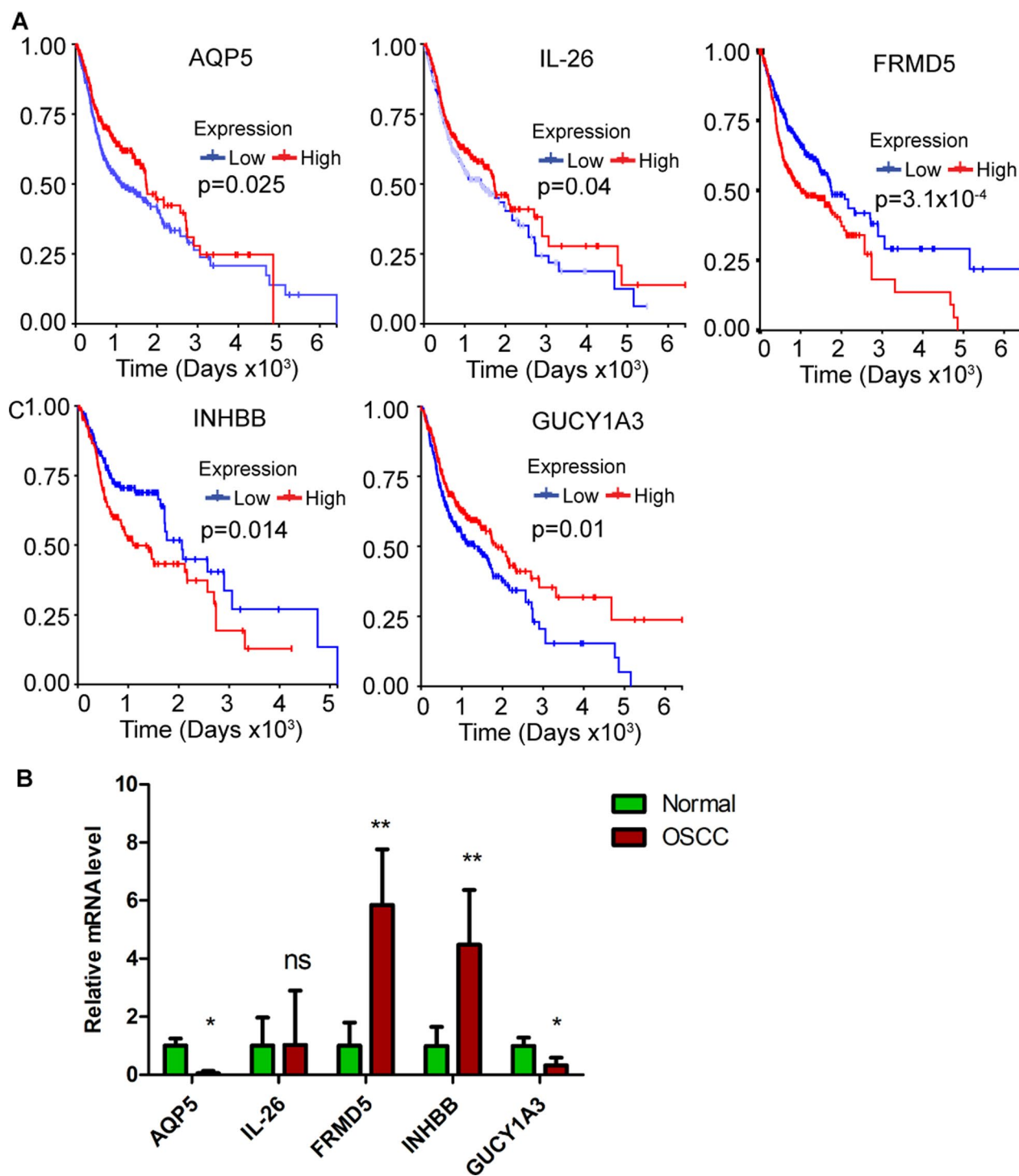
**Fig. 6** Expression of five hub genes in clinical OSCC versus normal tissues. **A** Overall survival analyses of five hub genes in the OSCC TCGA dataset. **B** Expression levels of AQP5, IL-26, FRMD5, INHBB and GUCYLA3 in OSCC tissues were detected by RT-qPCR. Statistically significant genes are highlighted in the figure, with red lines indicating high expression and blue lines indicating low expression

influence gemcitabine sensitivity, with low AQP5 potentially reducing drug efficacy and worsening prognosis [42, 43].

IL-26, a cytokine from the IL-10 family produced by Th17 cells, is involved in promoting inflammatory responses and microbial killing by inducing plasmocytoid

dendritic cells (pDCs) to secrete type I interferon [44–46]. Its role in cancer is debated: it can inhibit colorectal cancer cell proliferation [44] or enhance gastric cancer cell survival [47]. In this study, IL-26 was down-regulated in OSCC and associated with poor prognosis. While, IL-26 can stimulate type I interferon secretion and potentially activate dormant cancer stem cells, its diminished expression might reduce drug efficacy [48] and worsen survival outcomes. However, the lack of a significant expression difference between OSCC and normal tissues may be due to the small sample size, and further research with additional samples is ongoing.

Cox proportional hazards model multivariate analysis revealed that FRMD5, INHBB, and GUCY1A3, identified from the lncRNA–mRNA network, are significant prognostic factors for OSCC. FRMD5, a member of the FERM protein family, was first identified as a downstream target of the mutant p53R273H protein [49]. It is highly expressed in colorectal cancer and predicts poor prognosis [50], a finding consistent with our OSCC results. INHBB is a cytokine in the TGF-β family [51]. Overexpression of INHBB is associated with poor survival in OSCC. INHBB forms activin B, which promotes OSCC metastasis by regulating EMT-related genes [51]. GUCY1A3, which encodes the α1 subunit of soluble guanylate cyclase [52], has limited research linking it to cancers. One study [53] showed that GUCY1A3 overexpression increases cyclic GMP production, promoting angiogenesis in gliomas. Our study found GUCY1A3 down-regulated in OSCC, with low expression predicting poor prognosis. Additionally, the lncRNA C10orf85 regulates FRMD5, INHBB, and GUCY1A3. Although the precise role of C10orf85 and its regulatory mechanisms with these genes in OSCC are not fully understood, these findings suggest it is a key player.

We also constructed a ceRNA network with five hub lncRNAs (LOC441426, TTTY2, PART1, MEG3, and C10orf85). PART1, MEG3, and C10orf85 regulate top hub mRNAs through various miRNAs. For instance, PART1 positively regulates SLC38A3 and PANK1 in the salmon module, while ITGA11 in the green module is co-expressed with lncRNA C10orf85 and MEG3. These insights provide new strategies for exploring OSCC pathogenesis.

We acknowledge several limitations in our study that should be considered when interpreting the results. Although we performed extensive data collection and computational analysis, the findings are primarily based on data mining techniques and may require further validation. Specifically, while we validated the expression of five hub genes in clinical samples compared to normal tissues, these results have not been confirmed through experimental studies such as in vivo animal models or functional assays. Additionally, the number of patient samples tested is limited. To robustly establish the significance of our findings, it is essential to conduct comprehensive experimental research, including cellular and animal model studies. Such validation is crucial for corroborating our results and understanding their biological relevance in the context of OSCC.

## Conclusions

In summary, we used WGCNA to identify tumor-specific modules and hub genes in OSCC. We constructed a functional competing endogenous RNA (ceRNA) network based on these modules and identified key genes—AQP5, IL-26, FRMD5, INHBB, and GUCY1A3—that are associated with OSCC patient survival and validated the differential expression levels of these genes in patient and normal tissues. The findings may suggest that these genes could serve as novel biomarkers for prognostic assessment in OSCC and may offer new targets for therapeutic intervention.

**Abbreviations**

| | |
|---|---|
| OSCC | Oral squamous cell carcinoma |
| GEO | Gene Expression Omnibus |
| WGCNA | Weighted correlation network analysis |
| TCGA | Cancer Genome Atlas |
| FPKM | Fragments per kilobase of transcript per million mapped reads |
| PART1 | Prostate androgen-regulated transcript 1 |
| TYR | Tyrosinase |
| HPV | Human papillomavirus |
| DEGs | Differentially expressed genes |
| ceRNAs | Competing endogenous RNAs |
| lncRNAs | Long noncoding RNAs |
| circRNAs | Circular RNAs |
| MREs | MicroRNA response elements |
| GO | Gene ontology |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| DEmRNAs | Differentially expressed mRNAs |
| BP | Biological process |
| CC | Cellular component |
| MF | Molecular function |
| ChIP | Chromatin immunoprecipitation |
| HGNC | HUGO Gene Nomenclature Committee |
| OS | Overall survival |
| HNSCC | Head and neck squamous cell carcinoma |
| NSCLC | Non-small-cell lung cancer |
| EMT | Epithelial-mesenchymal transition |
| CAFs | Cancer-associated fibroblasts |
| GC | Gastric cancer |
| CYP | Cytochrome P450 oxidase |

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43042-024-00574-7.

Additional file 1

## Declarations

### Ethics approval and consent to participate
Not applicable.

### Consent for publication
Not applicable.

### Competing interests
The authors state that there are no potential competing interest regarding the authorship and/or publication of this article.

### Author details
[1]Department of Prosthodontics, Changsha Stomatological Hospital, 844 Wuyi Avenue, Changsha 410005, People's Republic of China. [2]Department of Oral and Maxillofacial Surgery, Changsha Stomatological Hospital, 389 Youyi Road, Changsha 410118, People's Republic of China. [3]Department of Stomatology, The Second Xiangya Hospital, Central South University, 139 Renmin Middle Road, Changsha 410011, People's Republic of China.

## References
1. Wang K, Ling T, Wu H, Zhang J (2013) Screening of candidate tumor-suppressor genes in 3p21. 3 and investigation of the methylation of gene promoters in oral squamous cell carcinoma. Oncol Rep 29:1175–1182
2. Cancer Genome Atlas Network (2015) Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature 517(7536):576
3. Jiang Y, Liao L, Shrestha C, Ji S, Chen Y, Peng J, Wang L, Liao E, Xie Z (2015) Reduced expression of E-cadherin and p120-catenin and elevated expression of PLC-γ1 and PIKE are associated with aggressiveness of oral squamous cell carcinoma. Int J Clin Exp Pathol 8:9042
4. Secretan B, Straif K, Baan R, Grosse Y, El Ghissassi F, Bouvard V, Benbrahim-Tallaa L, Guha N, Freeman C, Galichet L, Cogliano V (2009) A review of human carcinogens—part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. Lancet Oncol 10:1033–1034
5. Xia L, Tian-You L, Yi-Jun G, Dong-sheng T, Wen-Hui L (2009) Arecoline and oral keratinocytes may affect the collagen metabolism of fibroblasts. J Oral Pathol Med 38:422–426
6. Lee CH, Ko AMS, Warnakulasuriya S, Yin BL, Zain RB, Ibrahim SO, Liu ZW, Li WH, Zhang SS, Utomo B (2011) Intercountry prevalences and practices of betel-quid use in south, southeast and eastern Asia regions and associated oral preneoplastic disorders: an international collaborative study by Asian betel-quid consortium of south and east Asia. Int J Cancer 129:1741–1751
7. Dayyani F, Etzel CJ, Liu M, Ho CH, Lippman SM, Tsao AS (2010) Meta-analysis of the impact of human papillomavirus (HPV) on cancer risk and overall survival in head and neck squamous cell carcinomas (HNSCC). Head Neck Oncol 2:15
8. Jiang S, Dong Y (2017) Human papillomavirus and oral squamous cell carcinoma: a review of HPV-positive oral squamous cell carcinoma and possible strategies for future. Curr Probl Cancer 41:323–327
9. Bozinovic K, Sabol I, Dediol E, Milutin Gasperov N, Manojlovic S, Vojtechova Z, Tachezy R, Grce M (2019) Genome-wide miRNA profiling reinforces the importance of miR-9 in human papillomavirus associated oral and oropharyngeal head and neck cancer. Sci Rep 9:2306
10. Feng L, Houck JR, Lohavanichbutr P, Chen C (2017) Transcriptome analysis reveals differentially expressed lncRNAs between oral squamous cell carcinoma and healthy oral mucosa. Oncotarget 8:31521–31531
11. Eljabo N, Nikolic N, Carkic J, Jelovac D, Lazarevic M, Tanic N, Milasin J (2018) Genetic and epigenetic alterations in the tumour, tumour margins, and normal buccal mucosa of patients with oral cancer. Int J Oral Maxillofac Surg 47:976–982
12. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinform 9:559
13. Li B, Pu K, Wu X (2019) Identifying novel biomarkers in hepatocellular carcinoma by weighted gene co-expression network analysis. J Cell Biochem 120(7):11418–11431
14. Xiang Y, Zhang CQ, Huang K (2012) Predicting glioblastoma prognosis networks using weighted gene co-expression network analysis on TCGA data. BMC Bioinform 13(Suppl 2):S12
15. Qiu J, Du Z, Wang Y, Zhou Y, Zhang Y, Xie Y, Lv Q (2019) Weighted gene co-expression network analysis reveals modules and hub genes associated with the development of breast cancer. Medicine (Baltimore) 98:e14345
16. Yin X, Wang J, Zhang J (2018) Identification of biomarkers of chromophobe renal cell carcinoma by weighted gene co-expression network analysis. Cancer Cell Int 18:206
17. Zhang X, Feng H, Li Z, Li D, Liu S, Huang H, Li M (2018) Application of weighted gene co-expression network analysis to identify key modules and hub genes in oral squamous cell carcinoma tumorigenesis. Onco Targets Ther 11:6001–6021
18. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP (2011) A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell 146:353–358
19. Peng H, Lu M, Selaru FM (2015) The genome-wide gene expression profiling to predict competitive endogenous RNA network in hepatocellular cancer. Genom Data 4:93–95
20. Yang S, Ning Q, Zhang G, Sun H, Wang Z, Li Y (2016) Construction of differential mRNA–lncRNA crosstalk networks based on ceRNA hypothesis uncover key roles of lncRNAs implicated in esophageal squamous cell carcinoma. Oncotarget 7:85728–85740
21. Huang M, Zhong Z, Lv M, Shu J, Tian Q, Chen J (2016) Comprehensive analysis of differentially expressed profiles of lncRNAs and circRNAs with associated co-expression and ceRNA networks in bladder carcinoma. Oncotarget 7:47186–47200
22. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249–264
23. Diboun I, Wernisch L, Orengo CA, Koltzenburg M (2006) Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. BMC Genom 7:252
24. da Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57
25. Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, Brozell A, Schadt EE, Drake TA, Lusis AJ, Horvath S (2006) Integrating genetic and network analysis to characterize genes related to mouse weight. PLoS Genet 2:e130
26. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504
27. Paraskevopoulou MD, Vlachos IS, Karagkouni D, Georgakilas G, Kanellos I, Vergoulis T, Zagganas K, Tsanakas P, Floros E, Dalamagas T (2015) DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. Nucleic Acids Res 44:D231–D238

28. Li J-H, Liu S, Zhou H, Qu L-H, Yang J-H (2013) starBase v2. 0: decoding miRNA–ceRNA, miRNA–ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res 42:D92–D97
29. Therneau TM, Grambsch PM (2013) Modeling survival data: extending the cox model. Springer, New York
30. Kassambara A, Kosinski A, Biecek P, Fabian S (2019) Survminer: drawing survival curves using'ggplot2'. R package version 04 2019, 4
31. Rivera C, Oliveira AK, Costa RAP, De Rossi T, Paes Leme AF (2017) Prognostic biomarkers in oral squamous cell carcinoma: a systematic review. Oral Oncol 72:38–47
32. Scott SE, Grunfeld EA, McGurk M (2005) The idiosyncratic relationship between diagnostic delay and stage of oral squamous cell carcinoma. Oral Oncol 41:396–403
33. Almangush A, Pirinen M, Heikkinen I, Makitie AA, Salo T, Leivo I (2018) Tumour budding in oral squamous cell carcinoma: a meta-analysis. Br J Cancer 118:577–586
34. Wen Q, Alnemah MM, Luo J, Wang W, Chu S, Chen L, Li J, Xu L, Li M, Zhou J (2015) FLOT-2 is an independent prognostic marker in oral squamous cell carcinoma. Int J Clin Exp Pathol 8:8236
35. Huber VJ, Tsujita M, Nakada T (2012) Aquaporins in drug discovery and pharmacotherapy. Mol Aspects Med 33:691–703
36. Moosavi MS, Elham Y (2020) Aquaporins 1, 3 and 5 in different tumors, their expression, prognosis value and role as new therapeutic targets. Pathol Oncol Res 26(2):615–625
37. Woo J, Lee J, Chae YK, Kim MS, Baek JH, Park JC, Park MJ, Smith IM, Trink B, Ratovitski E et al (2008) Overexpression of AQP5, a putative oncogene, promotes cell growth and transformation. Cancer Lett 264:54–62
38. Jensen HH, Login FH, Koffman JS, Kwon TH, Nejsum LN (2016) The role of aquaporin-5 in cancer cell migration: a potential active participant. Int J Biochem Cell Biol 79:271–276
39. Liu S, Zhang S, Jiang H, Yang Y, Jiang Y (2013) Co-expression of AQP3 and AQP5 in esophageal squamous cell carcinoma correlates with aggressive tumor progression and poor prognosis. Med Oncol 30:636
40. Shimizu H, Shiozaki A, Ichikawa D, Fujiwara H, Konishi H, Ishii H, Komatsu S, Kubota T, Okamoto K, Kishimoto M, Otsuji E (2014) The expression and role of Aquaporin 5 in esophageal squamous cell carcinoma. J Gastroenterol 49:655–666
41. Lehnerdt GF, Bachmann HS, Adamzik M, Panic A, Koksal E, Weller P, Lang S, Schmid KW, Siffert W, Bankfalvi A (2015) AQP1, AQP5, Bcl-2 and p16 in pharyngeal squamous cell carcinoma. J Laryngol Otol 129:580–586
42. Sekine S, Shimada Y, Nagata T, Moriyama M, Omura T, Watanabe T, Hori R, Yoshioka I, Okumura T, Sawada S et al (2012) Prognostic significance of aquaporins in human biliary tract carcinoma. Oncol Rep 27:1741–1747
43. Mynhardt C, Damelin LH, Jivan R, Peres J, Prince S, Veale RB, Mavri-Damelin D (2018) Metformin-induced alterations in nucleotide metabolism cause 5-fluorouracil resistance but gemcitabine susceptibility in oesophageal squamous cell carcinoma. J Cell Biochem 119:1193–1203
44. Dambacher J, Beigel F, Zitzmann K, De Toni EN, Goke B, Diepolder HM, Auernhammer CJ, Brand S (2009) The role of the novel Th17 cytokine IL-26 in intestinal inflammation. Gut 58:1207–1217
45. Meller S, Di Domizio J, Voo KS, Friedrich HC, Chamilos G, Ganguly D, Conrad C, Gregorio J, Le Roy D, Roger T et al (2015) T(H)17 cells promote microbial killing and innate immune sensing of DNA via interleukin 26. Nat Immunol 16:970–979
46. Tengvall S, Che KF, Linden A (2016) Interleukin-26: an Emerging Player in Host Defense and Inflammation. J Innate Immun 8:15–22
47. You W, Tang Q, Zhang C, Wu J, Gu C, Wu Z, Li X (2013) IL-26 promotes the proliferation and survival of human gastric cancer cells by regulating the balance of STAT1 and STAT3 activation. PLoS ONE 8:e63588
48. Ma H, Jin S, Yang W, Tian Z, Liu S, Wang Y, Zhou G, Zhao M, Gvetadze S, Zhang Z, Hu J (2017) Interferon-alpha promotes the expression of cancer stem cell markers in oral squamous cell carcinoma. J Cancer 8:2384–2393
49. Brazdova M, Quante T, Togel L, Walter K, Loscher C, Tichy V, Cincarova L, Deppert W, Tolstonog GV (2009) Modulation of gene expression in U251 glioblastoma cells by binding of mutant p53 R273H to intronic and intergenic sequences. Nucleic Acids Res 37:1486–1500
50. Zhu C, Yamaguchi K, Ohsugi T, Terakado Y, Noguchi R, Ikenoue T, Furukawa Y (2017) Identification of FERM domain-containing protein 5 as a novel target of beta-catenin/TCF7L2 complex. Cancer Sci 108:612–619
51. Kita A, Kasamatsu A, Nakashima D, Endo-Sakamoto Y, Ishida S, Shimizu T, Kimura Y, Miyamoto I, Yoshimura S, Shiiba M et al (2017) Activin B

52. regulates adhesion, invasiveness, and migratory activities in oral cancer: a potential biomarker for metastasis. J Cancer 8:2033–2041
52. Mohammadoo Khorasani M, Karami Tehrani F, Parizadeh SMR, Atri M (2019) Differential expression of alternative transcripts of soluble guanylyl cyclase, GYCY1a3 and GUCY1b3 genes, in the malignant and benign breast tumors. Nitric Oxide 83:65–71
53. Saino M, Maruyama T, Sekiya T, Kayama T, Murakami Y (2004) Inhibition of angiogenesis in human glioma cell lines by antisense RNA from the soluble guanylate cyclase genes, GUCY1A3 and GUCY1B3. Oncol Rep 12:47–52

## Publisher's Note