Identification and validation of key modules and hub genes associated with the pathological stage of oral squamous cell carcinoma by weighted gene co-expression network analysis

Xuegang Hu[1,2], Guanwen Sun[2,3], Zhiqiang Shi[1], Hui Ni[1] and Shan Jiang[4]

[1] Department of Stomatology, Shenzhen Hospital, University of Chinese Academy of Sciences, Shenzhen, Guangdong, China
[2] Department of Endodontics and Operative Dentistry, School and Hospital of Stomatology, Fujian Medical University, Fuhou, Fujian, China
[3] Department of Stomatology, Ningbo Medical Center Lihuili Eastern Hospital, Ningbo, Zhejiang, China
[4] Restorative Dental Sciences, Faculty of Dentistry, The University of Hong Kong, Hong Kong SAR, China

ABSTRACT

Background. Oral squamous cell carcinoma (OSCC) is a major lethal malignant cancer of the head and neck region, yet its molecular mechanisms of tumourigenesis are still unclear.

Patients and methods. We performed weighted gene co-expression network analysis (WGCNA) on RNA-sequencing data with clinical information obtained from The Cancer Genome Atlas (TCGA) database. The relationship between co-expression modules and clinical traits was investigated by Pearson correlation analysis. Furthermore, the prognostic value and expression level of the hub genes of these modules were validated based on data from the TCGA database and other independent datasets from the Gene Expression Omnibus (GEO) database and the Human Protein Atlas database. The significant modules and hub genes were also assessed by functional analysis and gene set enrichment analysis (GSEA).

Results. We found that the turquoise module was strongly correlated with pathologic T stage and significantly enriched in critical functions and pathways related to tumourigenesis. PPP1R12B, CFD, CRYAB, FAM189A2 and ANGPTL1 were identified and statistically validated as hub genes in the turquoise module and were closely implicated in the prognosis of OSCC. GSEA indicated that five hub genes were significantly involved in many well-known cancer-related biological functions and signaling pathways.

Conclusion. In brief, we systematically discovered a co-expressed turquoise module and five hub genes associated with the pathologic T stage for the first time, which provided further insight that WGCNA may reveal the molecular regulatory mechanism involved in the carcinogenesis and progression of OSCC. In addition, the five hub genes may be considered candidate prognostic biomarkers and potential therapeutic targets for the precise early diagnosis, clinical treatment and prognosis of OSCC in the future.

Subjects Bioinformatics, Genomics, Dentistry, Oncology

Keywords Oral squamous cell carcinoma (OSCC), Weighted gene co-expression network analysis (WGCNA), Hub gene, Pathologic stage, Overall survival

Submitted 10 October 2019
Accepted 3 January 2020
Published 4 February 2020

Corresponding author
Xuegang Hu, hxueg2016@163.com

Academic editor
Vladimir Uversky

Additional Information and Declarations can be found on page 19

DOI **10.7717/peerj.8505**

How to cite this article Hu X, Sun G, Shi Z, Ni H, Jiang S. 2020. Identification and validation of key modules and hub genes associated with the pathological stage of oral squamous cell carcinoma by weighted gene co-expression network analysis. *PeerJ* **8**:e8505 http://doi.org/10.7717/peerj.8505

INTRODUCTION

Oral squamous cell carcinoma (OSCC) is the most common malignancy of head and neck squamous cell carcinoma (HNSCC) and has poor prognosis and survival (*Bozec et al., 2009*; *Ferlay et al., 2015*). The main treatment for OSCC is comprehensive treatment based on surgical operation, which prolongs the survival time of patients and improves the quality of life (*Kim et al., 2017*; *Verusingam et al., 2017*). However, due to the lack of early diagnostic markers, patients are often in an advanced clinical stage at the time of diagnosis, and the 5-year overall survival of patients with OSCC remains low (*Bland, Clarke & Harden, 1976*; *Omar, 2013*). Therefore, effective biomarkers are needed to explore diagnostic and therapeutic targets for OSCC (*Mehrotra & Gupta, 2011*).

The occurrence and development of OSCC is an extremely complex progressive process involving multiple molecular mechanisms (*Kamangar, Dores & Anderson, 2006*). Due to the limitations of traditional studies, most previous studies focused on individual genes or pathways, and the relationship between genes was ignored (*Sun et al., 2017*). With the advent and rapid development of RNA sequencing technologies in various tumours, bioinformatics analysis has been widely and rapidly used to identify novel and more effective potential biomarkers for the diagnosis, therapy and prognosis of many diseases (*Hinchcliff et al., 2019*). For example, one study by *Ahluwalia et al. (2019)* identified a novel 4-gene prognostic signature that has clinical utility in colorectal cancer using The Cancer Genome Atlas (TCGA) database and Gene Expression Omnibus (GEO) database.

Weighted gene co-expression network analysis (WGCNA) is an efficient systematic biological approach that can highlight co-expressed gene modules and investigate the relationships between gene modules and phenotypes more effectively (*Langfelder & Horvath, 2008*). WGCNA has been successfully and comprehensively used to explore targeted modules and hub genes in cancer-related research, such as clear cell renal cell carcinoma (*Wang et al., 2019a*; *Wang et al., 2019b*) and pancreatic carcinoma (*Zhou et al., 2018a*; *Zhou et al., 2018b*).

In the current study, we used WGCNA and other bioinformatics analysis methods to explore RNA-Seq data and clinical phenotypes of OSCC patients. Ultimately, we identified that the turquoise module was significantly associated with pathologic T stage for the first time. Five hub genes (PPP1R12B, CFD, CRYAB, FAM189A2 and ANGPTL1) related to prognosis at the transcriptional level were identified and validated in other independent datasets. Further functional analysis indicated that these genes were significantly enriched in critical biological functions and pathways related to the tumourigenesis and development of OSCC.

MATERIALS & METHODS

Study design

To clarify the research process, the workflow of our study is presented in Fig. 1.

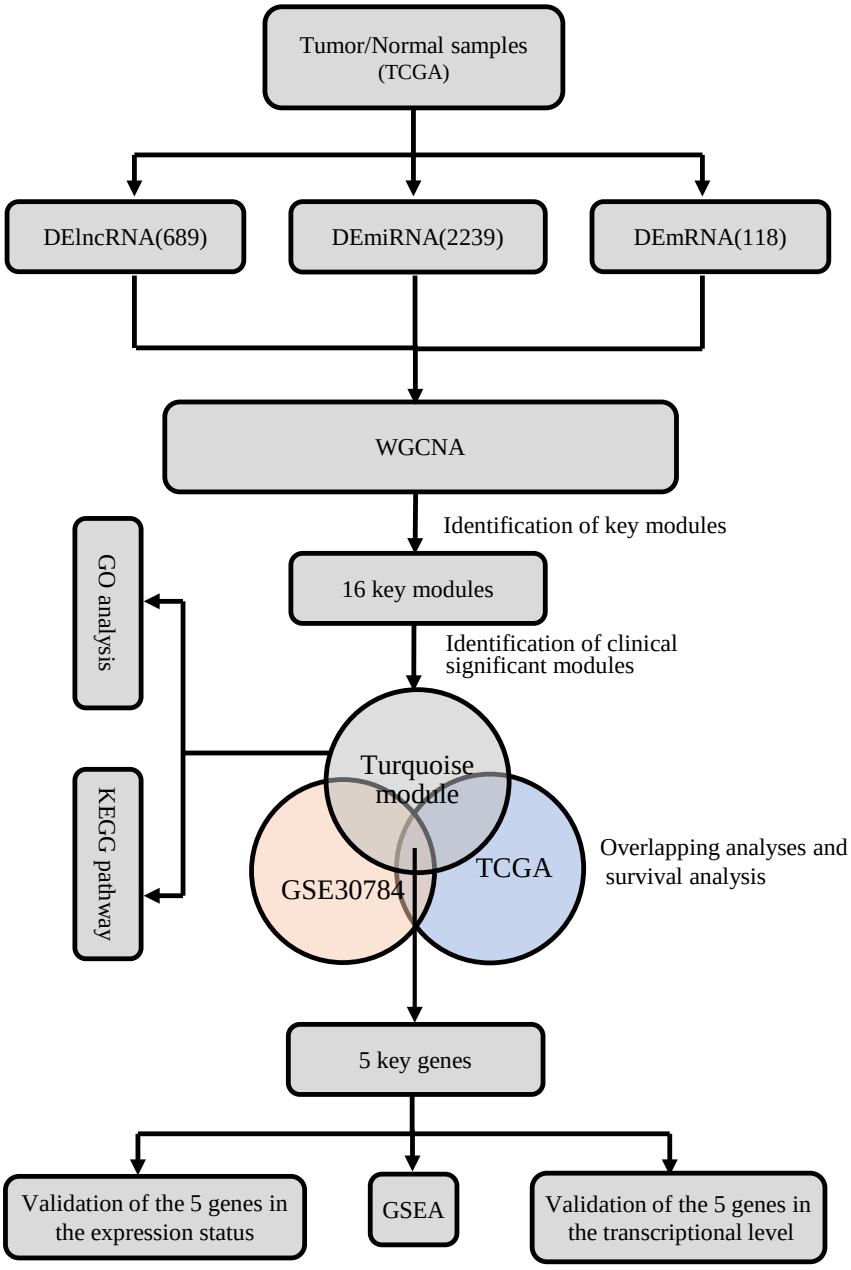

Figure 1 **The flow chart of data preparation, processing, analysis, and validation.**
Full-size DOI: 10.7717/peerj.8505/fig-1

Data acquisition

The RNA-seq expression data and relative clinical information of OSCC patients were retrieved from the TCGA database (https://portal.gdc.cancer.gov/) and GEO database (http://www.ncbi.nlm.nih.gov/geo/). A total of 373 patients were obtained from TCGA and used as a discovery group to construct a co-expression model of which 44 normal

ones and 329 OSCC. Also, 229 patients (167 OSCC samples, 17 dysplasia samples and 45 normal samples) from GSE30784 had available data on clinical characteristics.

Data preprocessing

Data preprocessing and analysis procedures were used to process the raw data, including robust multi-array average (RMA) background correction and the "affy" R package. The Affymetrix annotation files were used to annotate probes, and probes with no annotation were removed. False discovery rate (FDR) <0.05 and $|\log2FC| \geq 2$ were set as the cut-off values for screening differentially expressed mRNAs (DEmiRNA), lncRNAs (DElncRNA), and miRNA (DEmiRNA).

Differential gene expression analysis

The "edger" R package was used to screen differentially expressed genes (DEGs) in the TCGA dataset between normal and OSCC samples and "limma" R package was used to screen DEGs from GSE30784 (*Robinson, McCarthy & Smyth, 2010*). The threshold was set as log2FC ≥2, and FDR <0.05 was considered significant (*Lai, 2017*; *McCarthy, Chen & Smyth, 2012*; *Robinson, McCarthy & Smyth, 2010*). DEGs that met this criteria were chosen for further analysis. The construction of a volcano plot and hierarchical clustering analysis were also performed by the R packages "ggplot2" and "pheatmap", respectively.

Weighted gene co-expression network construction (WGCNA)

The construction of scale-free gene co-expression modules and identification of highly correlated genes of the DEGs, including lncRNAs, miRNAs and mRNAs, were conducted by the "WGCNA" package in R software (http://www.r-project.org/) (*Chen & Boutros, 2011*; *Zhang & Horvath, 2005*). Similar expression modules can be demonstrated by genes that have the same pathway or function. The cut-off of the co-expression module was set as $P < 0.05$. Then, we further calculated and visualized the dissimilarity of module eigengenes (MEs), chose a cut line for the module dendrogram and merged some modules.

Identification of clinically significant modules

Module-trait associations between MEs and clinical traits, including sex, age, grade, clinical stage (TNM) and pathological stage, were assessed by the Pearson test. In principal component analysis, MEs are considered the principal component of each gene module, and the expression patterns of all genes can be summarized as a single characteristic expression profile within a given module. The module with the absolute module significance (MS) ranked first among all the selected modules was considered to be related to a clinical trait (*Shi et al., 2015*). The module significantly correlated ($P < 0.05$) with the phenotype was selected for further investigation.

Identifying hub genes and survival analysis

To identify overlapping genes among the significant modules and the TCGA and GEO (GSE30784) datasets, a Venn diagram (http://jvenn.toulouse.inra.fr/app/example.html) was constructed. The overlapping genes were chosen as the potential genes for overall survival analysis and validation, which were performed using the log-rank test ($p < 0.05$).

Validation of the hub genes

The key genes overlapping among the significant modules and TCGA and GEO datasets that were also significant in survival analysis were chosen as the potential genes for further analysis and validation. *P*-values less than 0.05 were regarded as statistically significant. Furthermore, the Human Protein Atlas database (https://www.proteinatlas.org/) (*Uhlén et al., 2015*) was used to validate the protein expression level of the hub genes.

Functional enrichment analysis of meaningful modules and key genes

To further explore the biological functions of the clinically significant modules and hub genes, we used the "clusterprofiler" package to perform Gene Ontology (GO) term analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis and gene set enrichment analysis (GSEA, https://software.broadinstitute.org/gsea/index.jsp). The enrichment analysis of biological functions and pathways can be described and visualized. The significance level was set as p-value <0.01 and FDR <0.05.

RESULTS

Differentially expressed RNAs between OSCC and control samples

Data from 44 normal and 329 OSCC patients were obtained from the TCGA database. The clinical characteristics of patients with OSCC from TCGA were showed in Table 1. Based on the differential analysis, including 689 lncRNAs, 2239 mRNAs and 118 miRNAs, were left after filtering with thresholds of |log2FC|>2 and adjusted $P < 0.05$ with edger R package (*Robinson, McCarthy & Smyth, 2010*). The volcano map and expression heat map of the differential RNAs were constructed to illustrate the distribution in each category and are presented in Figs. 2 and 3.

Construction of the co-expression modules of OSCC

In order to explore the relationship between dysregulated RNAs and clinical parameters. The "WGCNA" package was used to construct co-expression networks and modules with differentially expressed RNAs based on OSCC from TCGA. Then, these samples were used for cluster analysis by the "flashClust tools" package, and the results are shown in Fig. 4. As shown in Figs. 5A and 5B, when the soft thresholding power value was chosen as 3 ($\beta = 3$), a hierarchical clustering tree (dendrogram) and consensus module eigengenes (Figs. 5C and 5D), including 8 merged co-expression modules, were produced.

Gene Co-expression Modules Correspond to Clinical Traits

Interaction relationships of the 8 co-expression modules were identified and are shown in Fig. 6A, which revealed that each co-expression module independently validated each other in the network. As shown in Fig. 7, the module-feature relationship of the turquoise module revealed a highly negative correlation ($r = -0.2$, $P = 4e{-}04$) with pathologic T stage compared with other modules by Pearson's correlation analysis. In addition, the eigengene dendrogram and heatmap were constructed to explore groups of correlated eigengenes and the dendrogram of all modules (Figs. 6B and 6C). The 7 modules were found to be mainly divided into two clusters. In addition, we constructed a scatterplot of

Table 1 **Clinical characteristics of patients with OSCC from TCGA.**

Variables	Patients, n (%)
Age, years	
≤50	52 (17.5)
>50	245 (82.5)
Gender	
Male	93(31.3)
Female	204(68.7)
Stage	
I	10 (3.4)
II	73 (24.6)
III	64 (21.5)
IV	150 (50.5)
Lymph Node	
N0	158 (53.2)
N1	56 (18.8)
N2	80 (26.9)
N3	3 (3.1)
T stage	
T1	19 (6.4)
T2	97 (32.6)
T3	75 (25.3)
T4	106 (35.7)
Metastasis	
Yes	295 (99.3)
No	2 (0.7)

Notes.
Abbreviations: OSCC, oral squamous cell carcinoma; TCGA, The Cancer Genome Atlas; TNM, tumor nodes metastasis.

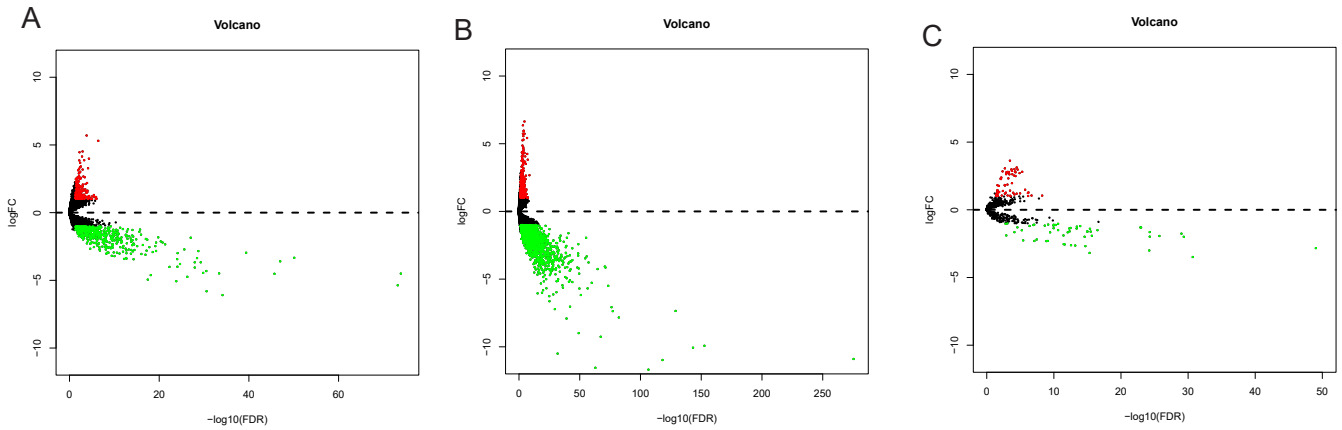

Figure 2 **The volcano plot of differentially expressed RNAs in patients with OSCC.** (A) DElncNAs; (B) DEmiRNAs; (C) DEmRNAs. Up-regulated RNA and down-regulated was represented in red dot and green dot respectively.

Full-size DOI: 10.7717/peerj.8505/fig-2

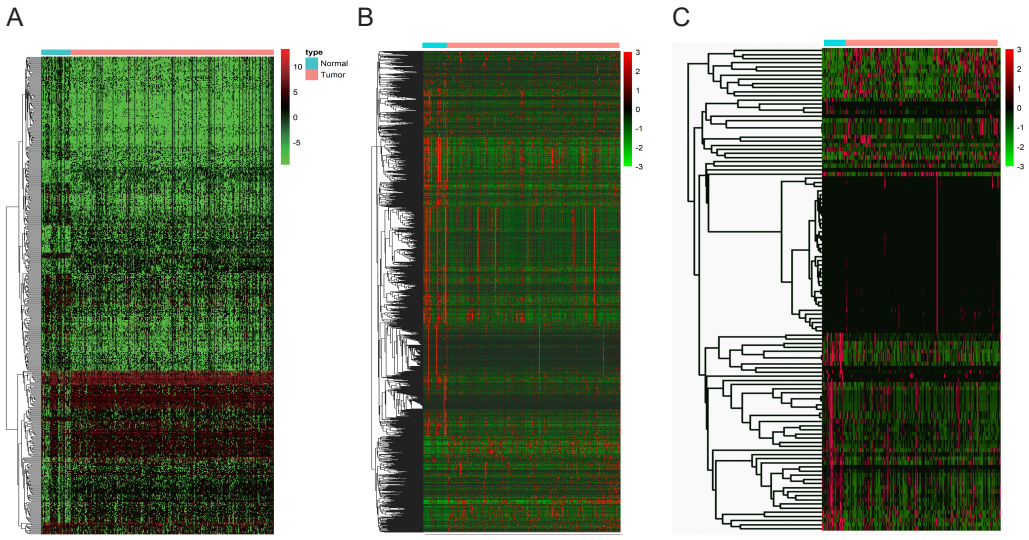

Figure 3 The heatmaps of DEGs between OSCC and normal tissues. (A) DElncNAs; (B) DEmiRNAs; (C) DEmRNAs. Each row represented a kind of RNA, and each column referred to one sample. Expression values are represented by the color scale. Redder grids are higher in expression values, while greener ones are lower.

Full-size DOI: 10.7717/peerj.8505/fig-3

pathologic T stage vs. module membership in the turquoise module, which illustrated that they were highly correlated (cor = 0.22, $p = 3.9e{-}12$) (Fig. 6D).

Functional enrichment analysis of genes in the turquoise module

To gain a primary understanding of the biological functions and pathway relevance of the turquoise module, GO enrichment analysis and KEGG pathway enrichment analysis were conducted. The results of GO enrichment analysis showed that the turquoise module was significantly enriched in critical biological functions, such as cell differentiation, cell–cell adhesion, and cell–cell junctions (Fig. 8). KEGG enrichment analysis revealed that the genes are significantly involved in many pathways that are correlated with tumourigenesis, including the MAPK signalling pathway, intrinsic apoptotic signalling pathway, AMPK signalling pathway, Wnt signalling pathway and Calcium signalling pathway (Fig. 9).

Identification of hub genes by overlapping analyses and survival analysis

To identify the hub genes of the turquoise module, overlapping analyses and survival analysis were performed. The differentially expressed genes in 212 patients (167 OSCC samples and 45 normal samples) were screened and extracted from the GEO database (GSE30784) with limma R package. The heat map of differential RNAs is shown in Fig. 10 which including xxx mRNAs and xxx lncRNAs. Then, a Venn diagram was constructed for overlapping analysis to identify overlapping genes among the turquoise module and the TCGA and GEO (GSE30784) databases. As shown in the Venn diagram, 1 lncRNA (Fig. 11A) and 123 mRNAs (Fig. 11C) were present in the turquoise module, the TCGA and GEO (GSE30784) datasets. 43 miRNAs were present in the turquoise module and

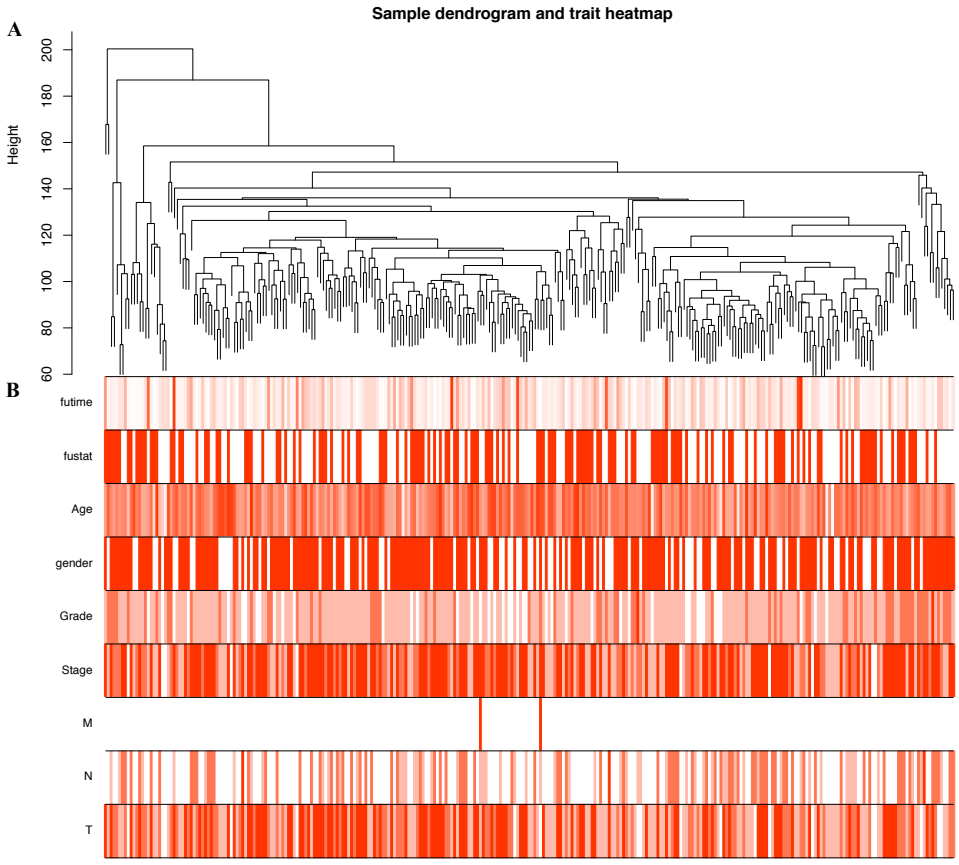

Figure 4 Construction of co-expression modules by the "WGCNA" package. (A) Sample clustering to detect outliers. (B) Sample cluster dendrogram and trait indicators. T, primary tumor; N, lymph node; M, distant metastasis.

Full-size DOI: 10.7717/peerj.8505/fig-4

the TCGA (Fig. 11B). Then, the overlapping genes including 1 lncRNA, 43 miRNAs and 123 mRNAs were selected as the potential genes by the log-rank test ($p < 0.05$) for further overall survival analysis. Eventually, 5 hub mRNAs (ANGPTL1) (Fig. 12A), CFD (Fig. 12B), CRYAB (Fig. 12C), FAM189A2 (Fig. 12D) and PPP1R12B (Fig. 12E) were identified. The Kaplan–Meier survival curve of the overall survival analysis revealed that OSCC patients with low expression levels of the 5 hub genes tended to have a poor outcome.

Validation of the hub genes

The GEO (GSE30784) and TCGA datasets were used to validate the expression status of the 5 hub genes (ANGPTL1, CFD, CRYAB, FAM189A2, and PPP1R12B). Compared with that in adjacent normal tissues, the 5 hub genes were significantly downregulated in tumour tissues from the GEO datasets (GSE30784) (Fig. 13A) and TCGA (Fig. 13B). This result mainly indicates that the expression status was consistent with the pathologic T stage, namely, that the expression of the 5 hub genes of the turquoise module was negatively correlated with the pathologic T stage. In addition, HNSCCs was analysed by the GEPIA online tool (519 OSCC samples and 44 normal samples) (http://gepia.cancer-pku.cn/)

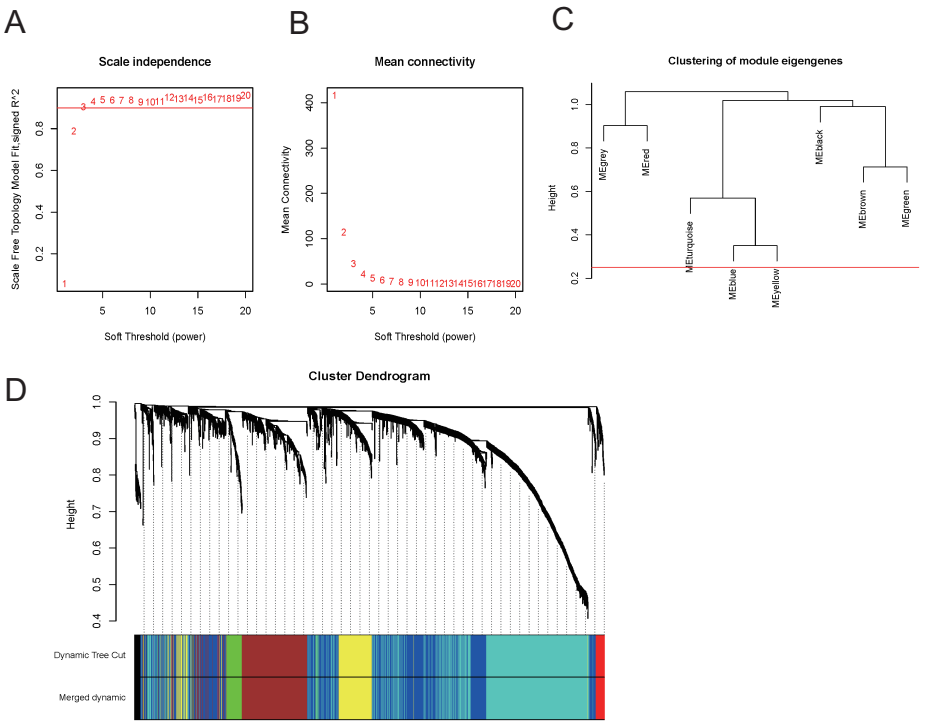

Figure 5 Construction of co-expression modules by the "WGCNA" package. (A) and (B) Analysis of network topology for various soft-thresholding powers. (A) The scale-free fit index as a function of the soft-thresholding power is shown. (B) The mean connectivity as a function of the soft-thresholding power is displayed. (C) The cluster dendrogram of module eigengenes. (D) The cluster dendrogram of genes in TCGA. Each branch in the figure represents one gene, and every colour below represents one co-expression module.

Full-size DOI: 10.7717/peerj.8505/fig-5

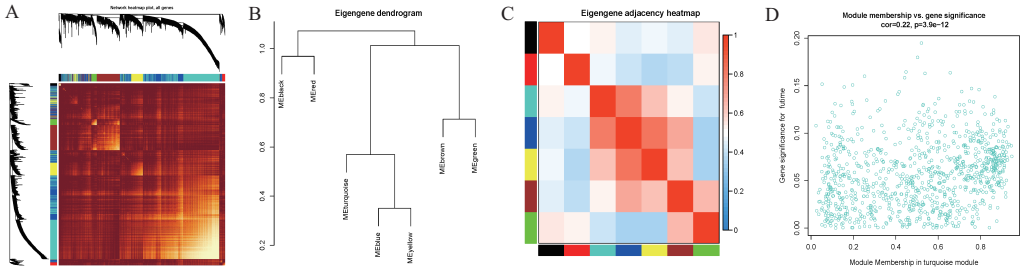

Figure 6 Functional gene modules detected by co-expression network analysis. (A) Topological overlap heatmap of the gene co-expression network. Each row and column represents a gene. Light color indicates low topological overlap, and on the contrary dark color denotes high topological overlap. The different side colors indicate different modules. The dendrogram suggests the clustering of these genes based on the similarity of their gene expression profiles. (B) and (C) The correlation between each module demonstrated based on eigengene. Blue represents a negative correlation, while red represents a positive correlation. (D) Scatter diagram of pathologic T stage vs. module membership for the significant genes in the turquoise module.

Full-size DOI: 10.7717/peerj.8505/fig-6

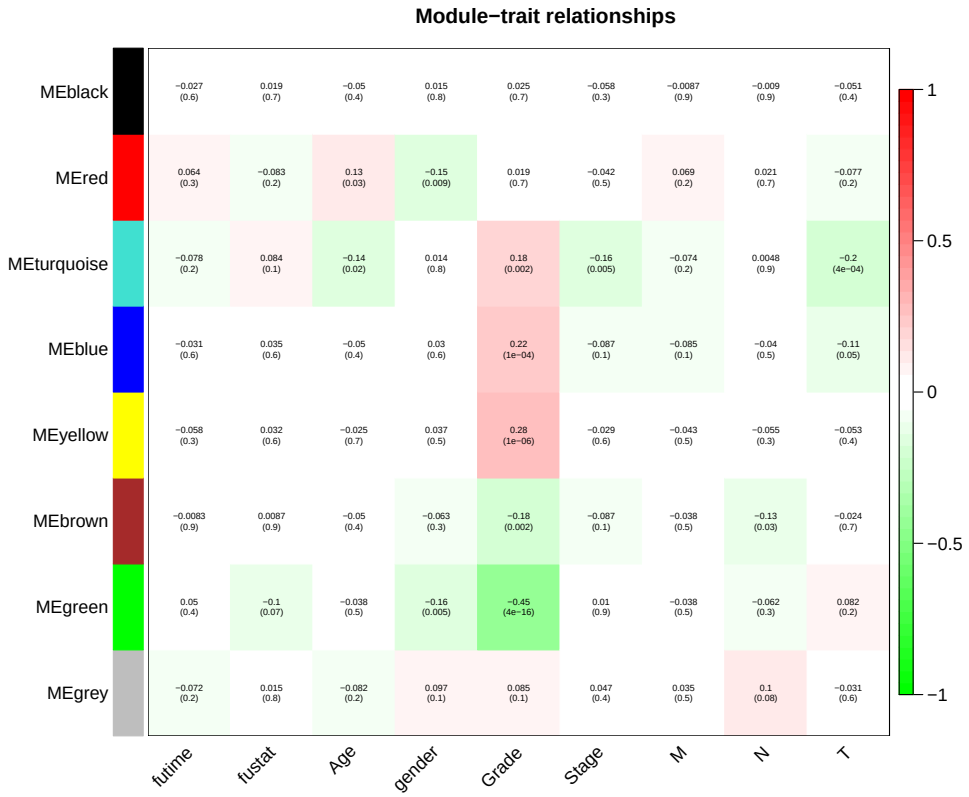

Figure 7 Module-trait associations were evaluated by correlations between MEs and clinical traits. Each row corresponds to a module eigengene, column to a trait. Each cell contains the corresponding correlation (first line) and *p*-value (second line). The cells are color coded by the correlation according to the color legend. Red module positively correlated to age ($p < 0.05$). Turquoise module negatively correlated to age ($p < 0.05$). Red and green modules negatively correlated to gender ($p < 0.05$). Turquoise, blue and yellow modules positively correlated to grade ($p < 0.05$). Green and brown modules negatively correlated to grade ($p < 0.05$). Turquoise module negatively correlated to stage ($p < 0.05$). Brown module negatively correlated to pathologic N stage ($p < 0.05$). Turquoise and blue modules negatively correlated to pathologic T stage ($p < 0.05$).

Full-size DOI: 10.7717/peerj.8505/fig-7

to validate the expression status of the 5 hub genes, which showed that the results were consistent with those described earlier, indicating that the results above are convincing and reliable (Figs. 14A–14E). Moreover, the protein levels of immunohistochemistry (IHC) staining obtained from the Human Protein Atlas (HPA) database showed that the expression of four of the hub genes (CRYAB, FAM189A2, ANGPTL1 and PPP1R12B) were significantly lower in tumour tissues than in normal tissues (Fig. 15), which was consistent with that at the transcriptional level. Among the 5 hub genes, one hub gene (CFD) was not reported in the HPA database.

GSEA of the Hub Genes

Moreover, GSEA was conducted to search the potential biological function and signaling pathway of the above five hub genes (PPP1R12B, CFD, CRYAB, FAM189A2 and ANGPTL1). The result of GESA indicated that the 5 hub genes were significantly involved

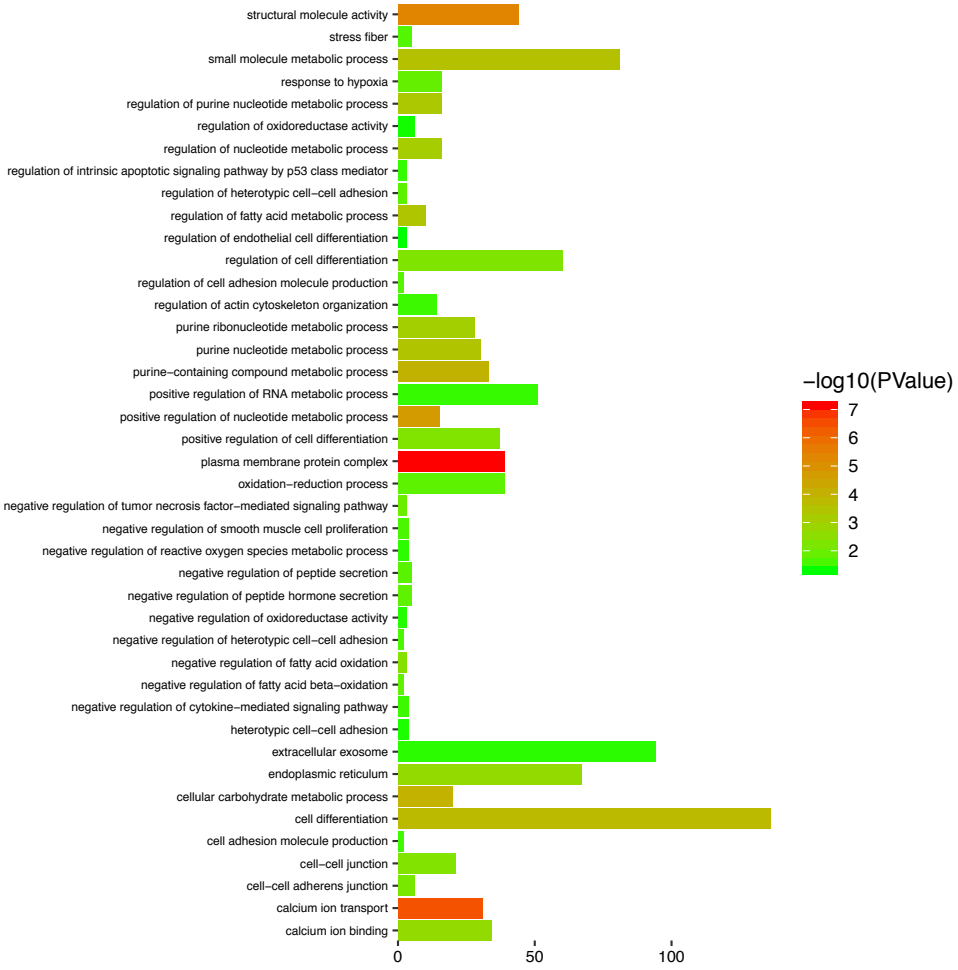

Figure 8 **Top enriched GO terms for the turquoise module.**
Full-size DOI: 10.7717/peerj.8505/fig-8

in critical biological functions and signal pathways that were correlated with carcinogenesis and progression of tumor, such as pathway in cancer, P53 signal pathway, MTOR signal pathway, Notch signal pathway, cell cycle, RRNA metabolic process, ribosome biogenesis and calcium ion transport (Figs. 16 and 17).

DISCUSSION

OSCC is one of the most complex and common malignant cancers. Despite significant improvements in the diagnosis, prognosis, and treatment of OSCC during the last decades, the 5-year overall survival rate is still very poor at approximately 50% due to local recurrence and metastasis (*Scott, Grunfeld & McGurk, 2005*; *Omar, 2013*). Therefore, to better explore novel and precise molecular biomarkers that could accurately and specifically predict the progression, recurrence and prognosis of OSCC patients (*Mehrotra & Gupta, 2011*), we used RNA sequencing data with clinical information from the TCGA and GEO databases

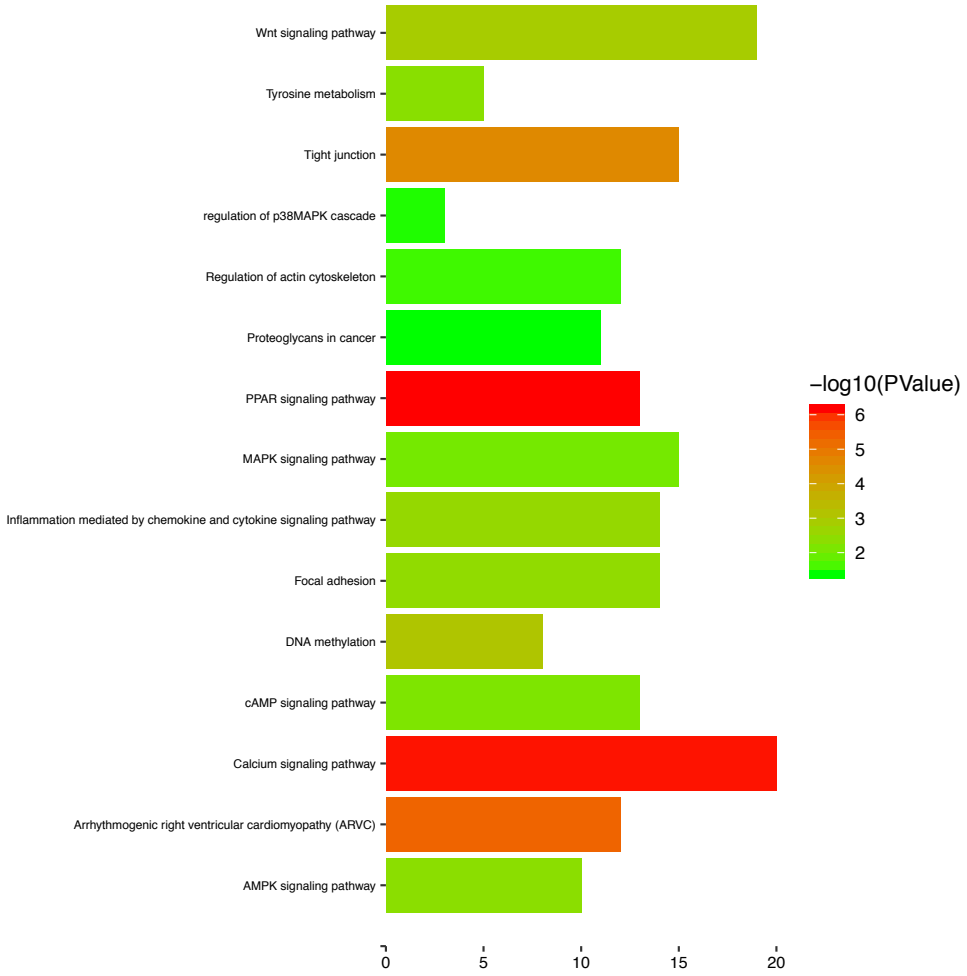

Figure 9 **The enriched KEGG pathways for the turquoise module.**
Full-size DOI: 10.7717/peerj.8505/fig-9

to investigate and validate potential key modules and hub genes by bioinformatics analysis with WGCNA.

WGCNA is a powerful tool for analysing multiple genes in large-scale datasets. It has been extensively used to explore gene co-expression modules and hub genes as potential target biomarkers in many cancers (*Foroughi et al., 2018*; *Chen et al., 2018*; *Giulietti et al., 2016*; *Liu et al., 2019*; *Wang et al., 2019a*; *Wang et al., 2019b*; *Xu et al., 2018*; *Zhai et al., 2019*; *Zhang et al., 2019a*; *Zhang et al., 2019b*; *Zhang et al., 2019c*). In the current study, we applied WGCNA and systematically identified the turquoise module as the most significantly negatively associated ($r = -0.2$, $P = 4e{-}04$) with pathologic stage for the first time. It is well known that the pathologic stage is significantly correlated with the survival of patients and mainly affects the proliferation rate and tissue invasion ability of tumours. One study by *Zhou et al. (2018a)* and *Zhou et al. (2018b)* suggested that patients with a higher pathologic stage were associated with a significantly higher risk of recurrence and

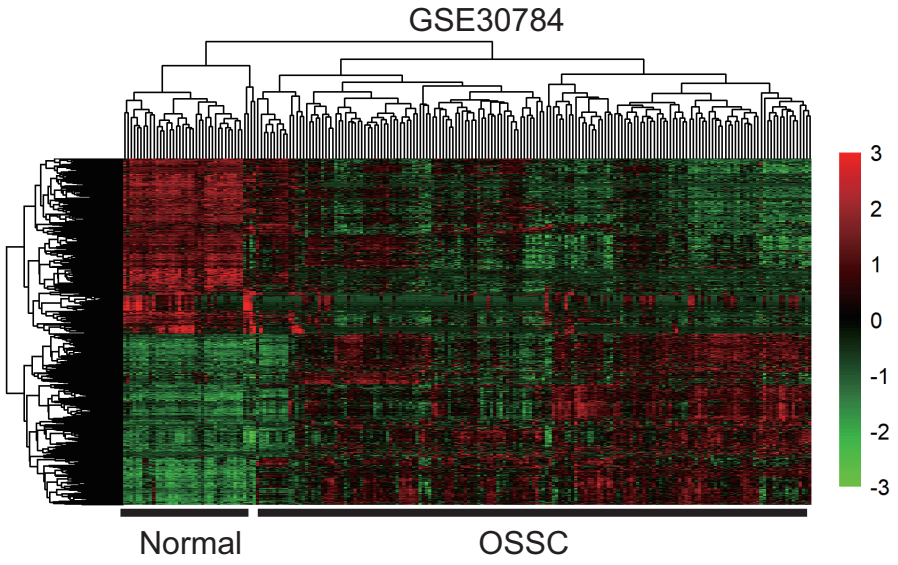

Figure 10 The hierarchical clustering heat map of differently expressed genes in the GEO database (GSE30784). (|log2FC|> 2 and $P < 0.05$). Significantly upregulated RNAs and downregulated RNAs are represented by red lines and green lines, respectively.

Full-size DOI: 10.7717/peerj.8505/fig-10

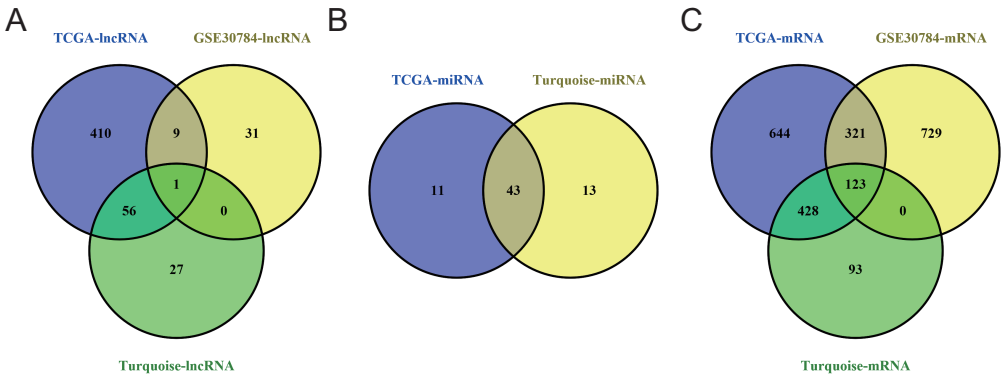

Figure 11 Venn diagram of overlapping genes among the turquoise module and TCGA and GEO (GSE30784) datasets. (A) lncRNA. (B) miRNA. (C) mRNA.

Full-size DOI: 10.7717/peerj.8505/fig-11

worse survival. Moreover, it has been reported that the clinicopathological stage in OSCC has survival implications (*Kılıç et al., 2018*).

Through overlap analysis and Kaplan–Meier survival analysis between the turquoise module and the TCGA and GEO (GSE30784) datasets, 5 common hub genes (PPP1R12B, CFD, CRYAB, FAM189A2 and ANGPTL1) had high connectivity with the overall survival of OSCC patients and were selected from the turquoise module. Survival analyses showed that low expression levels of the 5 hub genes were significantly correlated with poorer prognoses in OSCC patients.

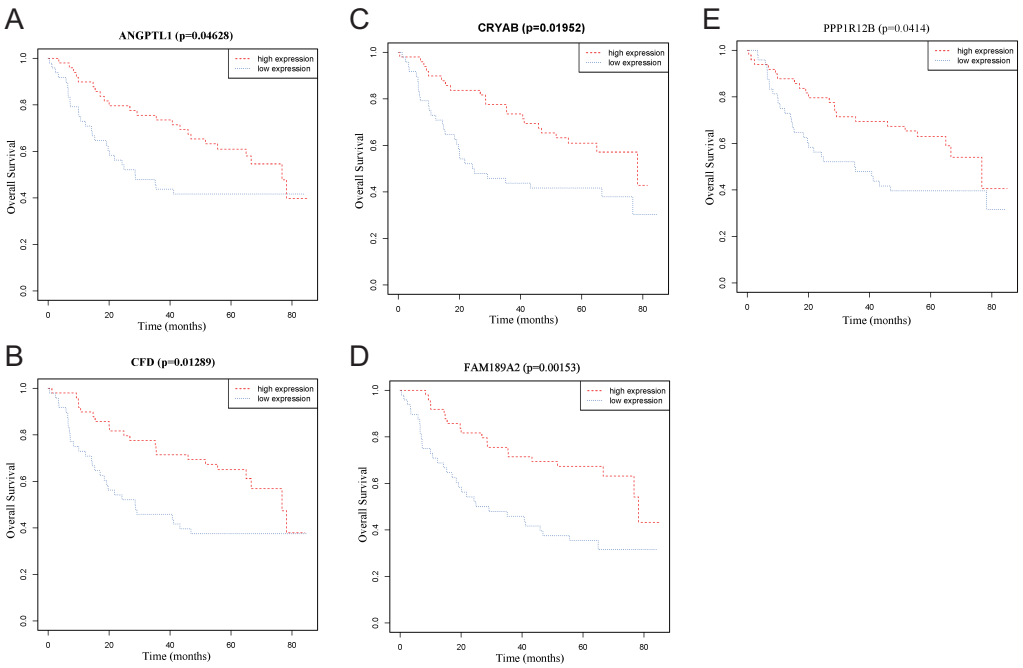

Figure 12 Overall survival analyses of hub genes in the turquoise module. (A) Overall survival analyses of ANGPTL1. (B) Overall survival analyses of CFD. (C) Overall survival analyses of CRYAB. (D) Overall survival analyses of FAM189A2. (E) Overall survival analyses of PPP1R12B. The red line represents samples with high gene expression, and the blue line represents samples with low gene expression.

Full-size DOI: 10.7717/peerj.8505/fig-12

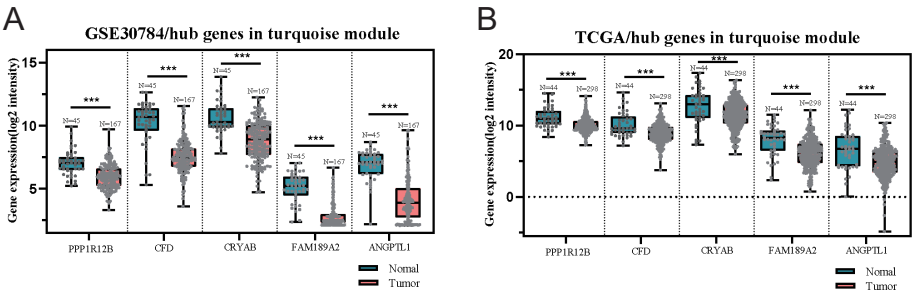

Figure 13 Validation of five hub genes between OSCC samples and normal tissue in the GSE30784 (A) and TCGA (B) datasets. The five hub genes were significantly downregulated in tumourtissues compared with normal samples.

Full-size DOI: 10.7717/peerj.8505/fig-13

At present, it has been reported in previous studies that these five hub genes are related to cancer, and their expression has been confirmed to play an important role in tumourigenesis, the malignant phenotype and disease prognosis. Another study (*Ding et al., 2019*) have also reported that the pseudopodium-enriched atypical kinase 1-PPP1R12B axis inhibits colorectal tumourigenesis and metastasis through the deactivation of the Grb2/PI3K/Akt pathway, which might provide a novel therapeutic strategy for CRC treatment. The hub gene CFD was identified and validated as a potential target gene

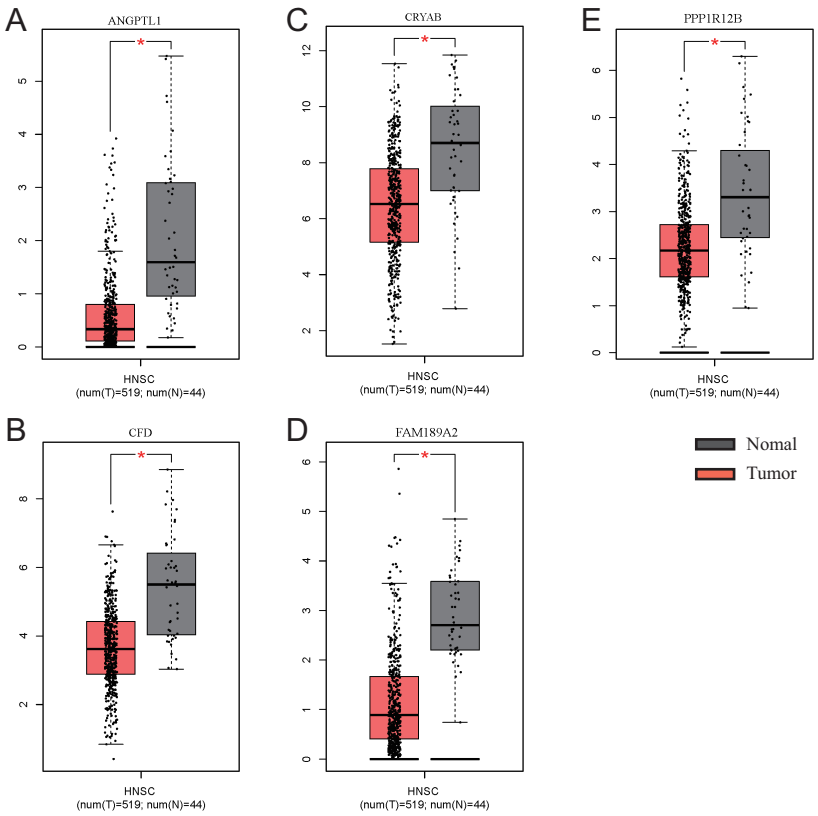

Figure 14 Validation of the gene expression of the five hub genes between OSCC samples and normal tissue from GEPIA. We obtained similar results as above. (A) The gene expression of ANGPTL1 between OSCC samples and normal tissue from GEPIA. (B) The gene expression of CFD between OSCC samples and normal tissue from GEPIA. (C) The gene expression of CRTAB between OSCC samples and normal tissue from GEPIA. (D) The gene expression of FAM189A2 between OSCC samples and normal tissue from GEPIA. (E) The gene expression of PPP1R12B between OSCC samples and normal tissue from GEPIA.

Full-size DOI: 10.7717/peerj.8505/fig-14

in papillary thyroid cancer (*Zhang et al., 2019a*; *Zhang et al., 2019b*; *Zhang et al., 2019c*). CRYAB is a very important protein involved in a variety of signal transduction pathways, including apoptosis, inflammation and oxidative stress (*Zhang et al., 2019a*; *Zhang et al., 2019b*; *Zhang et al., 2019c*). CRYAB is a member of the small heat shock protein family (*Zhang et al., 2019a*; *Zhang et al., 2019b*; *Zhang et al., 2019c*), and many studies have confirmed that CRYAB plays an important role in a variety of tumours, such as OSCC (*Annertz et al., 2014*), colorectal cancer (*Li et al., 2017*), breast cancer (*Kim et al., 2011*), and hepatocellular carcinoma (*Tang et al., 2009*). One study (*Wojtas et al., 2017*) confirmed that the differential expression of FAM189A2 can serve as a gene expression marker in thyroid tumours. ANGPTL1 repressed the migration and invasion of colorectal cancer cells and was inversely correlated with poor survival (*Chen et al., 2017*). The results show that OSCC may be regulated by multiple genes, which will provide more ideas for the evaluation of prognostic value.

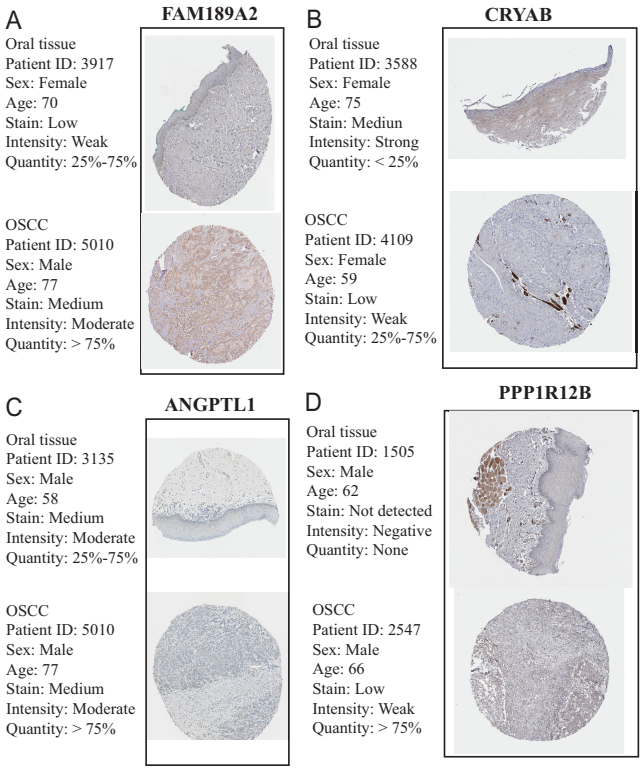

Figure 15 Validation of hub genes in the translational level. (A) Validation of FAM189A2 in turquoise module by The Human Protein Atlas database (IHC). (B) Validation of CRYAB in turquoise module by The Human Protein Atlas database (IHC). (C) Validation of ANGPTL1 in turquoise module by The Human Protein Atlas database (IHC). (D) Validation of PPP1R12B in turquoise module by The Human Protein Atlas database (IHC). There was no related IHC samples for CFD in The Human Protein Atlas database (IHC).

Full-size DOI: 10.7717/peerj.8505/fig-15

In the validation dataset of TCGA and GEO, the results indicated that the 5 hub genes were significantly downregulated in OSCC tissues. Moreover, two of the five genes in the turquoise module were also successfully validated by the HPA database, and the results were consistent with those at the transcriptional level. The above results indicate that the analysis results are reliable and convincing.

To further study the function and pathway regulation mechanism of tumourigenesis, we carried out GO annotation analysis, KEGG pathway analysis and GSEA. Then, some significant biological functions and signalling pathways related to tumourigenesis were identified. Functional annotation analysis showed that the genes were significantly enriched in cell differentiation, RRNA metabolic process, ribosome biogenesis, and cell–cell junctions. Moreover, KEGG pathway analysis showed that the genes are mostly involved in the MAPK signalling pathway, P53 signal pathway, MTOR signal pathway, Notch signal pathway, Wnt signalling pathway and calcium signalling pathway. At the same time, we also found that mutations or abnormal expression levels of these functional annotations and signalling pathways have been reported in OSCC and many other cancers (*Guo et al., 2017*; *Hu et al., 2018*; *Huang et al., 2017*; *Jiang et al., 2011*; *Kim et al., 2019*; *Mo et al., 2019*).

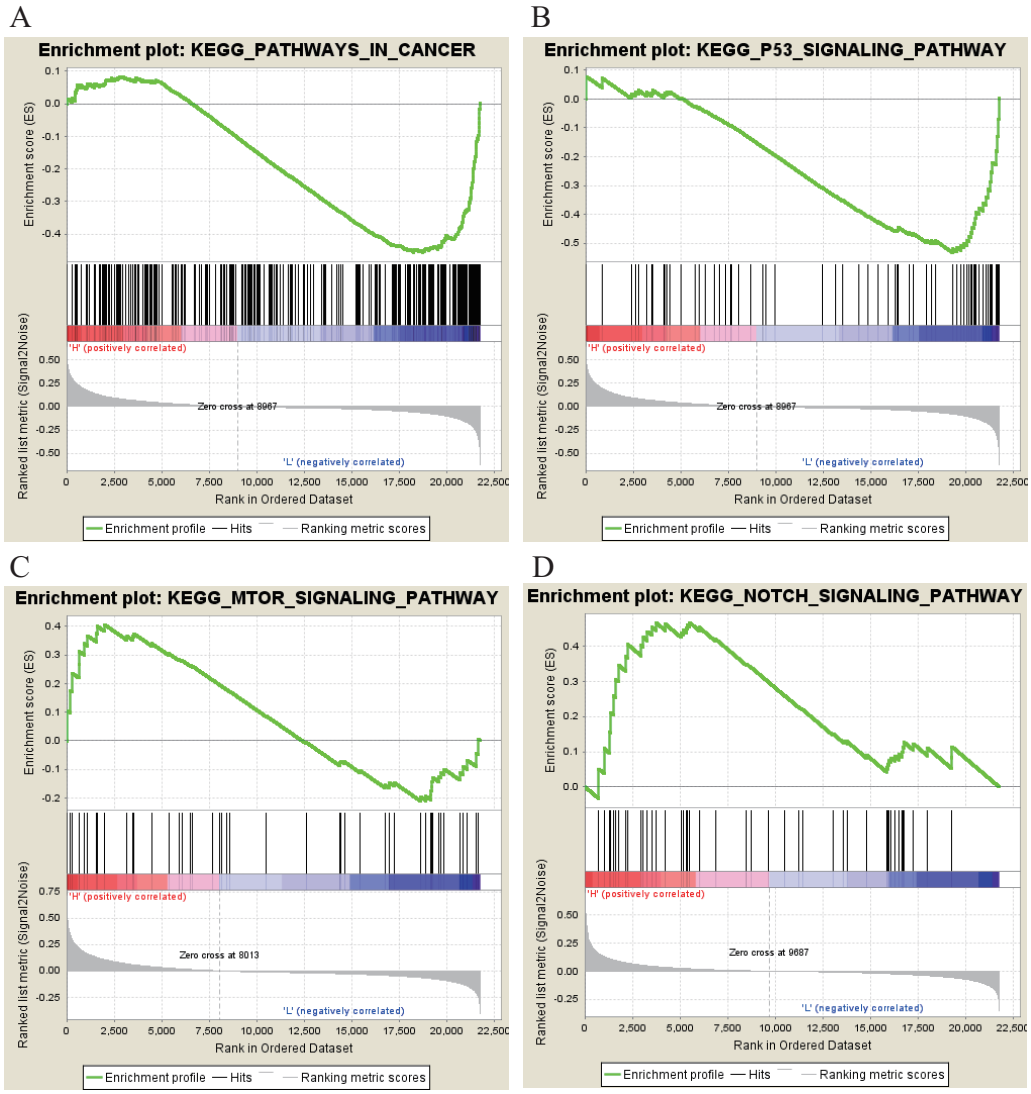

Figure 16 KEGG pathway enrichment analysis of five key genes. (A) Enrichment of genes in the KEGG PATHWAYS IN CANCER by GSEA. (B) Enrichment of genes in the KEGG P53 SIGNALING PATHWAY by GSEA. (C) Enrichment of genes in the KEGG MTOR SIGNALING PATHWAY by GSEA. (D) Enrichment of genes in the KEGG NOTCH SIGNALING PATHWAY by GSEA. The GSEA software was used to calculate enrichment levels.

Full-size DOI: 10.7717/peerj.8505/fig-16

These results provide more clues for further exploring the molecular regulation mechanism of the occurrence and development in OSCC.

CONCLUSIONS

In summary, our research attempts to explore the potential molecular regulatory mechanism of OSCC on the basis of comprehensive bioinformatics analysis. We first discovered that the turquoise module was significantly negatively correlated with the pathologic stage for the first time. Moreover, PPP1R12B, CFD, CRYAB, FAM189A2 and

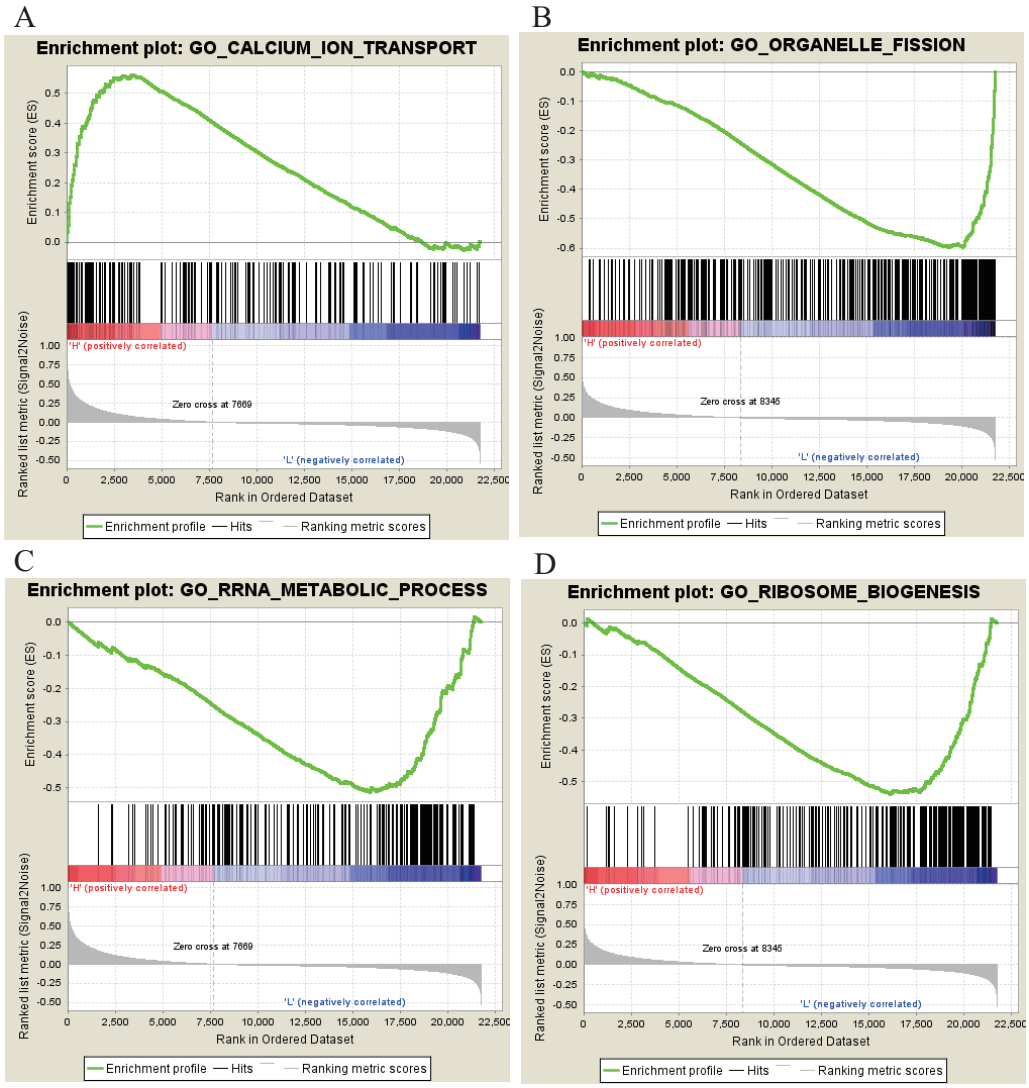

Figure 17 GO enrichment analysis of 5 key genes. (A) Enrichment of genes in GO CALCIUM ION TRANSPORT by GSEA. (B) Enrichment of genes in GO ORGANELLE FISSION by GSEA. (C) Enrichment of genes in GO RRRA METABOLIC PROCESS by GSEA. (D) Enrichment of genes in GO RIBOSOME BIOGENESIS by GSEA. The GSEA software was used to calculate enrichment levels.

Full-size DOI: 10.7717/peerj.8505/fig-17

ANGPTL1, as potential targets in OSCC, were from the turquoise module and their low expression levels were related to the poor survival prognosis of OSCC patients. Despite these findings having enormous potential value, there were some limitations to our study. These results still need further verification by detailed laboratory experiments and large-scale studies.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by Shenzhen Science and Technology Plan Project (No. JCYJ20180302145402866). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Shenzhen Science and Technology Plan Project: JCYJ20180302145402866.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Xuegang Hu conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Guanwen Sun performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Zhiqiang Shi and Hui Ni performed the experiments, prepared figures and/or tables, and approved the final draft.
- Shan Jiang conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw measurements are available in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.8505#supplemental-information.

REFERENCES

Ahluwalia P, Mondal AK, Bloomer C, Fulzele S, Jones K, Ananth S, Gahlay GK, Heneidi S, Rojiani AM, Kota V, Kolhe R. 2019. Identification and clinical validation of a novel 4 gene-signature with prognostic utility in Colorectal Cancer. *International Journal of Molecular Sciences* **20(15)**:E3818 DOI 10.3390/ijms20153818.

Annertz K, Enoksson J, Williams R, Jacobsson H, Coman WB, Wennerberg J. 2014. Alpha B-crystallin—a validated prognostic factor for poor prognosis in squamous cell carcinoma of the oral cavity. *Acta Oto-laryngologica* **134**:543–550 DOI 10.3109/00016489.2013.872293.

Bland RD, Clarke TL, Harden LB. 1976. Rapid infusion of sodium bicarbonate and albumin into high-risk premature infants soon after birth: a controlled, prospective trial. *American Journal of Obstetrics and Gynecology* **124**:263–267 DOI 10.1016/0002-9378(76)90154-x.

Bozec A, Peyrade F, Fischel JL, Milano G. 2009. Emerging molecular targeted therapies in the treatment of head and neck cancer. *Expert Opinion on Emerging Drugs* **14**:299–310 DOI 10.1517/14728210902997947.

Chen H, Boutros PC. 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12**:35 DOI 10.1186/1471-2105-12-35.

Chen H, Xiao Q, Hu Y, Chen L, Jiang K, Tang Y, Tan Y, Hu W, Wang Z, He J, Liu Y, Cai Y, Yang Q, Ding K. 2017. ANGPTL1 attenuates colorectal cancer metastasis by up-regulating microRNA-138. *Journal of Experimental & Clinical Cancer Research: CR* **36**:E78 DOI 10.1186/s13046-017-0548-7.

Chen J, Wang X, Hu B, He Y, Qian X, Wang W. 2018. Candidate genes in gastric cancer identified by constructing a weighted gene co-expression network. *PeerJ* **6**:e4692 DOI 10.7717/peerj.4692.

Ding C, Tang W, Wu H, Fan X, Luo J, Feng J, Wen K, Wu G. 2019. The PEAK1-PPP1R12B axis inhibits tumor growth and metastasis by regulating Grb2/PI3K/Akt signalling in colorectal cancer. *Cancer Letters* **442**:383–395 DOI 10.1016/j.canlet.2018.11.014.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer* **136**:E359–E386 DOI 10.1002/ijc.29210.

Foroughi K, Amini M, Atashi A, Mahmoodzadeh H, Hamann U, Manoochehri M. 2018. Tissue–specific down–regulation of the long non–coding RNAs PCAT18 and LINC01133 in gastric cancer development. *International Journal of Molecular Sciences* **19(12)**:E3881 DOI 10.3390/ijms19123881.

Giulietti M, Occhipinti G, Principato G, Piva F. 2016. Weighted gene co-expression network analysis reveals key genes involved in pancreatic ductal adenocarcinoma development. *Cellular Oncology* **39**:379–388 DOI 10.1007/s13402-016-0283-7.

Guo C, Hou J, Ao S, Deng X, Lyu G. 2017. HOXC10 up-regulation promotes gastric cancer cell proliferation and metastasis through MAPK pathway. *Chinese Journal of Cancer Research = Chung-kuo yen cheng yen chiu* **29**:572–580 DOI 10.21147/j.issn.1000-9604.2017.06.12.

Hinchcliff E, Paquette C, Roszik J, Kelting S, Stoler MH, Mok SC, Yeung TL, Zhang Q, Yates M, Peng W, Hwu P, Jazaeri A. 2019. Lymphocyte-specific kinase expression is a prognostic indicator in ovarian cancer and correlates with a prominent B cell transcriptional signature. *Cancer Immunology, Immunotherapy* **68**:1515–1526 DOI 10.1007/s00262-019-02385-x.

Hu X, Qiu Z, Zeng J, Xiao T, Ke Z, Lyu H. 2018. A novel long non-coding RNA, AC012456.4, as a valuable and independent prognostic biomarker of survival in oral squamous cell carcinoma. *PeerJ* **6**:e5307 DOI 10.7717/peerj.5307.

Huang Q, Cao H, Zhan L, Sun X, Wang G, Li J, Guo X, Ren T, Wang Z, Lyu Y, Liu B, An J, Xing J. 2017. Mitochondrial fission forms a positive feedback loop with cytosolic calcium signaling pathway to promote autophagy in hepatocellular carcinoma cells. *Cancer Letters* **403**:108–118 DOI 10.1016/j.canlet.2017.05.034.

Jiang R, Shi Z, Johnson JJ, Liu Y, Stack MS. 2011. Kallikrein-5 promotes cleavage of desmoglein-1 and loss of cell–cell cohesion in oral squamous cell carcinoma. *The Journal of Biological Chemistry* **286**:9127–9135 DOI 10.1074/jbc.M110.191361.

Kamangar F, Dores GM, Anderson WF. 2006. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* **24**:2137–2150 DOI 10.1200/JCO.2005.05.2308.

Kılıç SS, Kılıç S, Crippen MM, Varughese D, Eloy JA, Baredes S, Mahmoud OM, Park R. 2018. Predictors of clinical-pathologic stage discrepancy in oral cavity squamous cell carcinoma: a National Cancer Database study. *Head & Neck* **40**:828–836 DOI 10.1002/hed.25065.

Kim HM, Kang YH, Byun JH, Jang SJ, Rho GJ, Lee JS, Park BW. 2017. Midkine and NANOG have similar immunohistochemical expression patterns and contribute equally to an adverse prognosis of oral squamous cell carcinoma. *International Journal of Molecular Sciences* **18(11)**:E2339 DOI 10.3390/ijms18112339.

Kim HS, Lee Y, Lim YA, Kang HJ, Kim LS. 2011. αB-Crystallin is a novel oncoprotein associated with poor prognosis in Breast Cancer. *Journal of Breast Cancer* **14**:14–19 DOI 10.4048/jbc.2011.14.1.14.

Kim SA, Lee KH, Lee DH, Lee JK, Lim SC, Joo YE, Chung IJ, Noh MG, Yoon TM. 2019. Receptor tyrosine kinase, RON, promotes tumor progression by regulating EMT and the MAPK signaling pathway in human oral squamous cell carcinoma. *International Journal of Oncology* **55**:513–526 DOI 10.3892/ijo.2019.4836.

Lai Y. 2017. A statistical method for the conservative adjustment of false discovery rate (*q*-value). *BMC Bioinformatics* **18**:69 DOI 10.1186/s12859-017-1474-6.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**:559 DOI 10.1186/1471-2105-9-559.

Li Q, Wang Y, Lai Y, Xu P, Yang Z. 2017. HspB5 correlates with poor prognosis in colorectal cancer and prompts epithelial-mesenchymal transition through ERK signaling. *PLOS ONE* **12**:e0182588 DOI 10.1371/journal.pone.0182588.

Liu B, Huang G, Zhu H, Ma Z, Tian X, Yin L, Gao X, He X. 2019. Analysis of gene co-expression network reveals prognostic significance of CNFN in patients with head and neck cancer. *Oncology Reports* **41**:2168–2180.

McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* **40**:4288–4297 DOI 10.1093/nar/gks042.

Mehrotra R, Gupta DK. 2011. Exciting new advances in oral cancer diagnosis: avenues to early detection. *Head & Neck Oncology* **3**:E33 DOI 10.1186/1758-3284-3-33.

Mo Y, Wang Y, Zhang L, Yang L, Zhou M, Li X, Li Y, Li G, Zeng Z, Xiong W, Xiong F, Guo C. 2019. The role of Wnt signaling pathway in tumor metabolic reprogramming. *Journal of Cancer* **10**:3789–3797 DOI 10.7150/jca.31166.

Omar EA. 2013. The outline of prognosis and new advances in diagnosis of Oral Squamous Cell Carcinoma (OSCC): review of the literature. *Journal of Oral Oncology* **2013**:519312 DOI 10.1155/2013/519312.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**:139–140 DOI 10.1093/bioinformatics/btp616.

Scott SE, Grunfeld EA, McGurk M. 2005. The idiosyncratic relationship between diagnostic delay and stage of oral squamous cell carcinoma. *Oral Oncology* **41**:396–403 DOI 10.1016/j.oraloncology.2004.10.010.

Shi K, Bing ZT, Cao GQ, Guo L, Cao YN, Jiang HO, Zhang MX. 2015. Identify the signature genes for diagnose of uveal melanoma by weight gene co-expression network analysis. *International Journal of Ophthalmology* **8**:269–274 DOI 10.3980/j.issn.2222-3959.2015.02.10.

Sun CC, Zhang L, Li G, Li SJ, Chen ZL, Fu YF, Gong FY, Bai T, Zhang DY, Wu QM, Li DJ. 2017. The lncRNA PDIA3P interacts with mir-185-5p to modulate oral squamous cell carcinoma progression by targeting cyclin D2. *Molecular Therapy. Nucleic Acids* **9**:100–110 DOI 10.1016/j.omtn.2017.08.015.

Tang Q, Liu YF, Zhu XJ, Li YH, Zhu J, Zhang JP, Feng ZQ, Guan XH. 2009. Expression and prognostic significance of the alpha B-crystallin gene in human hepatocellular carcinoma. *Human Pathology* **40**:300–305 DOI 10.1016/j.humpath.2008.09.002.

Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, Von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, Von Heijne G, Nielsen J, Pontén F. 2015. Proteomics. Tissue-based map of the human proteome. *Science* **347**:1260419 DOI 10.1126/science.1260419.

Verusingam ND, Yeap SK, Ky H, Paterson IC, Khoo SP, Cheong SK, Ong A, Kamarul T. 2017. Susceptibility of Human Oral Squamous Cell Carcinoma (OSCC) H103 and H376 cell lines to Retroviral OSKM mediated reprogramming. *PeerJ* **5**:e3174 DOI 10.7717/peerj.3174.

Wang Y, Chen L, Wang G, Cheng S, Qian K, Liu X, Wu CL, Xiao Y, Wang X. 2019a. Fifteen hub genes associated with progression and prognosis of clear cell renal cell carcinoma identified by coexpression analysis. *Journal of Cellular Physiology* **234**:10225–10237 DOI 10.1002/jcp.27692.

Wang Y, Fu J, Wang Z, Lv Z, Fan Z, Lei T. 2019b. Screening key lncRNAs for human lung adenocarcinoma based on machine learning and weighted gene co-expression

network analysis. *Cancer Biomarkers: Section A of Disease Markers* **25**:313–324 DOI 10.3233/CBM-190225.

Wojtas B, Pfeifer A, Oczko-Wojciechowska M, Krajewska J, Czarniecka A, Kukulska A, Eszlinger M, Musholt T, Stokowy T, Swierniak M, Stobiecka E, Chmielik E, Rusinek D, Tyszkiewicz T, Halczok M, Hauptmann S, Lange D, Jarzab M, Paschke R, Jarzab B. 2017. Gene expression (mRNA) markers for differentiating between malignant and benign follicular thyroid tumours. *International Journal of Molecular Sciences* **18(6)**:E1184 DOI 10.3390/ijms18061184.

Xu P, Yang J, Liu J, Yang X, Liao J, Yuan F, Xu Y, Liu B, Chen Q. 2018. Identification of glioblastoma gene prognosis modules based on weighted gene co-expression network analysis. *BMC Medical Genomics* **11**:96 DOI 10.1186/s12920-018-0407-1.

Zhai T, Muhanhali D, Jia X, Wu Z, Cai Z, Ling Y. 2019. Identification of gene co-expression modules and hub genes associated with lymph node metastasis of papillary thyroid cancer. *Endocrine* **66(3)**:573–584 DOI 10.1007/s12020-019-02021-9.

Zhang B, Horvath S. 2005. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology* **4**:Article 17 DOI 10.2202/1544-6115.1128.

Zhang H, Guo L, Zhang Z, Sun Y, Kang H, Song C, Liu H, Lei Z, Wang J, Mi B, Xu Q, Guan H, Li F. 2019a. Co-expression network analysis identified gene signatures in osteosarcoma as a predictive tool for lung metastasis and survival. *Journal of Cancer* **10**:3706–3716 DOI 10.7150/jca.32092.

Zhang J, Liu J, Wu J, Li W, Chen Z, Yang L. 2019b. Progression of the role of CRYAB in signaling pathways and cancers. *OncoTargets and Therapy* **12**:4129–4139 DOI 10.2147/OTT.S201799.

Zhang K, Liu J, Li C, Peng X, Li H, Li Z. 2019c. Identification and validation of potential target genes in papillary thyroid cancer. *European Journal of Pharmacology* **843**:217–225 DOI 10.1016/j.ejphar.2018.11.026.

Zhou S, Liu S, Zhang L, Guo S, Shen J, Li Q, Yang H, Feng Y, Liu M, Lin SH, Xi M. 2018a. Recurrence risk based on pathologic stage after neoadjuvant chemoradiotherapy in esophageal squamous cell carcinoma: implications for risk-based postoperative surveillance strategies. *Annals of Surgical Oncology* **25**:3639–3646 DOI 10.1245/s10434-018-6736-7.

Zhou Z, Cheng Y, Jiang Y, Liu S, Zhang M, Liu J, Zhao Q. 2018b. Ten hub genes associated with progression and prognosis of pancreatic carcinoma identified by co-expression analysis. *International Journal of Biological Sciences* **14**:124–136 DOI 10.7150/ijbs.22619.