

Application of weighted gene co-expression network analysis to identify key modules and hub genes in oral squamous cell carcinoma tumorigenesis

Xiaoqi Zhang*

Hao Feng*

Ziyu Li

Dongfang Li

Shanshan Liu

Haiyun Huang

Minqi Li

Department of Bone Metabolism,
School of Stomatology, Shandong
University, Shandong Provincial
Key Laboratory of Oral Tissue
Regeneration, Jinan, China

*These authors contributed equally
to this work

Purpose: Oral squamous cell carcinoma (OSCC) is one of the most common malignant diseases worldwide, yet its molecular mechanisms are largely unknown. We aimed to construct gene co-expression networks to identify key modules and hub genes involved in the pathogenesis of OSCC.

Patients and methods: We used dataset GSE30784 to construct co-expression networks by weighted gene co-expression network analysis (WGCNA). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed by Database for Annotation, Visualization and Integrated Discovery (DAVID). Hub genes were screened and validated by other datasets.

Results: Turquoise and brown modules were found to be the most significantly related to tumorigenesis. Functional enrichment analysis showed that the turquoise module was associated with cell-cell adhesion, extracellular matrix and collagen catabolic process. A total of 10 hub genes (*MMP1*, *TNFRSF12A*, *PLAU*, *FSCN1*, *PDPN*, *KRT78*, *EVPL*, *GGT6*, *SMIM5* and *CYSRT1*) were identified and validated at transcriptional and translational levels. Their genetic alteration and survival analysis were also revealed.

Conclusion: We identified two modules and 10 hub genes, which were associated with the tumorigenesis of OSCC. The two modules provided references that will advance the understanding of mechanisms of tumorigenesis in OSCC. Moreover, the hub genes may serve as biomarkers and therapeutic targets for precise diagnosis and treatment of OSCC in the future.

Keywords: oral squamous cell carcinoma, co-expression, WGCNA, hub gene

Introduction

Head and neck cancer is a type of upper gastrointestinal cancer based on anatomical structure. Head and neck squamous cell carcinoma (HNSC) is a common malignancy in the world, which accounts for 90% of head and neck cancers, and oral squamous cell carcinoma (OSCC) is the main subtype of HNSC. OSCC is characterized by high incidence of masses in the oral cavity, which leads to high mortality.¹ The incidence of OSCC is associated with smoking, drinking and diet. In addition, viral infections, especially human papillomavirus (HPV), and the lack of vitamins and trace elements such as folic acid; vitamins A, C and E; zinc and selenium are also important etiological factors that can contribute to the incidence of OSCC.²⁻⁴ Accumulation of genetic alterations in oral epithelial cells is also considered to trigger the initiation and progression of OSCC.⁵ Considering recent developments in molecular biological research, it has been

Correspondence: Minqi Li; Haiyun Huang
Department of Bone Metabolism, School
of Stomatology, Shandong University,
Shandong Provincial Key Laboratory of
Oral Tissue Regeneration Wenhua West
Road 44-1, Jinan 250012, China
Tel +86 531 8838 2095;
+86 531 8838 2595
Fax +86 531 8838 2095;
+86 531 8838 2595
Email liminqi@sdu.edu.cn;
seacloud@sdu.edu.cn

discovered that the occurrence and development of OSCC are related to the activation of oncogenes such as *Bcl-2*, *c-Myc* and *EGFR*^{6,7} and the inactivation of *p53*, *Rb*, *p16* and *p21* tumor suppressor genes.^{8,9} However, the occurrence of tumors is a complex and progressive process that is multifactorial, multistage and multistep. The molecular pathogenesis of OSCC remains unclear, and therefore, it is difficult to prevent the pathogenesis of OSCC and improve the survival rate of OSCC patients. In order to diagnose OSCC at its early stage and improve its treatment effect, better understanding of the pathomechanism of OSCC development is needed.

With the discovery and development of high-throughput research methods, systematic description, high-throughput data analysis and screening of important information are the basis for subsequent research. Previous research has been largely focused on studying individual molecules with little focus on functional networks. However, the development of cancer is a systematic biological process (BP) that traverses different functional networks. Weighted gene co-expression network analysis (WGCNA)¹⁰ is a new tool that is used to analyze the potential gene modules that function in the gene expression data. It has been reported that WGCNA can be used to analyze many BPs, such as genetics, multiple cancers and brain imaging data analysis,¹¹ and hence can be useful to identify candidate biomarkers or therapeutic targets. Besides, it can be applied in comparing differentially expressed genes and help investigating the interactions among genes in different modules.¹² WGCNA has been successfully applied in various types of cancer, such as breast cancer,¹³ glioblastoma¹⁴ and prostate cancer,¹⁵ in which WGCNA is used to investigate the relationship between tissue microarray data and clinical traits to identify rules for predicting survival outcome of patients.

In this study, WGCNA was constructed on the basis of data from GSE30784, which included 167 OSCC samples, 17 dysplasia samples and 45 normal samples. Key gene modules associated with tumorigenesis of OSCC were identified, and the biological functions and pathways of genes in the two modules were detected and analyzed. Hub genes in turquoise and brown modules were also revealed and were validated by other datasets of GSE74530 and TCGA. Information on their genetic alterations and clinical features were also revealed. We postulated that these genes and modules might be potential biomarkers for the diagnosis and treatment of OSCC and might contribute to our understanding of the molecular mechanisms of its tumorigenesis.

Patients and methods

Data information

The OSCC datasets of GSE30784 and GSE74530 were obtained from NCBI Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>). GSE30784 consists of 167 OSCC samples, 17 dysplasia samples and 45 normal samples, and the platform is [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. As for the GSE74530, it contains six OSCC samples and six normal samples, and its platform is also [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. The R packages *affy* and *annotate* were used to process the raw data, make expression matrix and match the probe to their gene symbol.

Construction of WGCNA

The raw data of GSE30784 were downloaded from the GEO database. We used the *affy* package (under the R environment, version 3.4.3) to preprocess and normalize (RMA normalization) the raw data (.CEL files). The parameters were set as RMA (for background correction) and *impute* (for supplemental missing value). Ranked by SD from large to small (including normal, dysplasia and OSCC samples), we chose the top 5,000 genes for WGCNA. R package WGCNA was conducted, and the power parameter was pre-calculated by the *pickSoftThreshold* function. It can provide the appropriate soft-thresholding power for network construction by calculating the scale-free topology fit index for several powers. Turning adjacency into topological overlap was carried out, which could measure the network connectivity of a gene defined as the sum of its adjacency with all other genes for the network generation. Hierarchical clustering function was used to classify genes with similar expression profile into modules based on the TOM dissimilarity with a minimum size of 50 for the gene dendrogram. The dissimilarity of module eigengenes was calculated to choose a cutline to merge some modules. The gene network was visualized with randomly selected 1,000 genes. Last visualization in gene network of eigengenes was also carried out.

Identification of clinical significant modules

We calculated the correlation between MEs and clinical trait to identify the relevant module. Afterward, gene significance (GS) was defined as the \log_{10} transformation of the *P*-value ($GS = \lg P$) in the linear regression between gene expression and clinical information. In addition, module significance

(MS) was defined as the average GS for all the genes in a module. In general, the module with the absolute MS ranked first among all the selected modules was considered as the one related with clinical trait.

Function enrichment analysis

To get further insight into the function of genes in key modules, we performed the Gene Ontology (GO) enrichment analysis for modules by using the online Database for Annotation, Visualization and Integrated Discovery (DAVID; <https://david.ncifcrf.gov/summary.jsp>).^{16,17} The gene lists of modules were uploaded, and we got the results of BP, cellular component (CC), molecular function (MF) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. *P*-value <0.05 was regarded as significant.

Identification of hub genes in key modules

Hub genes that are highly interconnected with nodes in a module have been considered as functionally significant. In this study, key modules were selected and were visualized by Cytoscape and hub genes in networks were screened out by degree. Hub genes ranked top 30 in module networks were chosen as the candidates for further analysis and validation. GEPIA (<http://gepia.cancer-pku.cn/>)¹⁸ is a Web server for analyzing the RNA sequencing expression data of 9,736 tumors and 8,587 normal samples from the TCGA and the GTEx projects, using a standard processing pipeline. Survival analysis of candidate genes was carried out in GEPIA, and the significant results were picked out (logrank *P*<0.05). We chose the top five genes in each module with significant results of survival analysis and ranked them by node degree as hub genes in OSCC tumorigenesis.

Validation of hub genes

The “limma” R package was used to screen the differently expressed genes (DEGs) between OSCC samples and normal samples in dataset GSE74530 for validation. The cutoff value was $\log_2FC > |2|$, adjusted *P*-value <0.01. Volcano plot and hierarchical clustering analysis were carried out by R packages ggplot2 and pheatmap, respectively. Venn diagram was performed by online tool jvenn (<http://jvenn.toulouse.inra.fr/app/example.html>)¹⁹ to overlap the gene in key modules and DEGs. The expression statuses of hub genes in GSE30784, GSE74530 and TCGA HNSC were revealed for validation. *P*-value <0.05 was regarded as significant. The translational-level validation of hub genes was carried out using The Human Protein Atlas database (<https://www.proteinatlas.org/>).²⁰

Genetical alteration of hub genes

CBio Cancer Genomics Portal (<http://www.cbioportal.org/>)^{21,22} is an open platform that provides visualization, analysis and downloads of large-scale cancer genomic datasets of various cancers. Complex cancer genomic profiles can be easily obtained by using the query interface of the portal, enabling researchers to explore and compare genetic alterations across samples. We used CBioPortal to explore genetic alterations connected with the 10 hub genes and their correlation with other famous genes and drugs.

Results

Expression value analysis of microarray data

To construct the gene co-expression networks, the raw data of GSE30784 were downloaded from GEO database. Raw data were preprocessed identically by using R for background correction and normalization. R package annotation were conducted to match probes and gene symbols, and the probes matching several genes were removed, and for the gene matched by multiple probes, the median was regarded as the final expression value. At last, we obtained a total 20,027 genes. We calculated SD for each gene and ranked them from large to small, and then, the top 5,000 genes were chosen for WGCNA. The 5,000 genes were used for cluster analysis by fashClust function of the WGCNA package (Figure 1A). As we can see, 229 samples were divided into two clusters on the whole.

Construction of weighted co-expression network and identification of key modules

Selection of the soft-thresholding power is an important step when constructing a WGCNA. We performed the analysis of network topology for thresholding powers from 1 to 20 and identified the relatively balanced scale independence and mean connectivity of the WGCNA. As shown in Figure 1B and C, power value 9, which was the lowest power for the scale-free topology fit index on 0.9, was selected to produce a hierarchical clustering tree (dendrogram) of 5,000 genes. We set the MEDissThres as 0.25 to merge similar modules (Figure 2A), and 11 modules were generated (Figure 2B). There were 223 genes in the black module, 518 genes in the blue module, 954 genes in the brown module, 139 in genes in the greenyellow module, 172 genes in the magenta module, 302 genes in the pink module, 143 genes in the purple module, 379 genes in the red module, 112 genes in the salmon module, 530 genes in

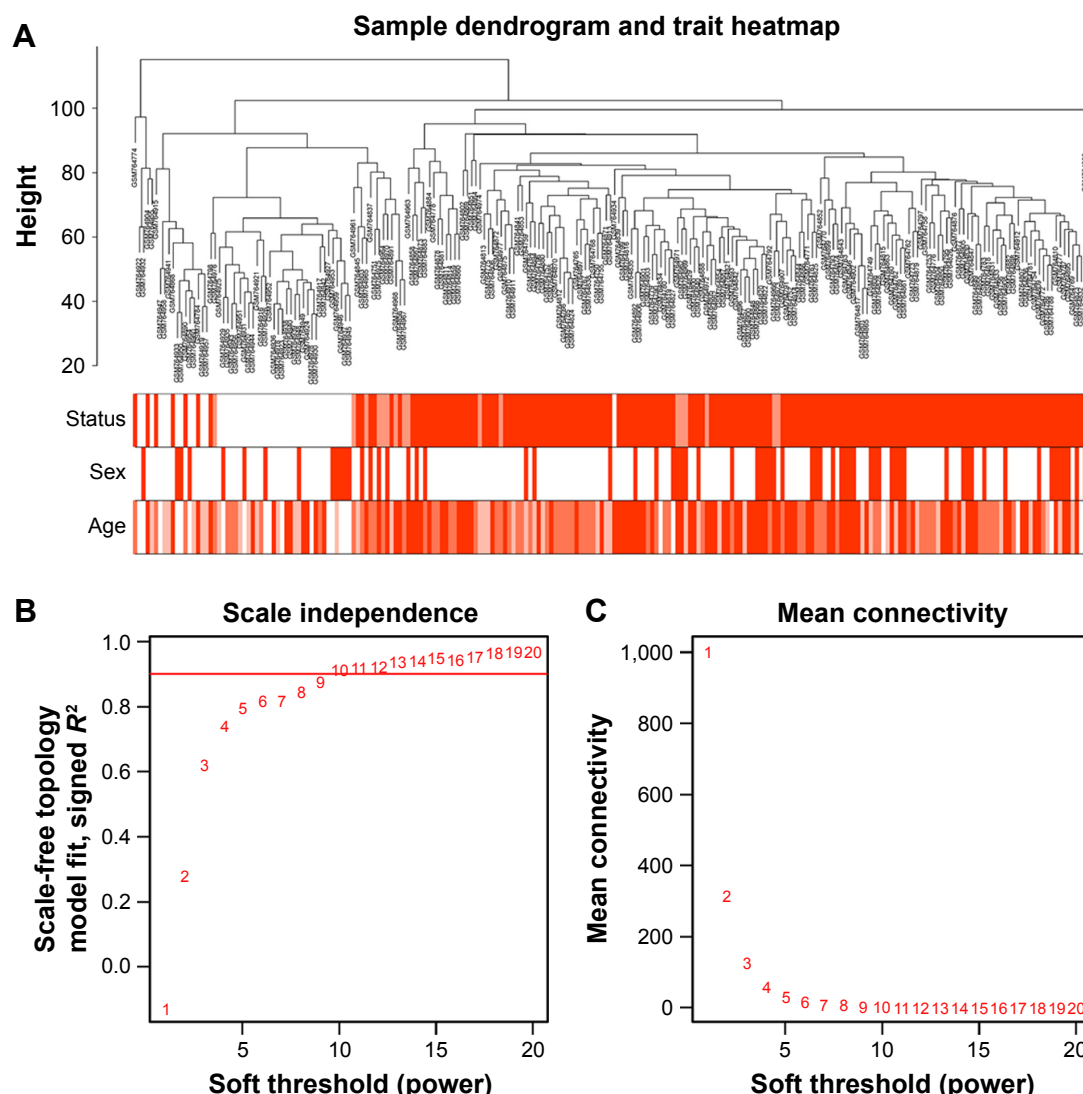


Figure 1 Clustering of samples and determination of soft-thresholding power.

Notes: (A) The clustering was based on the expression data of GSE30784, which contained 167 OSCC, 17 dysplasia and 45 normal samples. The top 5,000 genes with the highest SD values were used for the analysis by WGCNA. The color intensity was proportional to disease status (normal, dysplasia and OSCC), sex and age. (B) Analysis of the scale-free fit index for various soft-thresholding powers (β). (C) Analysis of the mean connectivity for various soft-thresholding powers. In all, 9 was the most fit power value.

Abbreviations: OSCC, oral squamous cell carcinoma; WGCNA, weighted gene co-expression network analysis.

the turquoise module and 471 genes in the yellow module. The genes that could not be included in any modules were put into the gray module, which was removed in the subsequent analysis. Module preservation analyses for the process of normal to dysplasia and dysplasia to neoplasm are shown in Figure S1A and B.

Correlation between modules and identification of key modules

Interaction relationships of the 11 modules were analyzed, and the network heatmap was plotted (Figure 3A). The results revealed that each module was an independent validation to each other, which demonstrated a high level

of independence among the modules and relative independence of genes expression in each module. Besides, we calculated eigengenes and clustered them according to their correlation in order to explore co-expression similarity of all modules (Figure 3C). We found that the 11 modules were mainly divided into two clusters. Similar results were demonstrated by the heatmap plotted according to adjacencies (Figure 3D). Furthermore, the ME of the turquoise module and brown module revealed a high correlation with disease status (normal, dysplasia and OSCC) compared with other modules (Figure 3B). The turquoise module was positively correlated with the disease status, while the brown module was negatively correlated with the disease status, indicating

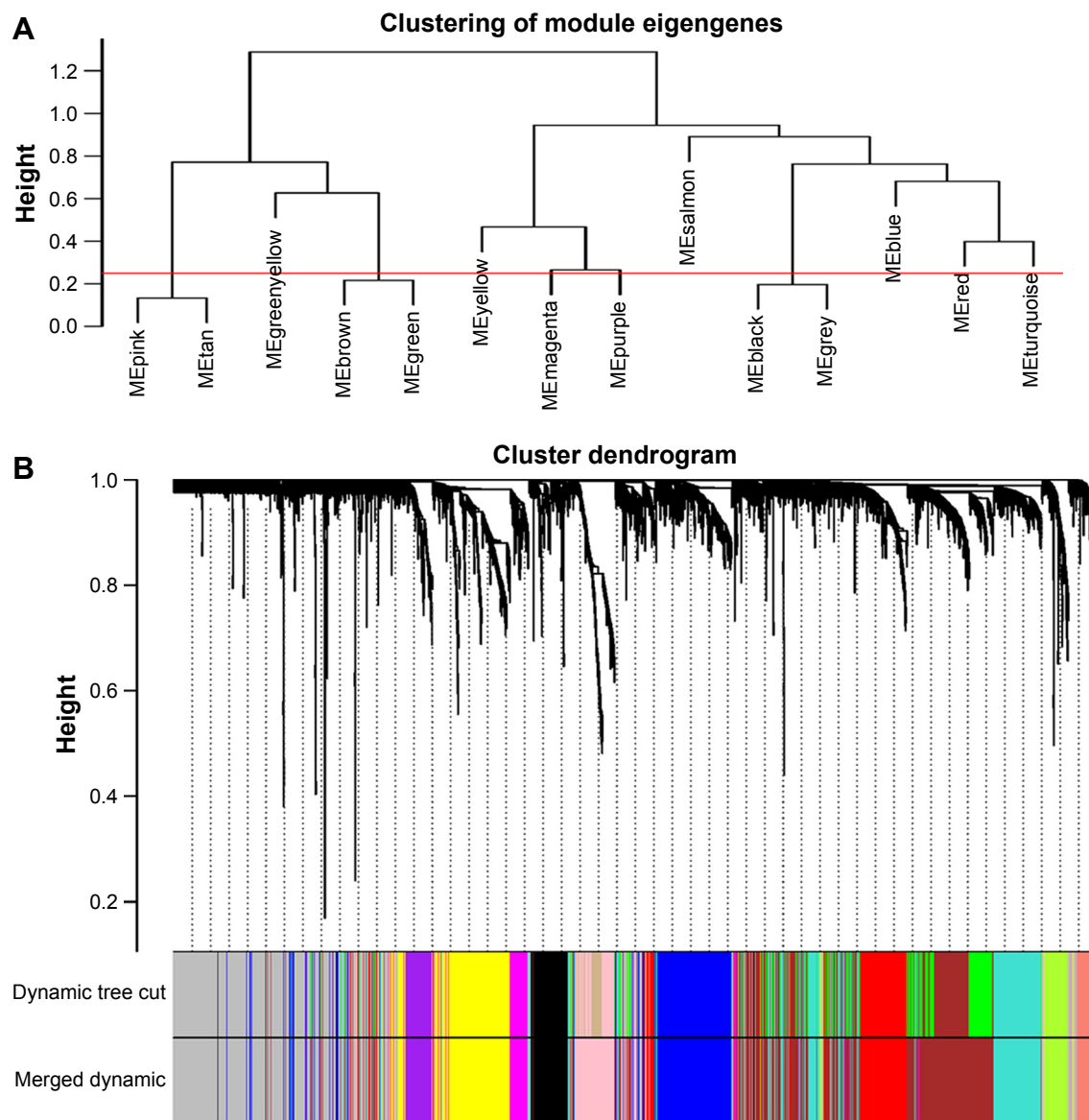


Figure 2 Construction of co-expression modules by WGCNA package in R.

Notes: (A) The cluster dendrogram of module eigengenes. (B) The cluster dendrogram of genes in GSE39784. Each branch in the figure represents one gene, and every color below represents one co-expression module.

Abbreviation: WGCNA, weighted gene co-expression network analysis.

that the turquoise module may play an important role in tumorigenesis of OSCC and the brown module may function as an antitumor. We identified the turquoise module and brown module as the modules most relevant to the disease status of OSCC. Figure 4A and B illustrates the correlation between module membership and GS in turquoise and brown modules, respectively.

Identification of hub genes in the turquoise module and brown module

We visualized the turquoise module and the brown module as networks in Cytoscape and screened out 30 genes by sorting

node degree candidate gene for further analysis. Figure 4C and D shows the top 30 hub genes in the turquoise module and brown module. We used GEPIA to perform survival analysis; $P < 0.05$ was regarded as statistically significant. The genes with a significant result of survival analysis were selected and ranked by node degree. The top five genes in each module were considered as hub genes. Figure 5 shows the survival analysis of hub genes (A–E were hub genes in the turquoise module and F–J were hub genes in the brown module). They were *MMP1*, *TNFRSF12A*, *PLAU*, *FSCN1* and *PDPN* in the turquoise module and *KRT78*, *EVPL*, *GGT6*, *SMIM5* and *CYSRT1* in the brown module. The survival

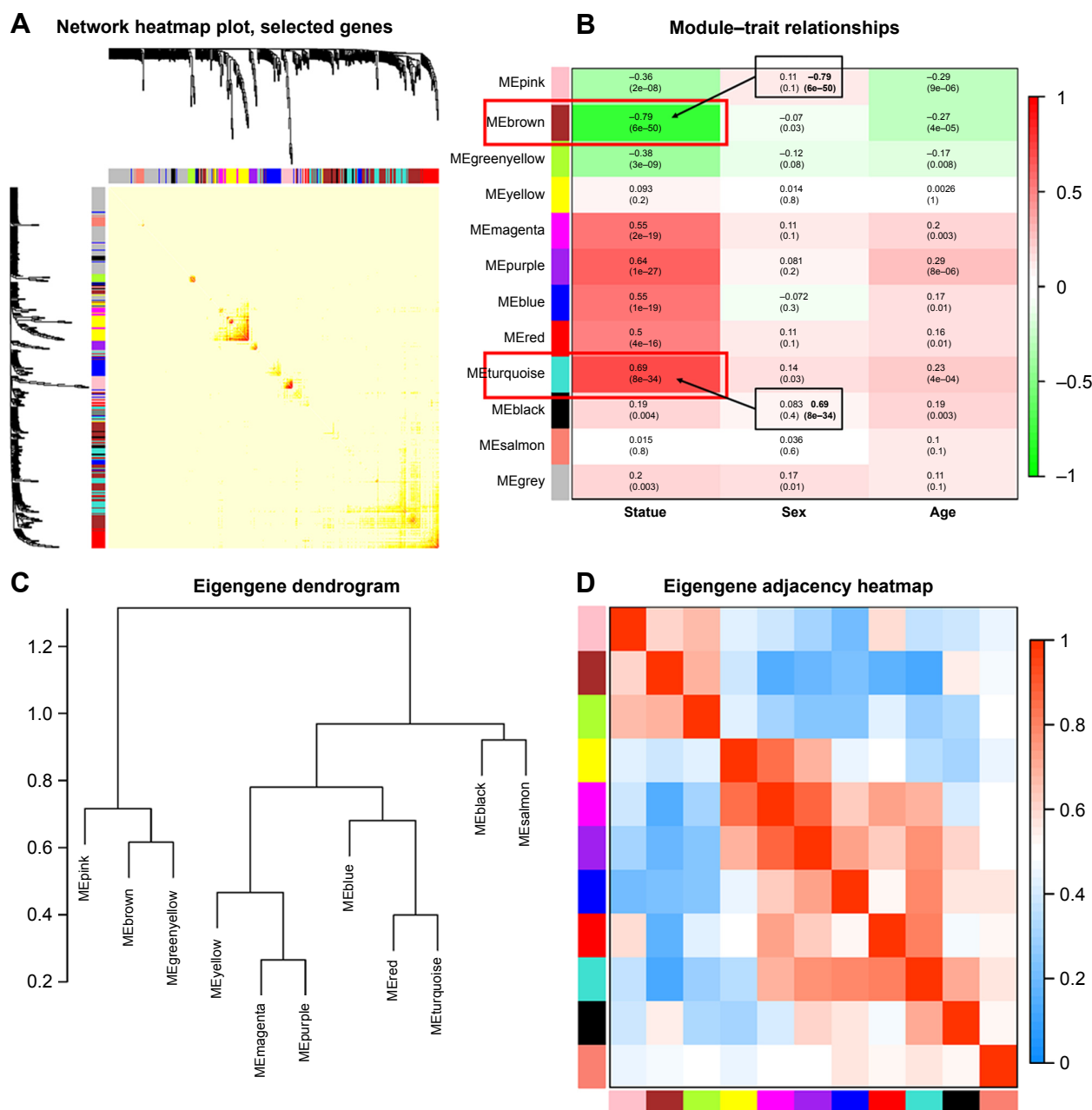


Figure 3 (A) Interaction relationship analysis of co-expression genes. Different colors of horizontal axis and vertical axis represent different modules. The brightness of yellow in the middle represents the degree of connectivity of different modules. There was no significant difference in interactions among different modules, indicating a high-scale independence degree among these modules. (B) Heatmap of the correlation between module eigengenes and the disease status of OSCC. The turquoise module was the most positively correlated with status, and the brown module was the most negatively correlated with status. (C) Hierarchical clustering of module hub genes that summarize the modules yielded in the clustering analysis. (D) Heatmap plot of the adjacencies in the hub gene network.

Abbreviation: OSCC, oral squamous cell carcinoma.

analysis of the 10 hub genes were validated in another GEO dataset, GSE41613 (Figure S3).

Function enrichment analysis in two key modules

We performed enrichment analysis to explore the BP and pathway in which the two key modules were involved.

GO enrichment of BP was conducted by DAVID, and the detailed information is given in Table 1. As the results show, the brown module was mainly enriched in regulation of epidermis such as epidermis development, epithelial cell differentiation and keratinocyte differentiation, which were negatively correlated with tumorigenesis. In the turquoise module, the results of enrichment analysis were mainly

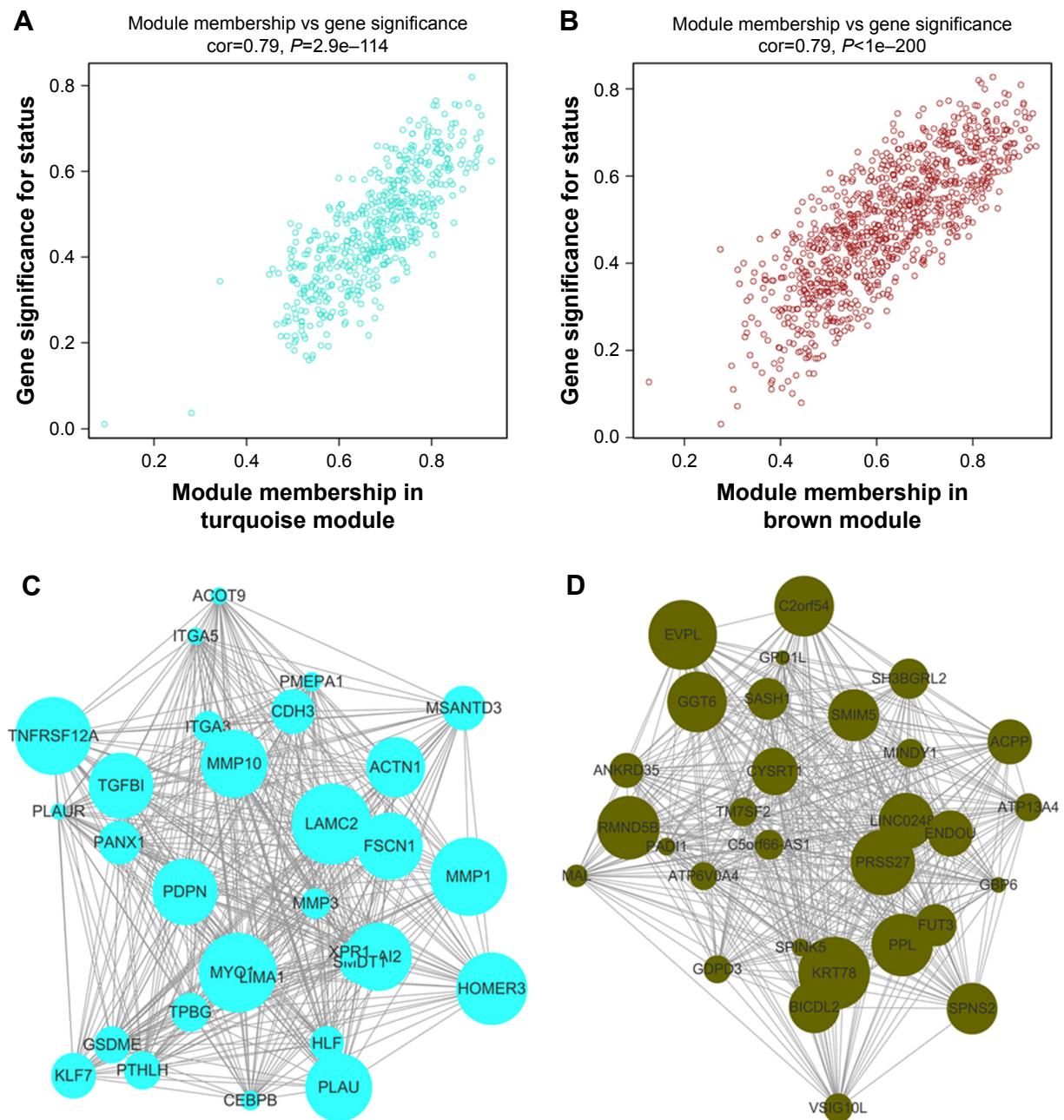


Figure 4 (A) Scatter plot of module eigengenes in the turquoise module. (B) Scatter plot of module eigengenes in the brown module. (C) The top 30 hub genes in the turquoise module. (D) The top 30 hub genes in the brown module. Nodes represent genes, and node size is correlated with connectivity of the gene by degree.

concerned with tumorigenesis of malignancy such as positive regulation of cell migration, angiogenesis, cell adhesion, cell–matrix adhesion and extracellular matrix (ECM) organization, which played an important role in tumorigenesis, invasion, migration and metastasis of OSCC.

As for the pathways in which these two modules were involved, we performed KEGG pathway analysis, and the details of the results are shown in Table 2. As we can see, the brown module enriched in pathways of metabolism of

biological compounds such as steroid biosynthesis, leucine and isoleucine degradation, biosynthesis of antibiotics and glycerolipid metabolism. In the turquoise module, pathways concerning with regulation of cell adhesion were significantly enriched such as focal adhesion, PI3K–Akt signaling pathway and ECM–receptor interaction, which were closely related with malignancy.

A direct comparison of the network constructed from tumor vs normal samples of GSE30784 for providing solid

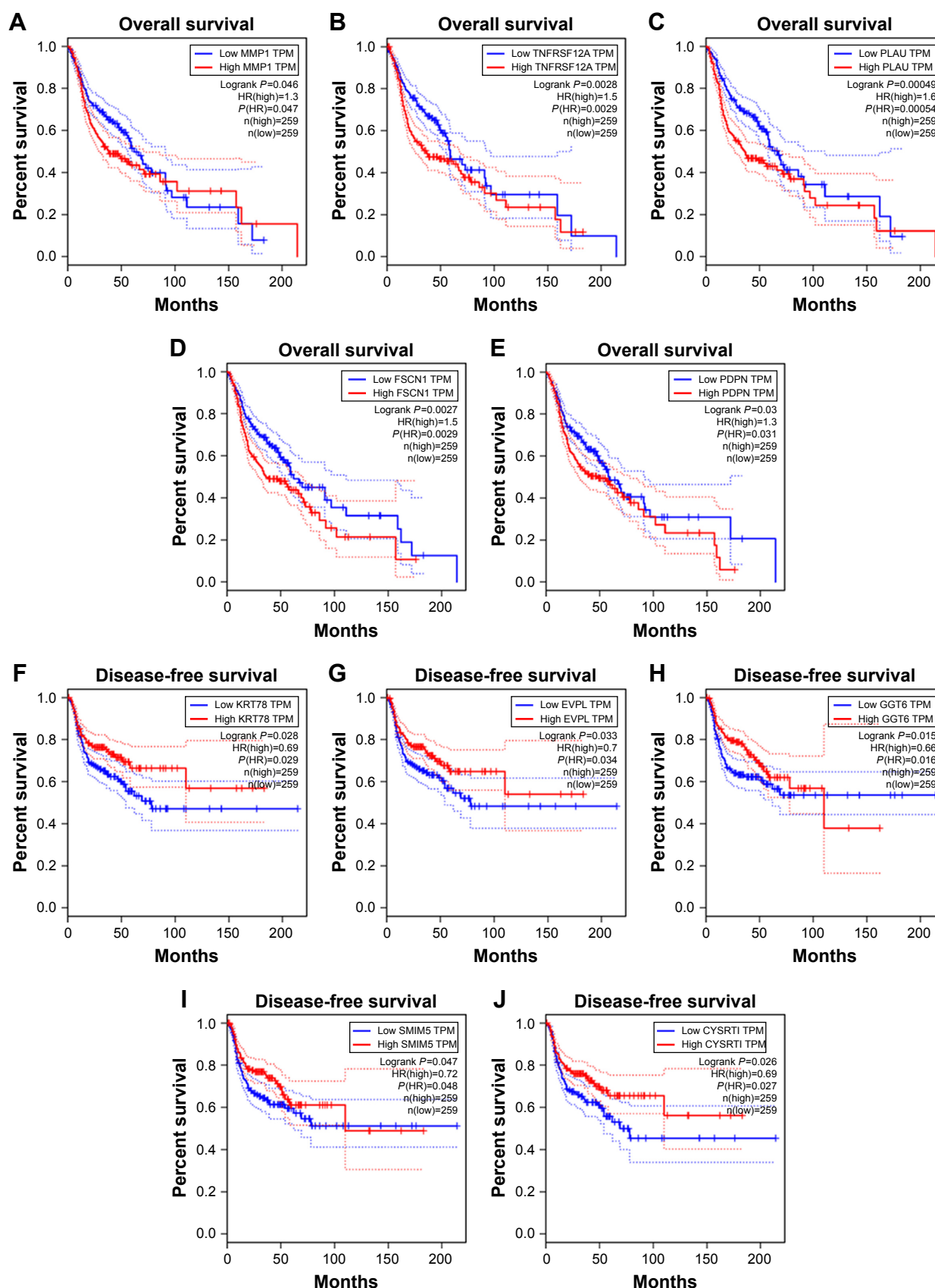


Figure 5 Survival analysis of hub genes.

Notes: (A–E) Five hub genes in the turquoise module with the highest node degree and significant results of survival analysis ($P<0.05$ was regarded as significant). They were *MMP1*, *TNFRSF12A*, *PLAUR*, *FSCN1* and *PDPN*, respectively. (F–J) Five hub genes in the brown module with the highest node degree and significant results of survival analysis ($P<0.05$ was regarded as significant). They were *KRT78*, *EVPL*, *GGT6*, *SMIM5* and *CYSRT1*, respectively.

Table I GO enrichment analysis of turquoise module and brown module

Category	Term	Count	%	P-value
Brown module				
BP	GO:0055114~oxidation–reduction process	66	7.03	2.20E–10
BP	GO:0008544~epidermis development	19	2.02	7.42E–08
BP	GO:0030855~epithelial cell differentiation	14	1.49	2.17E–05
BP	GO:0098609~cell–cell adhesion	29	3.09	9.53E–05
BP	GO:0006805~xenobiotic metabolic process	13	1.38	2.92E–04
BP	GO:0018149~peptide cross-linking	10	1.06	5.07E–04
BP	GO:0030216~keratinocyte differentiation	12	1.27	8.76E–04
BP	GO:0033539~fatty acid beta-oxidation using acyl-CoA dehydrogenase	6	0.63	1.22E–03
BP	GO:0030148~sphingolipid biosynthetic process	9	0.95	1.74E–03
BP	GO:0048662~negative regulation of smooth muscle cell proliferation	7	0.74	2.09E–03
CC	GO:0070062~extracellular exosome	240	25.58	7.17E–22
CC	GO:0005789~endoplasmic reticulum membrane	78	8.31	4.60E–08
CC	GO:0016324~apical plasma membrane	36	3.83	3.62E–07
CC	GO:0030057~desmosome	9	0.95	8.84E–06
CC	GO:0005913~cell–cell adherens junction	34	3.62	2.62E–05
MF	GO:0005198~structural molecule activity	31	3.30	2.55E–06
MF	GO:0016491~oxidoreductase activity	27	2.87	3.38E–06
MF	GO:0098641~cadherin binding involved in cell–cell adhesion	31	3.30	5.91E–05
MF	GO:0005509~calcium ion binding	56	5.97	3.58E–04
MF	GO:0003824~catalytic activity	21	2.23	6.79E–04
Turquoise module				
BP	GO:0050900~leukocyte migration	20	3.82	8.08E–10
BP	GO:0031581~hemidesmosome assembly	8	1.52	7.55E–09
BP	GO:0045926~negative regulation of growth	8	1.52	4.07E–07
BP	GO:0030335~positive regulation of cell migration	19	3.63	3.00E–06
BP	GO:0001525~angiogenesis	21	4.01	3.27E–06
BP	GO:0007155~cell adhesion	32	6.11	3.54E–06
BP	GO:0007160~cell–matrix adhesion	13	2.48	5.85E–06
BP	GO:0030198~extracellular matrix organization	19	3.63	7.27E–06
BP	GO:0071294~cellular response to zinc ion	7	1.33	7.99E–06
BP	GO:0001666~response to hypoxia	17	3.25	2.00E–05
CC	GO:0005925~focal adhesion	47	8.98	1.97E–17
CC	GO:0001726~ruffle	13	2.48	4.79E–06
CC	GO:0001725~stress fiber	10	1.91	1.18E–05
CC	GO:0008305~integrin complex	7	1.33	6.48E–05
CC	GO:0005938~cell cortex	13	2.48	1.14E–04
MF	GO:0005178~integrin binding	14	2.67	4.89E–06
MF	GO:0050431~TGF-beta binding	6	1.147	4.75E–05
MF	GO:0031418~L-ascorbic acid binding	6	1.14	1.98E–04
MF	GO:0001968~fibronectin binding	6	1.14	5.73E–04
MF	GO:0051015~actin filament binding	12	2.29	8.96E–04

Abbreviations: GO, Gene Ontology; BP, biological process; CC, cellular component; MF, molecular function.

insights (Figure S2A and B). The GO enrichment analysis and the KEGG pathway analysis were carried out for the DEGs mentioned earlier (Tables S1 and S2). The enrichment analysis results between modules from WGCNA and DEGs were similar, which means that the results were reliable.

Validation of hub genes

We used another dataset GSE74530 from GEO database to validate the expression status of the 10 hub genes.

We set the cutoff as $\log_{2}FC > |1|$ and $P < 0.05$ to screen DEGs. Figure 6A shows the volcano plot of DEGs, and Figure 6B shows hierarchical clustering heat map of DEGs. We overlapped the DEGs and genes in the turquoise module and brown module by venn diagram and found that the 10 hub genes were all present in DEGs and modules. The results are shown in Figure 6C and D. Figure 7A demonstrates the expression status of 10 hub genes in normal, dysplasia and OSCC samples of GSE30784. As shown in the figure, the

Table 2 KEGG pathway enrichment analysis of turquoise module and brown module

Category	Term	Count	%	P-value
Brown module				
KEGG	hsa01100: metabolic pathways	107	11.41	6.93E-10
KEGG	hsa00100: steroid biosynthesis	8	0.85	3.03E-05
KEGG	hsa00280: valine, leucine and isoleucine degradation	10	1.07	4.11E-04
KEGG	hsa01130: biosynthesis of antibiotics	22	2.35	1.78E-03
KEGG	hsa00561: glycerolipid metabolism	10	1.07	1.99E-03
KEGG	hsa03320: PPAR signaling pathway	10	1.07	5.40E-03
KEGG	hsa04071: sphingolipid signaling pathway	14	1.49	6.16E-03
KEGG	hsa00350: tyrosine metabolism	7	0.75	6.75E-03
KEGG	hsa04530: tight junction	15	1.60	7.62E-03
KEGG	hsa00600: sphingolipid metabolism	8	0.85	7.78E-03
Turquoise module				
KEGG	hsa04510: focal adhesion	27	5.16	3.64E-10
KEGG	hsa04151: PI3K-Akt signaling pathway	30	5.74	3.84E-07
KEGG	hsa04512: ECM-receptor interaction	14	2.68	1.64E-06
KEGG	hsa05200: pathways in cancer	30	5.74	5.52E-06
KEGG	hsa05205: proteoglycans in cancer	20	3.82	7.96E-06
KEGG	hsa05222: small-cell lung cancer	12	2.29	4.27E-05
KEGG	hsa04978: mineral absorption	9	1.72	5.74E-05
KEGG	hsa04810: regulation of actin cytoskeleton	15	2.87	0.004426
KEGG	hsa05202: transcriptional misregulation in cancer	13	2.49	0.004612
KEGG	hsa05412: ARVC	8	1.53	0.005263

Abbreviations: KEGG, Kyoto Encyclopedia of Genes and Genomes; ECM, extracellular matrix; ARVC, arrhythmic right ventricular cardiomyopathy.

expression status was in accordance with the disease status in which expression of five hub genes of the turquoise module was positively correlated with disease status and expression of five hub genes of the brown module was negatively correlated with disease status. Figure 7B illustrates the validation of 10 hub genes in GSE74530, which presents a similar conclusion with GSE30784, that the hub genes in the turquoise module were highly expressed in tumor samples, while hub genes in the brown module were highly expressed in normal samples. We further analyzed TCGA data of HNSCs to validate the results and found similar results as described earlier (Figure 7C). Moreover, immunohistochemistry (IHC) staining obtained from The Human Protein Atlas database also demonstrated the expression status of hub genes and the patient data as shown in Figure 8. Figure 8A–D shows hub genes in the turquoise module, whose results were in accordance with the transcriptional level, but there were no related IHC samples of MMP1 in the database. Figure 8E–I shows hub genes in the brown module, and results were in accordance with transcriptional level before. These results confirmed the validity of the hub genes that we found as reliable.

Genetical alteration of hub genes

OncoPrint of CBioPortal was used to show the 10 hub genes' alteration statuses in TCGA HNSC patients. The 10 hub

genes altered in 222 (44%) of 510 patients (Figure 9B), and the frequency of alteration of each hub gene is shown in Figure 9A. *FSCN1* and *MMP1* altered most (15% and 9%, respectively), amplification, and mRNA upregulation was the main type. Figure 9C demonstrates the relationship of the 10 genes and the other 50 most frequently altered neighbor genes (only *MMP1*, *PLAU*, *FSCN1* and *EVPL* had connection with these 50 genes). *PIK3CA* and *TP63* were significantly important in the network. The correlation of antitumor drugs and hub genes was exhibited; we found *MMP1* and *TNFRSF12A* were the target of cancer drugs, as for the other hub genes, there was no drug targeting to them, which might be the promising target of new cancer drugs.

Discussion

OSCC is a common malignancy caused by abnormal squamous cells, which mainly occurs in older adults. Clinical symptoms are not often present in the early stages of the disease but are observed in the late stages of the disease. Its 5-year survival rate is 50%.²³ In patients with high metastasized cancer, the 5-year survival rate drops to ~20%,²⁴ which can seriously harm human health and life. Although the new and advanced medical methods have greatly improved the quality of life for patients with OSCC, the 5-year survival rate has not yet improved significantly. Therefore, it is of great help for the prevention and treatment of OSCC to study the

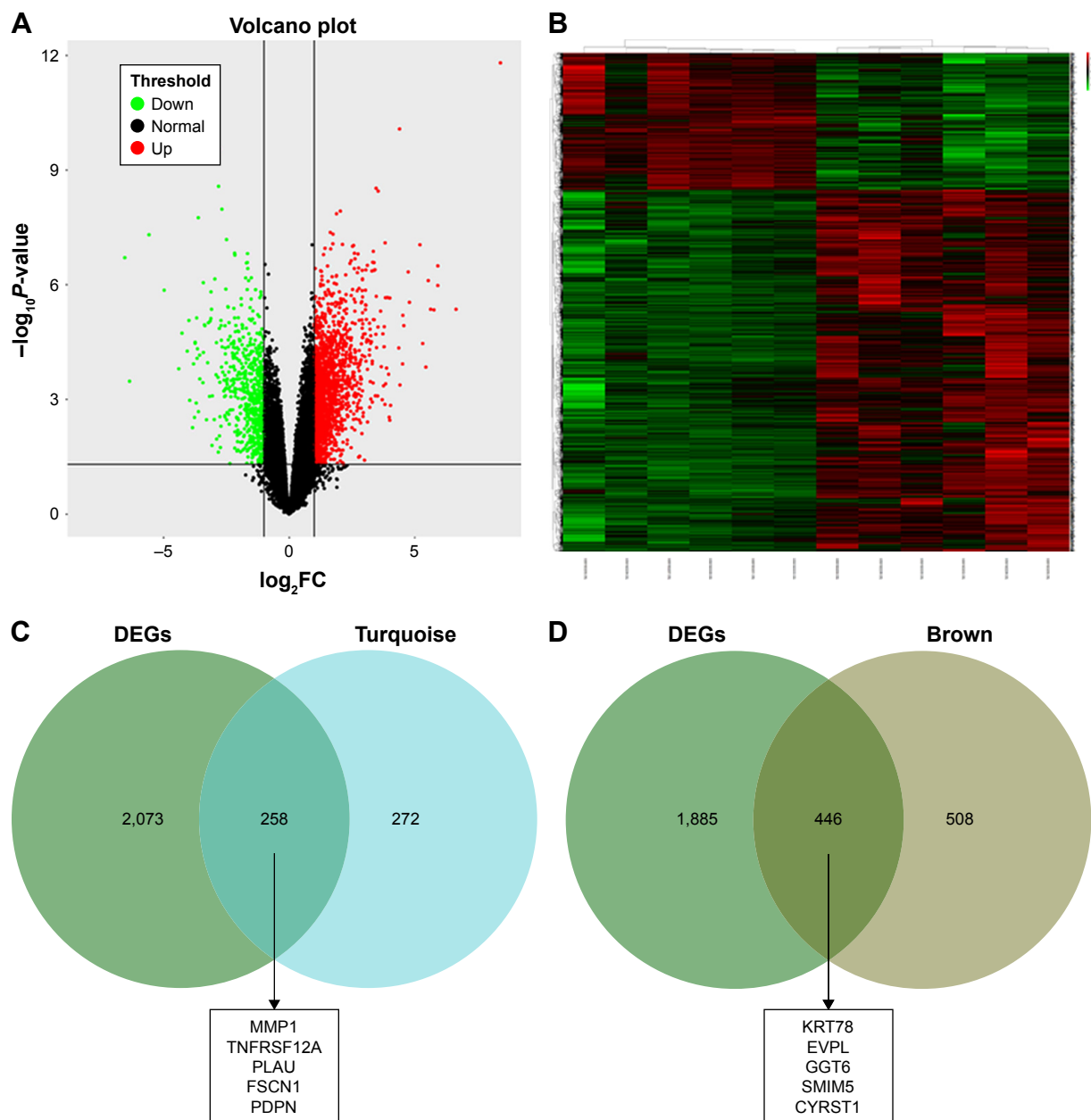


Figure 6 Validation of hub genes in GSE31056.

Notes: (A) Volcano plot visualizing DEGs in GSE31056 (23 normal samples and 21 OSCC). The vertical lines demarcate the fold change values. The right vertical line corresponds to more than or equal to twofold up changes and the left vertical line more than or equal to twofold down changes, while the horizontal line marks a $-\log_{10}$ adjusted P -value of 0.01. (B) Heat map hierarchical clustering reveals DEGs in OSCC groups compared with those in control groups. (C) Identification of common genes between DEGs and the turquoise module by overlapping them. The five hub genes in the turquoise module were also DEGs in GSE31056. (D) Identification of common genes between DEGs and the brown module by overlapping them. The five hub genes in the brown module were also DEGs in GSE31056.

Abbreviations: DEGs, differently expressed genes; OSCC, oral squamous cell carcinoma.

development of OSCC at the molecular level and develop effective measures to prevent and suppress metastasis of OSCC. In the present study, we applied WGCNA to identify the key modules and hub genes in tumorigenesis of OSCC by R and identified two modules that were most significantly associated with the OSCC disease status (the turquoise module was positively correlated with OSCC status, and the

brown module was negatively correlated with the disease status). Functional enrichment analysis of the two modules was performed by DAVID. A total of 10 hub genes with significant correlation with survival analysis in the two modules were identified and were further validated in another dataset and TCGA to confirm the reliability of the results. CBioPortal was used to show the correlation between hub genes and

genetic alteration. The results provide novel insights that will help to explain the pathogenesis of OSCC at the molecular level, and the hub genes might serve as biomarkers as well as therapeutic targets for precise diagnosis and treatment of OSCC in the future.

In this study, we used a global approach to construct co-expression network of OSCC to predict clusters of genes involved in the pathogenesis and tumorigenesis of OSCC. We aimed to find new and critical biomarkers and to understand the molecular mechanism of OSCC, which might contribute to the diagnosis and treatment of the disease. In this study, WGCNA was performed by extracting co-expression networks of group genes from a large expression data.²⁵ GSE30784 was downloaded from the GEO database, and 11 co-expression modules were obtained by WGCNA.

Among the 11 modules, we found that the turquoise module and brown module were most significantly related to the OSCC status, in which the turquoise module was positively correlated and the brown module was negatively correlated. There are many genome-wide gene expressions of OSCC datasets available, and there are many opportunities for researchers to use these data to carry out multiple analyses. However, there are challenges in data analysis. For example, data from different platforms cannot be compared together, but WGCNA methods avoid this limitation by focusing on a group of genes rather than on individual genes. Besides, the cutoff criterion is not necessary for WGCNA, and hence, important information that is omitted when using other methods can be retrieved. Until now, WGCNA has been applied in many cancers such as pancreatic cancer,²⁶

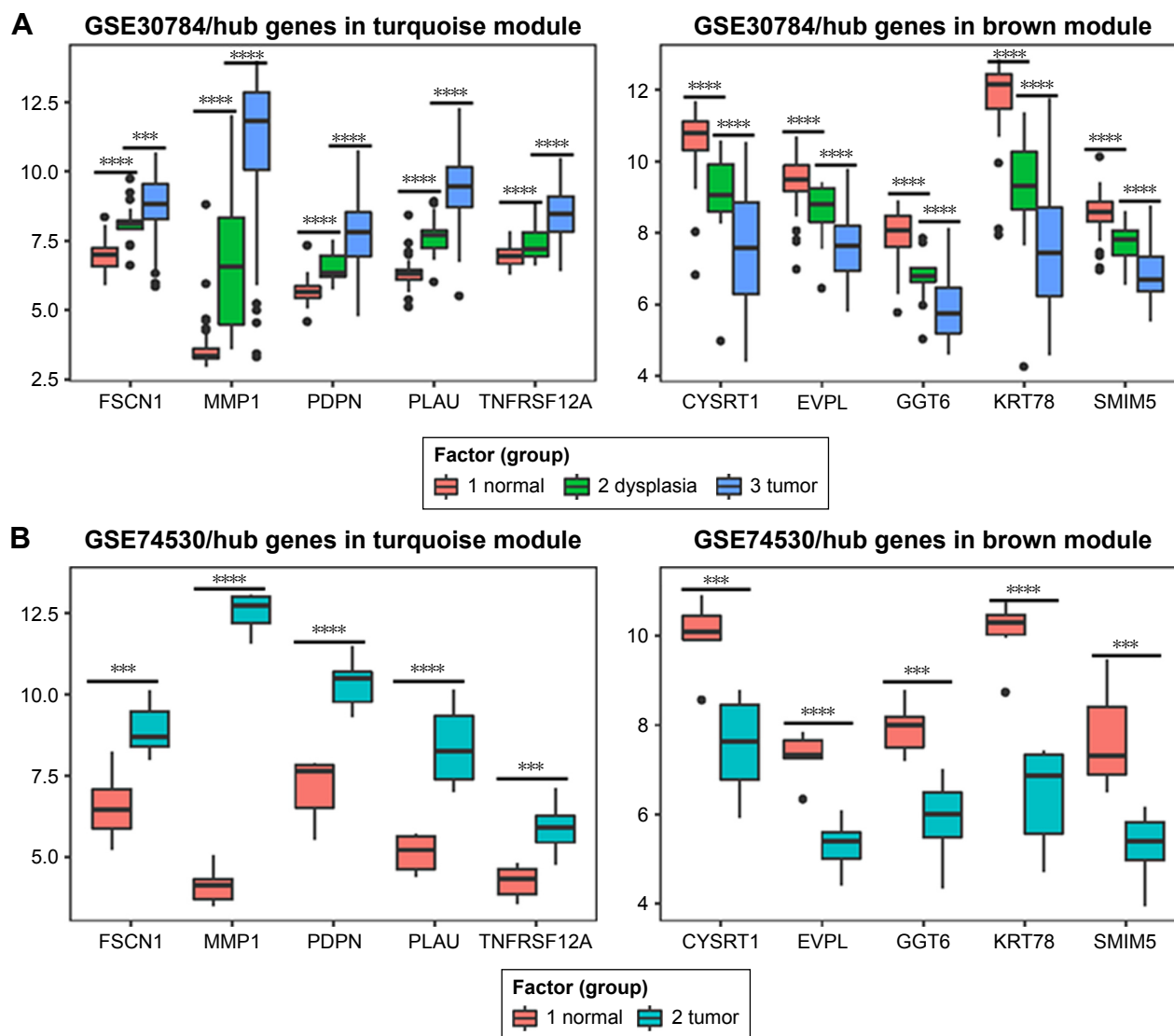


Figure 7 (Continued)

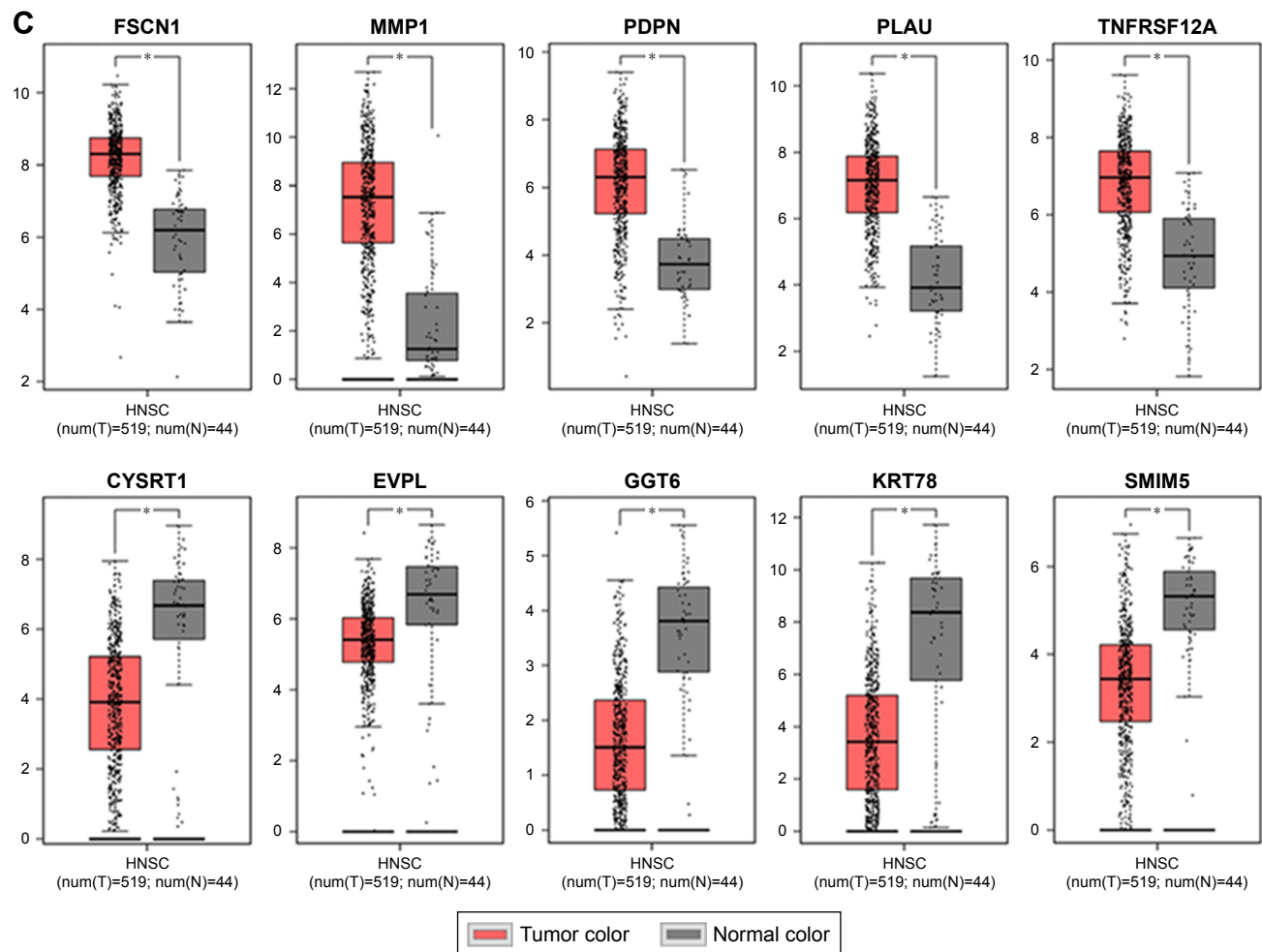


Figure 7 Validation of hub genes in the transcriptional level.

Notes: (A) Validation of hub genes in GSE30784. The expression status of five hub genes in the turquoise module was positively correlated with the disease status, indicating they played an important role in promoting tumorigenesis. This was in accordance with the results of WGCNA. The expression status of five hub genes in the brown module was negatively correlated with the disease status, indicating they played an important role in suppressing tumorigenesis. This was in accordance with the results of WGCNA. (B) Validation of hub genes in GSE74530 and the results were the same as earlier. (C) Validation of hub genes by TCGA HNSC data in GEPIA and we obtained similar results as above. These findings indicated that our results were reliable. *** indicates <0.001 , **** indicates <0.0001 .

Abbreviations: WGCNA, weighted gene co-expression network analysis; HNSC, head and neck squamous cell carcinoma.

breast cancer²⁷ and osteosarcoma.²⁸ Although there were some reports on OSCC based on expression profiles,²⁹ their disadvantages were obvious. There were no studies in which the expression profile of OSCC was analyzed by WGCNA. In this study, we used WGCNA to analyze OSCC from a new perspective.

From our research, the turquoise module and brown module identified by WGCNA were the most significantly related modules to OSCC disease status, indicating their importance in OSCC invasion, metastasis, migration and tumorigenesis. It has been reported that the epithelial–mesenchymal transition and the loss of cell adhesion are significant gene–molecular changes in the development and progression of OSCC.³⁰ We found that the turquoise module played a role

in the BPs and pathways, mainly in the regulation of ECM and collagen catabolic process, which play important roles in malignancy.³¹ The results were in accordance with previous reports³² that analyzed the DEGs between normal samples and OSCC in which the function of ECM was stressed and the MMP family and laminin family play an important role in tumorigenesis of OSCC.³³ These findings revealed the function of the turquoise module in OSCC, and further research is required to identify the role of the genes in the turquoise module because they might have diverse functions in OSCC malignant phenotypes. The KEGG pathway analysis revealed similar results. These pathways were concerned with cell adhesion such as focal adhesion, PI3K–Akt signaling pathway and ECM–receptor interaction. ECM proteolysis

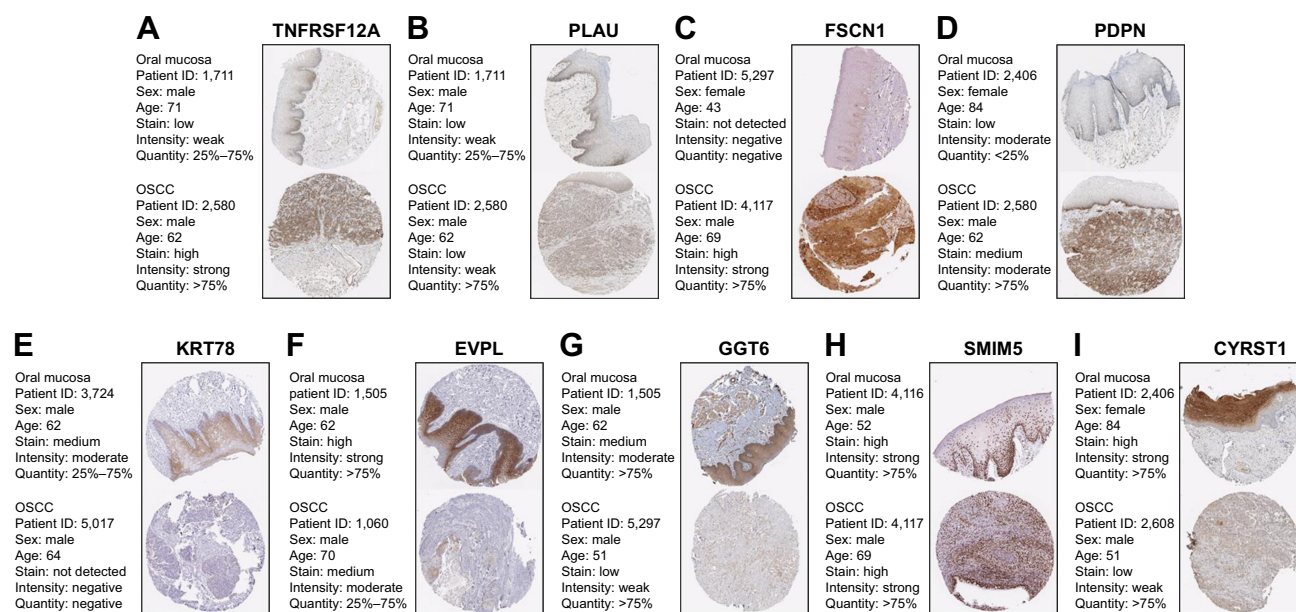


Figure 8 Validation of hub genes in the translational level.

Notes: (A–D) Validation of five hub genes in turquoise module by The Human Protein Atlas database (IHC). There was no related IHC samples of *MMP1* in the database. The translational expression level of the rest four hub genes was positively correlated with disease status as they were upregulated in OSCC samples. (E–I) Validation of five hub genes in the brown module by The Human Protein Atlas database (IHC). The translational expression level of five hub genes was negatively correlated with disease status, as they were downregulated in OSCC samples.

Abbreviations: IHC, immunohistochemistry; OSCC, oral squamous cell carcinoma.

confers cancer cells with invasive capability, and it influences migration in many types of cancers.³⁴ It has been reported that *COL11A1* promoted tumor progression in EOC via ECM–receptor interactions.³⁵ FAK is an important mediator of proliferation, cell survival and cell migration; the development of malignancy is often associated with abnormal FAK activity.³⁶ It has been reported in OSCC that focal adhesion is related to cell proliferation.³⁷ PI3K–Akt pathway plays an important role in cancer cell cycle progression, apoptosis and neoplastic transformation³⁸ and is involved in many of the mechanisms targeted by newer antitumor drugs.

As for the brown module, GO analysis revealed that it is involved in the regulation of epidermis development and epithelial cell differentiation, which oppose OSCC EMT. The brown module decreased in those BP that promotes the OSCC EMT, resulting in invasion and metastasis of tumor cells.³⁹ This illustrated an important role of the brown module as an antitumor.

Owing to the importance of turquoise and brown module in OSCC tumorigenesis, we screened the hub genes in the two modules. We found *MMP1*, *TNFRSF12A*, *PLAU*, *FSCN1* and *PDPN* in the turquoise module and *KRT78*, *EVPL*, *GGT6*, *SMIM5* and *CYRST1* in the brown module. They were significantly related to survival analysis results and were validated by other data at transcriptional and translational levels. Many studies have reported that some of the

10 hub genes were cancer-related genes, which function in tumorigenesis and malignant phenotypes such as *MMP1*, *TNFRSF12A* and *FSCN*. As for *PLAU*, *PDPN*, *KRT78*, *EVPL*, *GGT6*, *SMIM5* and *CYRST1*, there were few reports implicating them in cancers. *MMP1* encodes for a member of the peptidase M10 family of MMPs. It is involved in the breakdown of ECM in normal physiological processes, such as embryonic development, reproduction and tissue remodeling, as well as in disease processes, such as arthritis and metastasis. It was shown that *MMP1* associated with Ras, thereby initiating malignant tumor formation and leading to aberrant regulation of cell survival, proliferation and mobilization.⁴⁰ *TNFRSF12A* encodes receptor for TNFSF12/TWEAK. It can promote angiogenesis and modulate cellular adhesion to matrix proteins. Knockdown of *TNFRSF12A* would lead to inhibition of cell proliferation and migration in hepatocellular carcinoma,⁴¹ but no report has shown that it is linked to OSCC. *PLAU* encodes for a secreted serine protease that converts plasminogen into plasmin. Mutations in this gene may be associated with Quebec platelet disorder and late-onset Alzheimer's disease. Studies on *PLAU* were mainly focused on Alzheimer's disease, but several reports have shown that *PLAU* played an important role in breast cancer cell line aggression and metastasis.⁴² *PLAU* mutation occurs more or less in many cancer types, but there was no study that reported the effects of *PLAU* mutation in cancers.

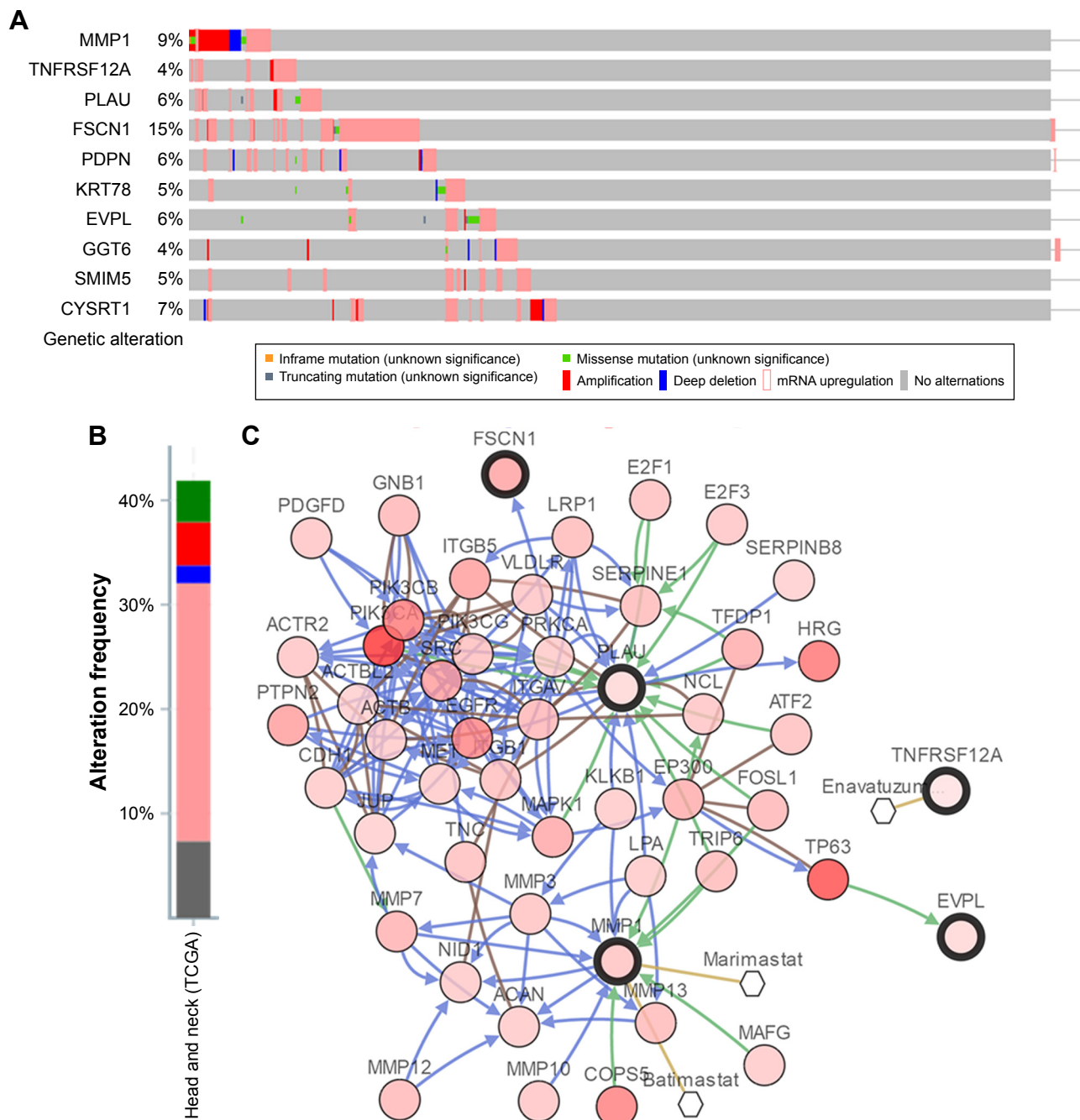


Figure 9 Genetic alterations associated with hub genes in TCGA HNSC.

Notes: (A) A visual summary across on a query of 10 hub genes showing genetic alteration of 10 hub genes in TCGA HNSC patients. (B) The total alteration frequency of 10 hub genes in TCGA HNSC is illustrated. (C) The network contains 55 nodes, including our 10 query genes and the 50 most frequently altered neighbor genes (only four out of 10 were correlated with the 50 genes). Relationship between hub genes and tumor drugs is also illustrated.

Abbreviation: HNSC, head and neck squamous cell carcinoma.

However, in Quebec platelet disorder, tandem duplication of a 78 kb region of chromosome 10 containing *PLAU* (the uPA gene) would lead to increasing production of normal *PLAU* transcripts.⁴³ *FSCN1* encodes for a member of the fascin family of actin-binding proteins, which plays a critical role in cell migration, motility, adhesion and cellular interactions. It has been reported that overexpression of this gene may

play a role in the metastasis of multiple types of cancer by increasing cell motility. For instance, in esophageal cancer, tumor-suppressive miRNAs target *FSCN1* and directly control oncogenic *FSCN1* gene.⁴⁴ *FSCN1* gene alteration was rarely reported, but there was a report showing that alteration of *FSCN1* increased when using demethylating agent, resulting in drug resistance of gastric cancer.⁴⁵ However, there are

no reports implicating it in OSCC at present. *PDPN* encodes for type I integral membrane glycoprotein, which has diverse distribution in human tissues. The specific function of this protein has not been determined, but it has been proposed to be a marker of lung injury. The role of *PDPN* in cancers has not been studied at present. Alteration of *PDPN* was reported to be associating with keratocystic odontogenic tumors' neoplastic invasion.⁴⁶ *EVPL* encodes for a member of the plakin family of proteins that forms a component of desmosomes and the epidermal cornified envelope. It has been reported that *EVPL* was underexpressed in esophageal squamous cell carcinoma compared to normal tissue, and hence, it may be used for early detection of ESCC.⁴⁷ The *EVPL* gene mutation was closely linked to the tylosis esophageal cancer locus in sporadic OSC and could be the target gene responsible for OSC.⁴⁸ OSCC was similar to OSC, which means the *EVPL* mutation also plays an important role in OSCC. As for *KRT78*, *GGT6*, *SMIM5* and *CYSRT1*, they are relatively new molecules with only few reports regarding their role in OSCC at present. Nevertheless, they played an important role in OSCC tumorigenesis and were significantly different between normal samples and OSCC. Further research is required to fully explore their roles in OSCC.

We acknowledge there were some limitations and shortcomings in this study. First, this study was mainly focused on the data mining and data analysis, which were based on methodology and the results were not validated by experiments. Further experiments are needed to better confirm the findings of this study. Second, the clinical parameters and prognosis were not well analyzed in this study because of the availability of data. Last, the datasets we could obtain were limited. We could only find one dataset that contained normal samples, dysplasia samples and OSCC samples. If there would have been another dataset that contained these three or more statuses of OSCC, it would be much better. Besides, we only validated our findings in one GEO dataset and TCGA data; more datasets should be involved to obtain a solid result.

Conclusion

We applied WGCNA to construct co-expression network for exploring the development of OSCC (from normal to dysplasia and normal to OSCC) for the first time and found two modules (turquoise module and brown module) and 10 hub genes (*MMP1*, *TNFRSF12A*, *PLAU*, *FSCN1*, *PDPN*, *KRT78*, *EVPL*, *GGT6*, *SMIM5* and *CYSRT1*), which played an important role in OSCC tumorigenesis. The turquoise module plays a positive role in OSCC tumorigenesis, which

includes functions of cell migration, cell adhesion and ECM, and the brown module plays a negative role in tumorigenesis, which functions opposite to the turquoise module. The two modules may provide a better understanding of the mechanisms of tumorigenesis in OSCC patients. Moreover, the 10 hub genes we newly found might serve as prognostic biomarkers and therapeutic targets in the future.

Acknowledgment

This study was partially supported by the National Nature Science Foundation of China (Grant Nos 81470719 and 81611140133) to Li M.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Mignogna MD, Fedele S, Lo Russo L. The World Cancer Report and the burden of oral cancer. *Eur J Cancer Prev*. 2004;13(2):139–142.
2. Hosokawa S, Takahashi G, Okamura J, et al. Risk and prognostic factors for multiple primary carcinomas in patients with head and neck cancer. *Jpn J Clin Oncol*. 2018;48(2):124–129.
3. Jacobs CD. Etiologic considerations for head and neck squamous cancers. *Cancer Treat Res*. 1990;52:265–282.
4. Falk RT, Pickle LW, Brown LM, Mason TJ, Buffler PA, Fraumeni JF. Effect of smoking and alcohol consumption on laryngeal cancer risk in coastal Texas. *Cancer Res*. 1989;49(14):4024–4029.
5. Choi S, Myers JN. Molecular pathogenesis of oral squamous cell carcinoma: implications for therapy. *J Dent Res*. 2008;87(1):14–32.
6. Marioni G, Staffieri A, Lionello M, et al. Relationship between anti-apoptotic proteins survivin and Bcl-2, and response to treatment in patients undergoing post-operative RT for laryngeal cancer: a pilot study. *J Oral Pathol Med*. 2013;42(4):339–344.
7. Fu S, Guo Y, Chen H, et al. MYCT1-TV, a novel MYCT1 transcript, is regulated by c-Myc and may participate in laryngeal carcinogenesis. *PLoS One*. 2011;6(10):e25648.
8. Jalali MM, Heidarzadeh A, Zavarei MJ, Sarmast H. p53 overexpression impacts on the prognosis of laryngeal squamous cell carcinomas. *Asian Pac J Cancer Prev*. 2011;12(7):1731–1734.
9. Boonyaphiphat P, Pruegsanusak K, Thongsuksai P. The prognostic value of p53, Bcl-2 and Bax expression in laryngeal cancer. *J Med Assoc Thai*. 2012;95(10):1317–1320.
10. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;4:Article17.
11. Ivliev AE, 't Hoen PA, Sergeeva MG. Coexpression network analysis identifies transcriptional modules related to proastrocytic differentiation and sprouty signaling in glioma. *Cancer Res*. 2010;70(24):10060–10070.
12. Clarke C, Madden SF, Doolan P, et al. Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis. *Carcinogenesis*. 2013;34(10):2300–2308.
13. Presson AP, Yoon NK, Bagryanova L, et al. Protein expression based multimarker analysis of breast cancer samples. *BMC Cancer*. 2011;11:230.
14. Xiang Y, Zhang CQ, Huang K. Predicting glioblastoma prognosis networks using weighted gene co-expression network analysis on TCGA data. *BMC Bioinformatics*. 2012;13(Suppl 2):S12.
15. Wang L, Tang H, Thayanithy V, et al. Gene networks and microRNAs implicated in aggressive prostate cancer. *Cancer Res*. 2009;69(24):9490–9497.

16. Huang Daw, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009;37(1):1–13.
17. Huang Daw, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1):44–57.
18. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 2017;45(W1):W98–W102.
19. Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics.* 2014;15:293.
20. Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347(6220):1260419.
21. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* 2013;6(269):pl1.
22. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012;2(5):401–404.
23. Scott SE, Grunfeld EA, McGurk M. The idiosyncratic relationship between diagnostic delay and stage of oral squamous cell carcinoma. *Oral Oncol.* 2005;41(4):396–403.
24. Smith RA, Cokkinides V, Eyre HJ. American Cancer Society guidelines for the early detection of cancer, 2006. *CA Cancer J Clin.* 2006; 56(1):1149–2550 quiz.
25. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
26. Zhou Z, Cheng Y, Jiang Y, et al. Ten hub genes associated with progression and prognosis of pancreatic carcinoma identified by co-expression analysis. *Int J Biol Sci.* 2018;14(2):124–136.
27. Zhao Q, Song W, He DY, Li Y. Identification of key gene modules and pathways of human breast cancer by co-expression analysis. *Breast Cancer.* 2018;25(2):213–223.
28. Liu X, Hu AX, Zhao JL, Chen FL. Identification of Key Gene Modules in Human Osteosarcoma by Co-Expression Analysis Weighted Gene Co-Expression Network Analysis (WGCNA). *J Cell Biochem.* 2017; 118(11):3953–3959.
29. Fountzilas E, Kotoula V, Angouridakis N, et al. Identification and validation of a multigene predictor of recurrence in primary laryngeal cancer. *PLoS One.* 2013;8(8):e70429.
30. Chung CH, Parker JS, Ely K, et al. Gene expression profiles identify epithelial-to-mesenchymal transition and activation of nuclear factor-kappaB signaling as characteristics of a high-risk head and neck squamous cell carcinoma. *Cancer Res.* 2006;66(16):8210–8218.
31. Wu JS, Sheng SR, Liang XH, Tang YL. The role of tumor microenvironment in collective tumor cell invasion. *Future Oncol.* 2017;13(11): 991–1002.
32. Yang B, Chen Z, Huang Y, Han G, Li W. Identification of potential biomarkers and analysis of prognostic values in head and neck squamous cell carcinoma by bioinformatics analysis. *Onco Targets Ther.* 2017; 10:2315–2321.
33. Tanis T, Cincin ZB, Gokcen-Rohlig B, et al. The role of components of the extracellular matrix and inflammation on oral squamous cell carcinoma metastasis. *Arch Oral Biol.* 2014;59(11):1155–1163.
34. Wolf K, Friedl P. Mapping proteolytic cancer cell-extracellular matrix interfaces. *Clin Exp Metastasis.* 2009;26(4):289–298.
35. Wu YH, Chang TH, Huang YF, Huang HD, Chou CY. COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer. *Oncogene.* 2014;33(26):3432–3440.
36. Mclean GW, Carragher NO, Avizienyte E, Evans J, Brunton VG, Frame MC. The role of focal-adhesion kinase in cancer – a new therapeutic opportunity. *Nat Rev Cancer.* 2005;5(7):505–515.
37. Abu-Ali S, Fotovati A, Shirasuna K. Tyrosine-kinase inhibition results in EGFR clustering at focal adhesions and consequent exocytosis in uPAR down-regulated cells of head and neck cancers. *Mol Cancer.* 2008;7:47.
38. Chang F, Lee JT, Navolanic PM, et al. Involvement of PI3K/Akt pathway in cell cycle progression, apoptosis, and neoplastic transformation: a target for cancer chemotherapy. *Leukemia.* 2003;17(3):590–603.
39. Lumerman H, Freedman P, Kerpel S. Oral epithelial dysplasia and the development of invasive squamous cell carcinoma. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 1995;79(3):321–329.
40. Uhlírova M, Bohmann D. JNK- and Fos-regulated Mmp1 expression cooperates with Ras to induce invasive tumors in Drosophila. *Embo J.* 2006;25(22):5294–5304.
41. Wang T, Ma S, Qi X, et al. Knockdown of the differentially expressed gene TNFRSF12A inhibits hepatocellular carcinoma cell proliferation and migration in vitro. *Mol Med Rep.* 2017;15(3):1172–1178.
42. Moquet-Torcy G, Tolza C, Piechaczyk M, Jariel-Encontre I. Transcriptional complexity and roles of Fra-1/AP-1 at the uPA/Plau locus in aggressive breast cancer. *Nucleic Acids Res.* 2014;42(17):11011–11024.
43. Hayward CP, Liang M, Tasneem S, et al. The duplication mutation of Quebec platelet disorder dysregulates PLAUI, but not C10orf55, selectively increasing production of normal PLAUI transcripts by megakaryocytes but not granulocytes. *PLoS One.* 2017;12(3):e0173991.
44. Kano M, Seki N, Kikkawa N, et al. miR-145, miR-133a and miR-133b: Tumor-suppressive miRNAs target FSCN1 in esophageal squamous cell carcinoma. *Int J Cancer.* 2010;127(12):2804–2814.
45. Maeda O, Ando T, Ohmiya N, et al. Alteration of gene expression and DNA methylation in drug-resistant gastric cancer. *Oncol Rep.* 2014; 31(4):1883–1890.
46. Zhang X, Wang J, Ding X, et al. Altered expression of podoplanin in keratocystic odontogenic tumours following decompression. *Oncol Lett.* 2014;7(3):627–630.
47. Hu N, Qian L, Hu Y, et al. Quantitative real-time RT-PCR validation of differential mRNA expression of SPARC, FADD, Fascin, COL7A1, CK4, TGM3, ECM1, PPL and EVPL in esophageal squamous cell carcinoma. *BMC Cancer.* 2006;6(1):33.
48. Iwaya T, Maesawa C, Kimura T, et al. Infrequent mutation of the human envoplakin gene is closely linked to the tylosis oesophageal cancer locus in sporadic oesophageal squamous cell carcinomas. *Oncol Rep.* 2005;13(4):703–707.

Supplementary materials

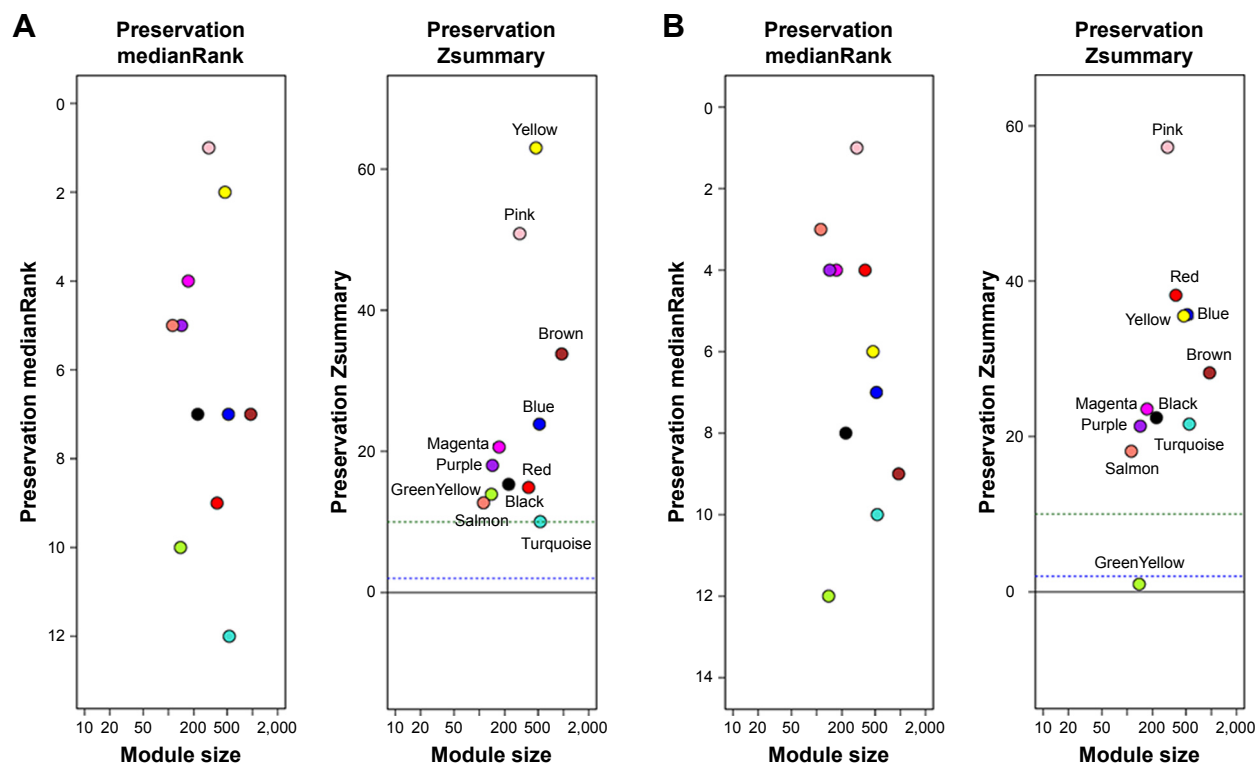


Figure S1 The medianRank and Zsummary statistics of module preservation.

Notes: Module preservation was evaluated by medianRank and Zsummary statistics, which correlated to connectivity and density of networks. If Zsummary >10, there is a strong evidence that the module is preserved. The module with a lower medianRank tends to exhibit stronger observed preservation than the module with a higher medianRank if both of them are preserved. **(A)** Module preservation analysis for the process of normal to dysplasia. **(B)** Module preservation analysis for the process of dysplasia to neoplasm. The pink module was found to be highly preserved in both processes of normal to dysplasia and dysplasia to OSCC.

Abbreviation: OSCC, oral squamous cell carcinoma.

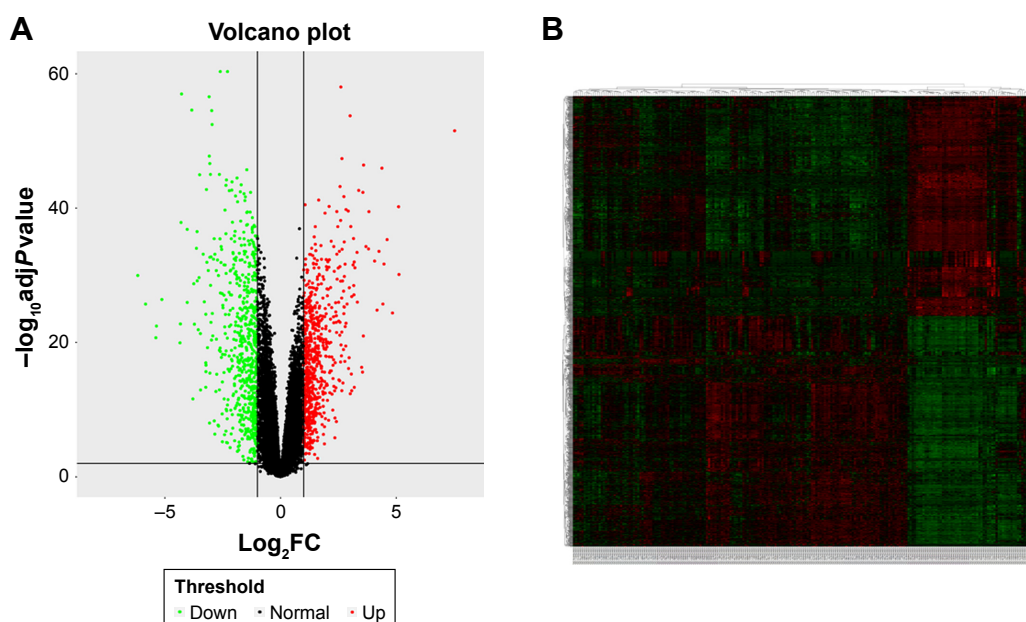


Figure S2 **(A)** Volcano plot visualizing DEGs in GSE30784 (45 normal samples and 167 OSCC). The vertical lines demark the fold change values. The right vertical line corresponds to more than or equal to twofold up changes, and the left vertical line corresponds to twofold down changes, while the horizontal line marks a $-\log_{10}$ adjusted P-value of 0.01. **(B)** Heat map hierarchical clustering reveals DEGs in OSCC groups compared with control groups.

Abbreviations: DEG, differently expressed genes; OSCC, oral squamous cell carcinoma.

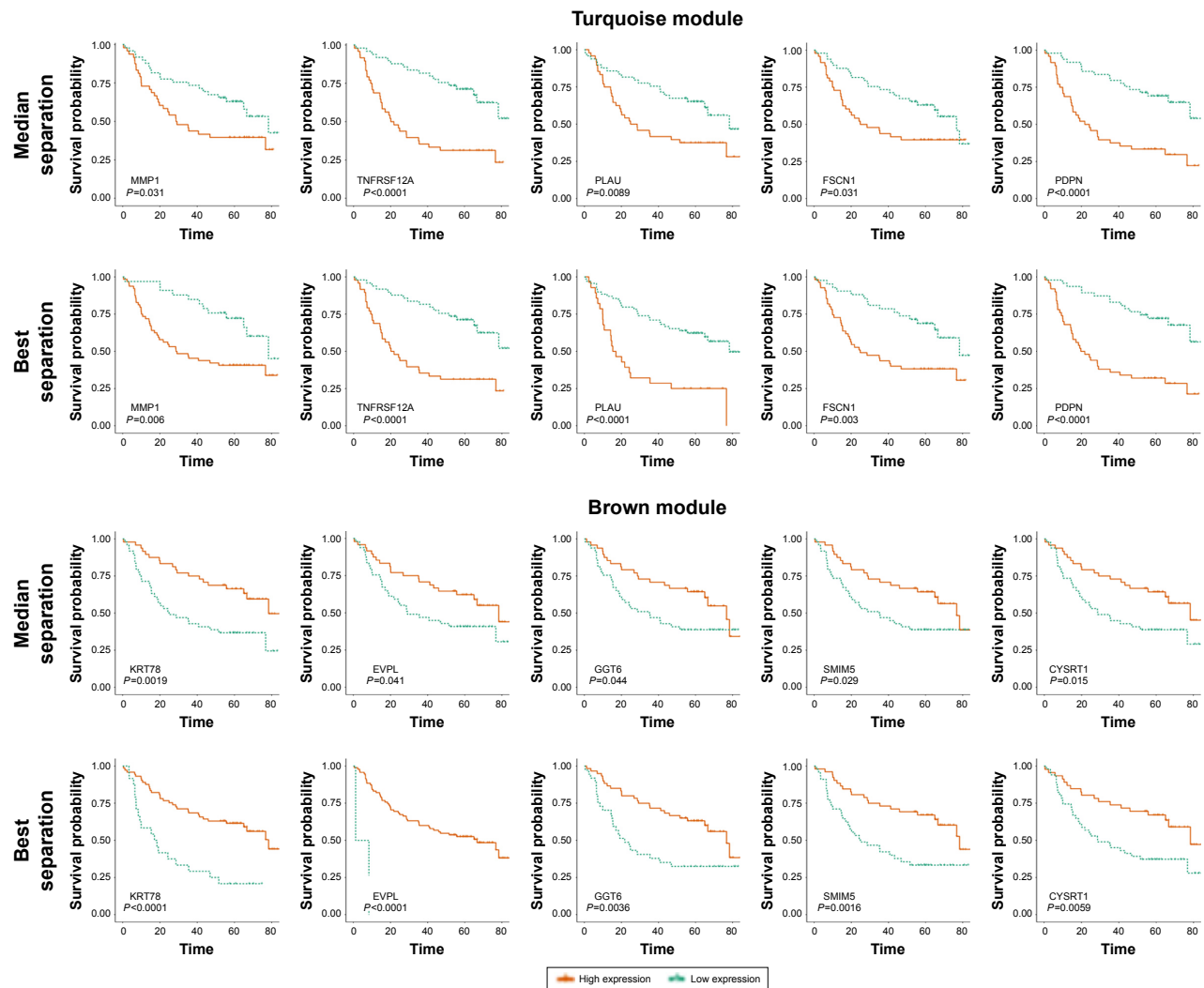


Figure S3 Survival analysis of the 10 hub genes using GSE41613 for validation.

Table S1 GO enrichment analysis of DEGs

Category	Term	Count	%	P-value
Upregulated genes				
BP	GO:0030198~ECM organization	60	8.39	6.57E-36
BP	GO:0007155~cell adhesion	78	10.91	9.15E-28
BP	GO:0030574~collagen catabolic process	32	4.48	8.50E-27
BP	GO:0006955~immune response	63	8.81	1.59E-19
BP	GO:0006954~inflammatory response	59	8.25	3.88E-19
BP	GO:0022617~ECM disassembly	25	3.50	8.43E-16
BP	GO:0035987~endodermal cell differentiation	16	2.24	8.80E-15
BP	GO:0009615~response to virus	28	3.92	1.36E-14
BP	GO:0060337~type I interferon signaling pathway	22	3.08	2.17E-14
BP	GO:0030199~collagen fibril organization	18	2.52	2.68E-14
CC	GO:0005576~extracellular region	175	24.48	4.88E-39
CC	GO:0031012~extracellular matrix	63	8.81	6.30E-29
CC	GO:0005615~extracellular space	140	19.58	1.26E-28
CC	GO:0005578~proteinaceous ECM	58	8.11	5.00E-27
CC	GO:0005581~collagen trimer	30	4.20	2.27E-19
MF	GO:0005201~ECM structural constituent	23	3.22	3.95E-15

(Continued)

Table S1 (Continued)

Category	Term	Count	%	P-value
MF	GO:0005518~collagen binding	21	2.94	5.06E-14
MF	GO:0008009~chemokine activity	18	2.52	1.92E-12
MF	GO:0005178~integrin binding	23	3.22	8.68E-11
MF	GO:0008201~heparin binding	28	3.92	1.25E-10
Downregulated genes				
BP	GO:0030049~muscle filament sliding	12	1.76	3.40E-08
BP	GO:0006805~xenobiotic metabolic process	15	2.21	3.32E-07
BP	GO:0008544~epidermis development	15	2.21	9.90E-07
BP	GO:0006936~muscle contraction	16	2.35	3.36E-06
BP	GO:0055114~oxidation-reduction process	42	6.18	1.37E-05
BP	GO:0030855~epithelial cell differentiation	12	1.76	2.27E-05
BP	GO:0060048~cardiac muscle contraction	9	1.32	1.22E-04
BP	GO:0016338~calcium-independent cell-cell adhesion via plasma membrane cell-adhesion molecules	6	0.88	5.83E-04
BP	GO:0006941~striated muscle contraction	5	0.74	7.45E-04
BP	GO:0016266~O-glycan processing	9	1.32	9.28E-04
CC	GO:0070062~extracellular exosome	186	27.35	3.12E-20
CC	GO:0005615~extracellular space	89	13.09	2.98E-09
CC	GO:0016324~apical plasma membrane	29	4.26	9.39E-07
CC	GO:0005859~muscle myosin complex	7	1.03	9.48E-06
CC	GO:0030018~Z disc	16	2.35	1.25E-05
MF	GO:0003779~actin binding	32	4.71	8.33E-09
MF	GO:0008307~structural constituent of muscle	13	1.91	1.04E-08
MF	GO:0005198~structural molecule activity	24	3.53	1.51E-05
MF	GO:0016491~oxidoreductase activity	20	2.94	6.25E-05
MF	GO:0051371~muscle alpha-actinin binding	5	0.74	1.51E-04

Abbreviations: GO, Gene Ontology; DEG, differently expressed genes; BP, biological process; ECM, extracellular matrix; CC, cellular component; MF, molecular function.

Table S2 KEGG pathway enrichment analysis of turquoise module and brown module

Category	Term	Count	%	P-value
Upregulated genes				
KEGG	hsa04512: ECM-receptor interaction	33	4.62	6.07E-21
KEGG	hsa04510: focal adhesion	46	6.43	9.77E-19
KEGG	hsa04151: PI3K-Akt signaling pathway	50	6.99	2.07E-12
KEGG	hsa05323: rheumatoid arthritis	24	3.36	9.34E-12
KEGG	hsa05146: amoebiasis	26	3.64	1.35E-11
KEGG	hsa04974: protein digestion and absorption	23	3.22	6.97E-11
KEGG	hsa04060: cytokine-cytokine receptor interaction	35	4.90	2.31E-09
KEGG	hsa05200: pathways in cancer	45	6.29	6.30E-08
KEGG	hsa05144: malaria	13	1.82	2.08E-06
KEGG	hsa04062: chemokine signaling pathway	26	3.64	2.09E-06
Downregulated genes				
KEGG	hsa01100: metabolic pathways	68	10.00	1.80E-06
KEGG	hsa04530: tight junction	17	2.50	5.91E-06
KEGG	hsa04972: pancreatic secretion	13	1.91	3.28E-05
KEGG	hsa00982: drug metabolism – cytochrome P450	11	1.62	4.84E-05
KEGG	hsa00340: histidine metabolism	6	0.88	6.44E-04
KEGG	hsa00590: arachidonic acid metabolism	9	1.32	6.93E-04
KEGG	hsa00350: tyrosine metabolism	7	1.03	7.11E-04
KEGG	hsa05204: chemical carcinogenesis	10	1.47	8.98E-04
KEGG	hsa00980: metabolism of xenobiotics by cytochrome P450	9	1.32	0.002231
KEGG	hsa04670: leukocyte transendothelial migration	11	1.62	0.004088

Abbreviations: KEGG, Kyoto Encyclopedia of Genes and Genomes; ECM, extracellular matrix.

OncoTargets and Therapy**Dovepress****Publish your work in this journal**

OncoTargets and Therapy is an international, peer-reviewed, open access journal focusing on the pathological basis of all cancers, potential targets for therapy and treatment protocols employed to improve the management of cancer patients. The journal also focuses on the impact of management programs and new therapeutic agents and protocols on

patient perspectives such as quality of life, adherence and satisfaction. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/oncotargets-and-therapy-journal>