A key genomic signature associated with lymphovascular invasion in head and neck squamous cell carcinoma

Jian Zhang
Affiliated Cancer hospital & Institute of Guangzhou Medical University
Huali Jiang
Affiliated Donghua Hospital of Sun Yat-sen University
Tao Xie
Affiliated Cancer Hospital of Guangzhou Medical University
Baiyao Wang
Affiliated Cancer Hospital of Guangzhou Medical Unversity
Xiaoting Huang
Affiliated Cancer Hospital & Institute of Guangzhou Medical University
Jie Lin
Affiliated Cancer Hospital & Institute of Guangzhou Medical University
Anan Xu
Affiliated Cancer Hospital of Guangzhou Medical University
Rong Li
Affiliated Cancer Hospital & Institute of Guangzhou Medical University
Yawei Yuan (yuanyawei@gzhmu.edu.cn)
Guangzhou Medical University Affiliated Cancer Hospital

Research article

Keywords: lymphovascular invasion, head and neck squamous cell carcinoma, hub genes, TCGA, weighted gene co-expression network analysis

Posted Date: January 16th, 2020

DOI: https://doi.org/10.21203/rs.2.18349/v2

Abstract

Objective: Lymphovascular invasion (LOI), a key pathological feature of head and neck squamous cell carcinoma (HNSCC), predicts poor survival. However, the associated clinical characteristics remain uncertain, and the molecular mechanisms are largely unknown.

Methods: Weighted gene co-expression network analysis was performed to construct gene co-expression networks and investigate the relationship between modules and LOI clinical trait. Functional enrichment and KEGG pathway enrichment analysis were performed for differentially expressed genes using DAVID database. The protein-protein interaction network was constructed using Cytoscape software, and module analysis was performed using MCODE. Prognosis role and expression analysis was further validated by survival analysis, GEPIA analysis and HPA database. Multivariable Cox regression analysis was used to establish a prognostic risk formula and the areas under the receiver operating characteristic curve (AUCs) were used to evaluate prediction efficiency. And the potential targeted LOI molecular agents were identified with DrugBank.

Results: 10 co-expression modules in two key modules (turquoise and pink) associated with tumor LOI were identified. Functional enrichment and KEGG analysis identified turquoise and pink modules played significant roles in the progression of HNSCC. The seven hub genes (CNFN, KIF18B, KIF23, PRC1, CCNA2, DEPDC1 and TTK) in two modules were identified and validated by survival analysis and expression analysis. Multivariable Cox regression analysis was used to establish a prognostic risk formula (risk score = EXP DEPDC1 * 0.32636 + EXP CNFN * (-0.07544). The low-risk group had a better OS than the high-risk group ($P < 0.001$), and the areas under the receiver operating characteristic curve (AUCs) of 1-, 3- and 5-year OS were 0.582, 0.634 and 0.636, respectively. Eight small molecular agents, including XL844, AT7519, AT9283, Alvocidib, Nelarabine, Benzamidine, L-Glutamine, and Zinc, may be a candidate drug for treating LOI ($P < 0.05$).

Conclusions: Our research revealed the two-mRNA signature could serve as an independent biomarker to predict LOI risk, which provide some new insights into LOI of HNSCC. Additionally, the small molecular agents may be a candidate drug for treating LOI.

Background

Head and neck squamous cell carcinoma (HNSCC), one of the most common pathological subtypes, nearly reach 90% of head and neck cancer [1]. Metastasis is the main cause of treatment failure and a important factor that affects the prognosis of HNSCC [2]. Thus, understanding the genomic changes of metastasis may be a valuable way to reduce the metastasis of lymph nodes.

In HNSCC, advanced TNM stage, histological grade and lymph node status, which indicate several high risk factors of metastatic disease, poor overall survival and disease free survival, are poor prognosis indicators [3-5]. Lymphovascular invasion (LOI) has been found to be associated with lymph node

metastasis of HNSCC [6-8]. Thus, knowing effective molecular prognosticators of LOI can be a useful way to decrease the metastasis risk in HNSCC.

According to the recent studies, the clinical characteristics and parameters about LOI are still uncertain. Such as, the incidence of LOI in HNSCC was not consistent, which varied from 14 to 47% [9, 10]. This huge difference could be due to the small sample sizes, distribution difference and the heterogeneity of HNSCC. Meanwhile, it's urgent to analyze genomic and clinical characteristics of LOI based on large sample sizes clinical studies. Thus, it's critically systematic to elucidating the genomic changes and mechanism of LOI for facilitating the development of new therapeutic targets and can enhance HNSCC survival with LOI.

The Cancer Genome Atlas (TCGA) has provided comprehensive molecular features of HNSCC and offered histopathological annotations and clinical survival information of HNSCC with 10 years follow-up time. The clinical resource enabled us to systematically evaluate the relationships between LOI and gene signatures and clarify key gene modules involved in LOI of HNSCC, which provide us a comprehensive and systemic understanding of LOI not only from the genomic level but also to the prognosis level.

Methods

Patient selection and data pre-processing

Data information of HNSCC patients was downloaded from TCGA database. As shown in Table 1, RNA expression profiles and clinical survival data of 500 patients were obtained. Among these 500 patients, clinical prognosis data of 339 patients is available. According to the difference multiples ($|logFC| > 1$) and the significance threshold ($P < 0.05$), 2248 genes that met the criteria were screened out as differentially expressed genes (DEGs) (Additional file 1: Table S1). The intersection of DEGs from the NCBI-gene database (Additional file 2: Table S2) and OMIM database (Additional file 3: Table S3) was performed using the Venn Diagram package in R language. The study was approved by the Human Ethics Approval Committee of Affiliated Cancer Hospital & Institute of Guangzhou Medical University (2019-290).

Construction of co-expression module network

As the method used by previous studies [11, 12], based on mRNA expression, the scale-free gene modules of co-expression were constructed by WGCNA. To ensure the reliability of the co-expression network, hierarchical clustering was performed based on Euclidean distance and 2 outlier samples were removed. The module–trait associations were considered as important clinical characteristics between the clinical phenotype and module eigengenes (MEs). We analyzed the module–trait correlation and clarified the relevant modules, which were closely associated with the LOI clinical traits. An adequate soft-threshold power that met the scale-free topology criterion was selected for transforming the former correlation matrix into an adjacency matrix, which was subsequently converted into a Topological Overlap Matrix

(TOM) using the 'TOM similarity' function in R. TOM-based dissimilarity was computed as measure distance, and a mRNA clustering dendrogram and module colors were obtained. In the clustering dendrogram, the minimum module size and cut height were separately set to 30 and 0.25, respectively. For the key gene modules, gene significance (GS) and module membership (MM) mean a positive correlation level between RNA expression profiles and LOI clinical phenotype, and between RNA expression profiles and clinical MEs.

Enrichment of key co-expression modules analysis

As the method used by previous studies [12, 13], the aberrantly expressed mRNAs in the key gene modules were selected to perform GO function analysis and KEGG pathway analysis. For GO analysis, the corresponding genes were divided into by biological process (BPs) analysis. Using the KEGG analysis, genes of the key co-expression modules were used to detect the gene modules functions, and $P < 0.05$ was considered as statistically significant.

PPI network analysis and identification of hub genes

As method used by previous studies [14, 15] , the key gene co-expression module was further explored to predict gene function correlation using STRING database with a confidence score >0.9. Cytoscape was employed to screen significant gene pairs in the PPI network [16]. We further screen the modules of the PPI network by molecular complex detection (MCODE) analysis. The criteria of MCODE were as follows: degree cutoff=2, node cutoff=0.2, maximum depth=100 and k-core=2. At last, the 24 genes were selected as hub genes and further analyzed using univariate survival analysis. Seven genes with significant prognostic differences were selected as characteristic genes by $P < 0.05$.

Survival analysis of Hub genes

According to the expression profiles of characteristic genes, Kaplan Meier analysis was further used to explore the prognostic differences, the Cox proportional hazard ratio and the 95% confidence interval were used for analysis. A $P < 0.05$ was considered statistically significant. Then the least absolute shrinkage and selection operator (LASSO) model were used to find vital mRNAs from the prognostic hub genes. The LASSO method was utilized by the package "glmnet" in the R (version 3.5.1) software.

mRNA expression analysis

To know the protein expression of seven hub genes, as the method performed by Ze et al. [17], GEPIA databse (http://gepia.cancer-pku.cn/), a web-based tool to deliver fast and customizable functionalities based on TCGA and GTEx data, was used to analyze the mRNA expression of seven hub genes.

Immunohistochemistry analysis

To validate the protein expression of seven hub genes, as the method described by Jian et al. [18], The HPA database (https://www.proteinatlas.org/) was used to perform protein expression analysis of seven hub genes in head and neck squamous cell carcinoma (n= 519) and normal tissues (n= 44) (Scar bar = 200μm). All IHC images have been manually annotated by certified pathologists.

Establishment of prognostic risk score formula

In light of the expression level of hub genes and regression coefficients, a prognostic risk formula was established by multivariable Cox regression analysis. The risk scores of each patient were calculated by the formula as mentioned above. Finally, all patients were divided into a high-risk group and a low-risk group by utilizing the median risk score as the cutoff value. Next, Kaplan–Meier survival curve was used to evaluate the prognosis between the low-risk group and high-risk group. And a time-dependent receiver operating characteristic (ROC) curve was performed to assess the diagnostic accuracy based on the risk score for 1-, 3- and 5-year OS probability. $P < 0.05$ was recognized as statistically significant.

Identification of small molecular drugs

DrugBank is a comprehensive and systematic resource to explore detailed drug-target interaction information [19]. The turquoise and pink modules in PPI network were mapped onto the DrugBank database. The |connectivity score| > 2 was used as cutoff value to identify molecular drugs targeted with HNSCC.

Statistical Analysis

Univariate analysis was performed using SPSS 17.0 (SPSS Inc., Chicago, IL, USA). Cumulative survival time was calculated and analyzed by the Kaplan-Meier and log-rank test. The difference between two groups was tested by a chi-square test or Fisher's exact test. *P*-value< 0.05 was considered statistically significant.

Results

Weighted co-expression network construction and key module analysis

We first performed the initial quality assessment using the average linkage method. Two outlier samples were removed after clustering. The remaining 339 cancer samples and 44 control samples with clinical information of LOI were used in subsequent analysis. The 2,601 variant genes in samples were with the largest variance via average linkage hierarchical clustering.

To estabilish a scale-free network, the scale index (Fig. 1a) and mean connectivity (Fig. 1b, c) were calculated. We found that the power value of β = 7 for the scale-free topology for which the fitting index reached 0.85 (Fig. 1d). Different genes were subsequently grouped into modules according to the association of expression. Moreover, genes with similar expression patterns could be placed into different modules via average linkage clustering. Finally, a total of 10 modules were identified (Fig. 2). The correlations between MEs and LOI trait were explored. As shown in Fig. 3a, the result indicated that 10 co-expression modules were correlated with LOI phenotypes. Figure 3B showed 10 co-expression modules was associated with cancer status, especially turquoise and pink key modules. Then, scatter diagrams of GS for MM *vs*. LOI status in the turquoise and pink modules were plotted, respectively (Fig. 3c, d), which showed the genes in two modules were significantly related with LOI status. The correlation values were 0.4 (turquoise) and 0.59 (pink), *P*-values were 1.4×10^{-30} (turquoise) and 1.8×10^{-8} (pink), which indicated the turquoise and pink modules were high correlations with LOI status.

Enrichment analysis of key co-expression modules

To know the function of genes in the key co-expression modules, GO and KEGG analysis were performed. GO analysis showed that the turquoise module was associated with DNA replication, mitotic nuclear division, chromosome segregation, nuclear division and DNA–dependent DNA replication. KEGG analysis found that the turquoise module was associated with cell cycle, DNA replication, mismatch repair and p53 signaling pathway ($P < 0.05$) (Fig. 4a, b). Similarly, GO analysis indicated that the pink module was not only involved in squamous cell function, such as, epidermal cell differentiation, keratinocyte differentiation skin development, epidermis development and cornification ($P < 0.05$), but also involved in regulation of protein secretion, such as, peptidase activity, negative regulation of proteolysis, negative regulation of peptidase activity and negative regulation of endopeptidase activity ($P < 0.05$) (Fig. 4c). These results indicated that the turquoise module and pink modules played an important role in LOI of HNSCC.

PPI analysis and hub genes

To know the hub genes in the key modules, PPI analysis of STRING database were analyzed. Connect threshold was used to define the hub genes, 89 genes including top 5 genes, KIF18B, BUB1, BUB1B, KIF4A and EXO1 in turquoise module (connect threshold ≥ 0.25) and 38 genes including top 5 genes, KRT78, CNFN, SLURP1, PRSS27 and CRCT1 in pink module (connect threshold ≥ 0.10) were screened as candidate hub genes (Fig. 5 and Additional file 4: Table S4 and Additional file 5: Table S5). In addition, connect degree (≥6) was further used to define the hub genes and then 24 genes (18 genes in turquoise module and 6 genes in pink module) were defined as hub genes.

Prognosis and expression analysis of hub genes

After excluding samples with no survival information/survival time less than 1 month, 339 cancer samples were used to evaluate the prognosis of 24 hub genes. The prognosis analysis showed that HNSCC with LOI had a poor clinical outcome than HNSCC without LOI ($P < 0.05$), which indicated that LOI was an important histological characteristic of HNSCC (Fig. 6a). The univariate survival analysis was further performed among hub genes by using R-package survival and the result indicated that CNFN was associated good survival (Fig. 6a), nevertheless, KIF18B, KIF23, PRC1, CCNA2, DEPDC1 and TTK were associated poor survival in LOI of HNSCC ($P < 0.05$) (Fig. 6c-h).

To determine the mRNA expression level of seven hubgenes (CNFN, KIF18B, KIF23, PRC1, CCNA2, DEPDC1 and TTK), we used the GEPIA database to validate the mRNA expression and found that CNFN was significantly downregulated in HNSCC and KIF18B, KIF23, PRC1, CCNA2, DEPDC1 and TTK were significantly upregulated ($P < 0.05$) (Fig. 6i). To further explore the protein expression of seven genes (CNFN, KIF18B, KIF23, PRC1, CCNA2, DEPDC1 and TTK), we further performed the protein expression analysis using the HPA databse (Fig. 6j). Statistical analysis indicated that CNFN expression was low and not detected (100%; n=4); KIF18B (66.7%; n=3), KIF23 (100%; n=4), PRC1 (75.0%; n=4), CCNA2 (66.7%; n=3), DEPDC1 (100%; n=3) and TTK (66.7%; n=3) were significantly medium and high expression (Fig. 6k).

Establishment of prognostic risk score formula

Using the LASSO method and multivariable Cox regression analysis, two mRNAs (CNFN and DEPDC1) were identified as integrated prognostic biomarkers for HNSCC patients. We then established a prognostic risk score formula based on the expression profiles of two prognostic mRNAs and their regression coefficients. The prognostic risk score formula was as follows: risk score = EXP_{DEPDC1} * 0.32636 + EXP_{CNFN} * (-0.07544). The risk scores were calculated for all patients and split patients into a high-risk group (n = 165) and a low-risk group (n = 165) by using the median risk score as the cutoff value (Additional file 6: Table S6). The distribution of risk scores and survival status of the patients were shown in Fig. 7a and b.

We then assessed the prognostic value of the above risk formula by using Kaplan-Meier analysis. We found that the low-risk group had a better OS than the high-risk group (P < 0.001) (Fig. 7c). Moreover, time-dependent ROC analysis was also utilized to evaluate the prognostic capacity of the risk formula. The areas under the ROC curve at 1, 3 and 5 years were 0.582, 0.634 and 0.636, respectively, which suggested that the integrated two-mRNA signature had better utility than each single one (Fig. 7d).

Identification of small molecular agents

To know the small molecular agents targeted LOI in the turquoise and pink modules, we searched all drug-gene interaction in DrugBank database. Connection degree ≥ 2 and $P < 0.05$ were used to screen the drug-module, five drug-module interactions (XL844, AT7519, AT9283, Alvocidib and Nelarabine) in turquoise module and three drug-module interactions (Benzamidine, L-Glutamine and Zinc) in pink module may be used to target LOI ($P < 0.05$) (Table 2). To further know the clinical application of 8 small molecular agents on head and neck cancer or solid tumor, we analyzed clinical trial registration of these small molecular agents using the clinicaltrials (https://clinicaltrials.gov/ct2/home). And as shown in Additional file 7: Table S7, although the study on Benzamidine was still unexplored, three clinical trials of L-Glutamine (NCT03015077, NCT02282839, NCT00006994) and three clinical trials of Zinc (NCT00036881, NCT03531190, NCT02868151) on head and neck cancer had been conducted. Meanwhile, XL844 (NCT00475917), AT7519 (NCT00390117, NCT02503709), AT9283 (NCT00443976, NCT00985868), Alvocidib (NCT00080990) and Nelarabin (NCT01376115) on solid tumor or cancer had also been studied. These results indicated that XL844, AT7519, AT9283, Alvocidib, Nelarabine, Benzamidine, L-Glutamine and Zinc may provide us new approach to block metastasis of lymph nodes.

Discussion

Metastasis is the main cause of HNSCC treatment failure [20]. Nodal metastatic disease is considered as a independent factor for poor survival in HNSCC [21-23]. Several clinicopathologic parameters have validated as associated with nodal metastasis, such as tumor size [9], tumor depth [24], tumor differentiation [25], histological grade [26] and LOI [4]. In this study, we performed a systematic analysis of LOI in HNSCC from the molecular level to the clinical level by comprehensive integrating genomic analysis. We established a novel two-mRNA signature was constructed for predicting LOI risk of HNSCC. Survival curves indicated HNSCC with low-risk and high-risk groups stratified by mRNA signature had a significant prognosis difference. Time-dependent ROC analysis indicated that the mRNA signature had a high predictive accuracy in predicting OS. The small molecular agents including XL844, AT7519, AT9283, Alvocidib, Nelarabine, Benzamidine, L-Glutamine and Zinc may provide us new approach to block LOI.

With the application of sequencing techniques, genomic studies have transformed from individual genes aberrant expression to systematically integrating genomics mutation and chromatin remodel study. However, the molecular mechanisms of LOI are still unclear. TCGA database has listed several genomic landscape study of HNSCC from worldwide, which provides us a chance to integrate genomics data and

understand the molecular changes of LOI. In this study, we conducted a co-expression network module of HNSCC and found the turquoise and pink modules were significantly associated with LOI. Function enrichment analysis indicated that the key gene modules function was not only involved in squamous cell function, such as, epidermal cell differentiation, keratinocyte differentiation skin development, epidermis development and cornification, but also involved in regulation of protein secretion, such as, peptidase activity, negative regulation of proteolysis, negative regulation of peptidase activity and negative regulation of endopeptidase activity and DNA function, such as, DNA replication, mitotic nuclear division, nuclear division and DNA–dependent DNA replication. Pathway enrichment analysis validated that the genes in the key module were enriched in cell cycle, DNA replication, mismatch repair and p53 signaling pathway, which indicated that the key gene modules play an important role in LOI in HNSCC.

Lymphatic vessels are remodeled by the tumor microenvironment, including cancer cell, the mutation of oncogenic driver genes, and the interaction of immune-check point signals and their receptors [27]. In our study, we systematically analyzed the mRNA expression of 339 HNSCC with LOI and 44 normal sample, 2522 genes were found significantly differential expression. Protein-protein interaction network and module analysis showed 18 genes in turquoise module, such as KIF18B, BUB1, BUB1B, KIF4A, EXO1 and 6 genes in pink module, such as KRT78, CNFN, SLURP1, PRSS27, CRCT1, were associated with LOI of HNSCC. However, the role and mechanism of 24 genes on the metabolic and immune reprogramming of the tumor microenvironment were still need further explored.

It is essential to early diagnosis of LOI, as HNSCC patients with LOI may require more timely treatment [28, 29]. Despite the development and application of MRI and PET-CT to assess LOI in HNSCC, the detection rate of early-stage LOI is still rarely low [30]. In this study, the hub genes in the key modules related to LOI were screened and prognosis analysis and expression analysis showed that CNFN was downregulated and associated with good prognosis, whereas, KIF18B, KIF23, PRC1, CCNA2, DEPDC1 and TTK were upregualted and associated poor prognosis. The two-mRNA signature could stratify the risk of LOI and predict the OS of HNSCC. However, there are also some limitations. First, the two-mRNA signature need be further explored. The second limitation was that the prognostic value of the mRNA panel is not very satisfactory and be further explored. Third, the biological functions and mechanism of the two mRNAs were not assessed in this study.

Although the targeted treatment for LOI is lacking and unreliable, DrugBank provides comprehensive molecular information about drugs and their targets for the treatment of LOI. Based on interactions of drug and key modules, we found 8 small molecular agents that could target LOI, including Benzamidine, L-Glutamine, Zinc, XL844, AT7519, AT9283, Alvocidib and Nelarabine. A recent study found that AT7519 and Alvocidib, a cyclin-dependent kinase inhibitor, has been shown to have potential for anticancer effects for cancer treatment by targeting CDK1 [31-35]. XL844, a specific inhibitor of mitotic spindle checkpoint kinase-1 and mitotic spindle checkpoint kinase-2 kinase, has been found to effectively sensitize cancer cells to induce cell cycle arrest [36]. Clinical trial registration analysis of 8 small molecular agents has also indicated that these small molecular agents have also been widely conducted in head and neck cancer and solid tumor. These results indicated the 8 small molecular agents may be

used as treating LOI of HNSCC by targeting DNA damage, which was consistent with function enrichment analysis and pathway analysis. However, the role and mechanism of 8 small molecular agents targeted LOI need further be explored.

Conclusions

In summary, the current study was intended to identify LOI with comprehensive bioinformatics analysis and we contructed a key two-mRNA signature that was significantly associated with OS in HNSCC patients, which could accurately identify patients with low LOI risk from those with high LOI risk. Benzamidine, L-Glutamine, Zinc, XL844, AT7519, AT9283, Alvocidib and Nelarabine may be a candidate drug for treating LOI.

Abbreviations

LOI: Lymphovascular invasion (LOI), HNSCC: head and neck squamous cell carcinoma, DEGs: differentially expressed genes, MEs: module eigengenes, GS: gene significance, MM: module membership, WGCNA: Weighted gene co-expression network analysis, MCODE: Molecular Complex Detection, MRI: magnetic resonance imaging, PET-CT: positron emission tomography-computed tomography.

Declarations

Ethics approval and consent to participate

Written Informed consent was obtained from all patients before treatment and this study was approved by the institutional research ethics committee of Affiliated Cancer Hospital & Institute of Guangzhou Medical University (2019-290).

Consent for publication

Not applicable.

Availability of data and materials

All of the data of this study has been downloaded from TCGA database (https://cancergenome.nih.gov/), OMIM database (https://www.omim.org/), NCBI gene databse (https://www.ncbi.nlm.nih.gov/gene/) and DrugBank database (https://www.drugbank.ca/).

Competing interest

The authors declare that they have no competing interests.

Funding

This work was supported by grants from the Social Science and Technology Development Key Project of Dongguan (201750715046462); Guangzhou Key Medical Discipline Construction Project Fund (B195002004042); Open Funds of State Key Laboratory of Oncology in South China (KY013711).

Acknowledgements

We would like to thank the native English-speaking scientists of Elixigen Company (Huntington Beach, California) for editing our manuscript.

Author contributions

JZ, HLJ and TX designed the research. JZ, RL, BYW, JL, AAX and XTH acquired and analyzed the data. JZ and YY wrote the manuscript.

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015;65(2):87-108. doi: 10.3322/caac.21262.
2. Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. Oral Oncol. 2009;45(4-5):309-16. doi: 10.1016/j.oraloncology.2008.06.002.
3. Vasan K, Low TH, Gupta R, Ashford B, Asher R, Gao K, Ch'ng S, Palme CE, Clark JR. Lymph node ratio as a prognostic factor in metastatic cutaneous head and neck squamous cell carcinoma. Head Neck. 2018;40(5):993-999. doi: 10.1002/hed.25066.
4. Wreesmann VB, Katabi N, Palmer FL, Montero PH, Migliacci JC, Gönen M, Carlson D, Ganly I, Shah JP, Ghossein R, et al. Influence of extracapsular nodal spread extent on prognosis of oral squamous cell carcinoma. Head Neck. 2016;38 Suppl 1:E1192-9. doi: 10.1002/hed.24190.
5. Liu SA, Wang CC, Jiang RS, Lee FY, Lin WJ, Lin JC. Pathological features and their prognostic impacts on oral cavity cancer patients among different subsites - A singe institute's experience in Taiwan. Sci Rep. 2017;7(1):7451. doi: 10.1038/s41598-017-08022-w.
6. Moore BA, Weber RS, Prieto V, El-Naggar A, Holsinger FC, Zhou X, Lee JJ, Lippman S, Clayman GL. Lymph node metastases from cutaneous squamous cell carcinoma of the head and neck.

Laryngoscope. 2005;115(9):1561-7.

7. Yilmaz T, Hosal AS, Gedikoglu G, Onerci M, Gürsel B. Prognostic significance of vascular and perineural invasion in cancer of the larynx. Am J Otolaryngol. 1998;19(2):83-8.
8. Karahatay S, Thomas K, Koybasi S, Senkal CE, Elojeimy S, Liu X, Bielawski J, Day TA, Gillespie MB, Sinha D, et al. Clinical relevance of ceramide metabolism in the pathogenesis of human head and neck squamous cell carcinoma (HNSCC): attenuation of C(18)-ceramide in HNSCC tumors correlates with lymphovascular invasion and nodal metastasis. Cancer Lett. 2007;256(1):101-11.
9. Kurokawa H, Yamashita Y, Takeda S, Zhang M, Fukuyama H, Takahashi T. Risk factors for late cervical lymph node metastases in patients with stage I or II carcinoma of the tongue. Head Neck. 2002;24(8):731-6.
10. Hahn SS, Spaulding CA, Kim JA, Constable WC. The prognostic significance of lymph node involvement in pyriform sinus and supraglottic cancers. Int J Radiat Oncol Biol Phys. 1987;13(8):1143-7.
11. Lu JM, Chen YC, Ao ZX, Shen J, Zeng CP, Lin X, Peng LP, Zhou R, Wang XF, Peng C, et al. System network analysis of genomics and transcriptomics data identified type 1 diabetes-associated pathway and genes. Genes Immu. 2019;20(6):500-508. doi: 10.1038/s41435-018-0045-9.
12. Yuan L, Chen L, Qian K, Qian G, Wu CL, Wang X, Xiao Y. Co-expression network analysis identified six hub genes in association with progression and prognosis in human clear cell renal cell carcinoma (ccRCC). Genom Data. 2017;14:132-140. doi: 10.1016/j.gdata.2017.10.006.
13. Zhang Y, Wang J, Ji LJ, Li L, Wei M, Zhen S, Wen CC. Identification of Key Gene Modules of Neuropathic Pain by Co-Expression Analysis. J Cell Biochem 2017;118(12):4436-4443. doi: 10.1002/jcb.26098.
14. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res. 2003;31(1):258-61.
15. Xia WX, Yu Q, Li GH, Liu YW, Xiao FH, Yang LQ, Rahman ZU, Wang HT, Kong QP. Identification of four hub genes associated with adrenocortical carcinoma progression by WGCNA. Peer J. 2019;7:e6555. doi: 10.7717/peerj.6555.
16. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.
17. 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.
18. Zhang J, Zheng Z, Zheng J, Xie T, Tian Y, Li R, Wang B, Lin J, Xu A, Huang X et al. Epigenetic-Mediated Downregulation of Zinc Finger Protein 671 (ZNF671) Predicts Poor Prognosis in Multiple Solid Tumors. Front Oncol 2019; 9:342.
19. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018;46(D1):D1074-D1082. doi: 10.1093/nar/gkx1037.

20. Leeman JE, Li JG, Pei X, Venigalla P, Zumsteg ZS, Katsoulakis E, Lupovitch E, McBride SM, Tsai CJ, Boyle JO, et al. Patterns of Treatment Failure and Postrecurrence Outcomes Among Patients With Locally Advanced Head and Neck Squamous Cell Carcinoma After Chemoradiotherapy Using Modern Radiation Techniques. JAMA Oncol. 2017; 3(11):1487-1494. doi: 10.1001/jamaoncol.2017.0973.
21. Layland MK, Sessions DG, Lenox J. The influence of lymph node metastasis in the treatment of squamous cell carcinoma of the oral cavity, oropharynx, larynx, and hypopharynx: N0 versus N+. Laryngoscope. 2005;115(4):629-39.
22. Sessions DG, Spector GJ, Lenox J, Parriott S, Haughey B, Chao C, Marks J, Perez C. Analysis of treatment results for floor-of-mouth cancer. Laryngoscope. 2000;110(10 Pt 1):1764-72.
23. Sessions DG, Lenox J, Spector GJ, Chao C, Chaudry OA. Analysis of treatment results for base of tongue cancer. Laryngoscope. 2003; 113(7):1252-61.
24. Pentenero M, Gandolfo S, Carrozzo M. Importance of tumor thickness and depth of invasion in nodal involvement and prognosis of oral squamous cell carcinoma: a review of the literature. Head Neck. 2005;27(12):1080-91.
25. Byers RM, El-Naggar AK, Lee YY, Rao B, Fornage B, Terry NH, Sample D, Hankins P, Smith TL, Wolf PJ. Can we detect or predict the presence of occult nodal metastases in patients with squamous carcinoma of the oral tongue? Head Neck. 1998;20(2):138-44.
26. Umeda M, Yokoo S, Take Y, Omori A, Nakanishi K, Shimada K. Lymph node metastasis in squamous cell carcinoma of the oral cavity: correlation between histologic features and the prevalence of metastasis. Head Neck. 1992;14(4):263-72.
27. Achen MG, Stacker SA. Molecular control of lymphatic metastasis. Ann N Y Acad Sci. 2008;1131(1):225-234.
28. Solares CA, Mason E, Panizza BJ. Surgical Management of Perineural Spread of Head and Neck Cancers. J Neurol Surg B Skull base. 2016;77(2):140-9. doi: 10.1055/s-0036-1579751.
29. Bur AM, Lin A, Weinstein GS. Adjuvant radiotherapy for early head and neck squamous cell carcinoma with perineural invasion: A systematic review. Head Neck. 2016;38 Suppl 1:E2350-7. doi: 10.1002/hed.24295.
30. Lee H, Lazor JW, Assadsangabi R, Shah J. An Imager's Guide to Perineural Tumor Spread in Head and Neck Cancers: Radiologic Footprints on (18)F-FDG PET, with CT and MRI Correlates. J Nucl Med. 2019;60(3):304-311. doi: 10.2967/jnumed.118.214312.
31. Dolman ME, Poon E, Ebus ME, den Hartog IJ, van Noesel CJ, Jamin Y, Hallsworth A, Robinson SP, Petrie K, Sparidans RW, et al. Cyclin-Dependent Kinase Inhibitor AT7519 as a Potential Drug for MYCN-Dependent Neuroblastoma. Clin Cancer Res. 2015;21(22):5100-9. doi: 10.1158/1078-0432.CCR-15-0313.
32. Kang MA, Kim W, Jo HR, Shin YJ, Kim MH, Jeong JH. Anticancer and radiosensitizing effects of the cyclin-dependent kinase inhibitors, AT7519 and SNS032, on cervical cancer. Int J Oncol. 2018;53(2):703-712. doi: 10.3892/ijo.2018.4424.

33. Kong Y, Sheng X, Wu X, Yan J, Ma M, Yu J, Si L, Chi Z, Cui C, Dai J, et al. Frequent Genetic Aberrations in the CDK4 Pathway in Acral Melanoma Indicate the Potential for CDK4/6 Inhibitors in Targeted Therapy. Clin Cancer Res. 2017;23(22):6946-6957. doi: 10.1158/1078-0432.CCR-17-0070.
34. Hafner M, Mills CE, Subramanian K, Chen C, Chung M, Boswell SA, Everley RA, Liu C, Walmsley CS, Juric D, et al. Multiomics Profiling Establishes the Polypharmacology of FDA-Approved CDK4/6 Inhibitors and the Potential for Differential Clinical Activity. Cell Chem Biol. 2019;26(8):1067-1080. doi: 10.1016/j.chembiol.2019.05.005.
35. Roskoski R, Jr. Cyclin-dependent protein serine/threonine kinase inhibitors as anticancer drugs. Pharmacol Res. 2019;139:471-488. doi: 10.1016/j.phrs.2018.11.035.
36. Matthews DJ, Yakes FM, Chen J, Tadano M, Bornheim L, Clary DO, Tai A, Wagner JM, Miller N, Kim YD, et al. Pharmacological abrogation of S-phase checkpoint enhances the anti-tumor activity of gemcitabine in vivo. Cell cycle. 2007; 6(1):104-10.

Additional Files

Additional file 1: Table S1 The differentially expressed genes in TCGA database.

Additional file 2: Table S2 The genes related to HNSCC in NCBI gene database.

Additional file 3: Table S3 The genes related to HNSCC in OMIM database.

Additional file 4: Table S4 The candidate hub genes in turquoise module.

Additional file 5: Table S5 The candidate hub genes in pink module.

Additional file 6: Table S6 The risk score.

Additional file 6: Table S7 Clinical trials of small molecular agents.

Tables

Table 1
Clinical pathological characteristics of 500 patients with HNSCC

Parameters	Subtype	Patients
Age	> 61	234
	≤ 61	265
	Unknow	1
Gender	Male	367
	Female	133
Lymphovascular invasion	Yes	120
	No	219
	Unknow	161
Pathologic stage	Stage I-II	125
	Stage III-IV	337
	Unknow	68
OS times(months)	< 1	8
	≥ 1	491
	Unknow	1

Table 2
Enriched significant of small molecules

module	drug	connection	P value
pink	Benzamidine	2	8.29E-07
pink	L-Glutamine	2	1.28E-05
pink	Zinc	2	0.001325331
turquoise	XL844	2	0
turquoise	AT7519	2	0
turquoise	AT9283	2	0
turquoise	Alvocidib	5	3.55E-06
turquoise	Nelarabine	2	0.000198791

Figures

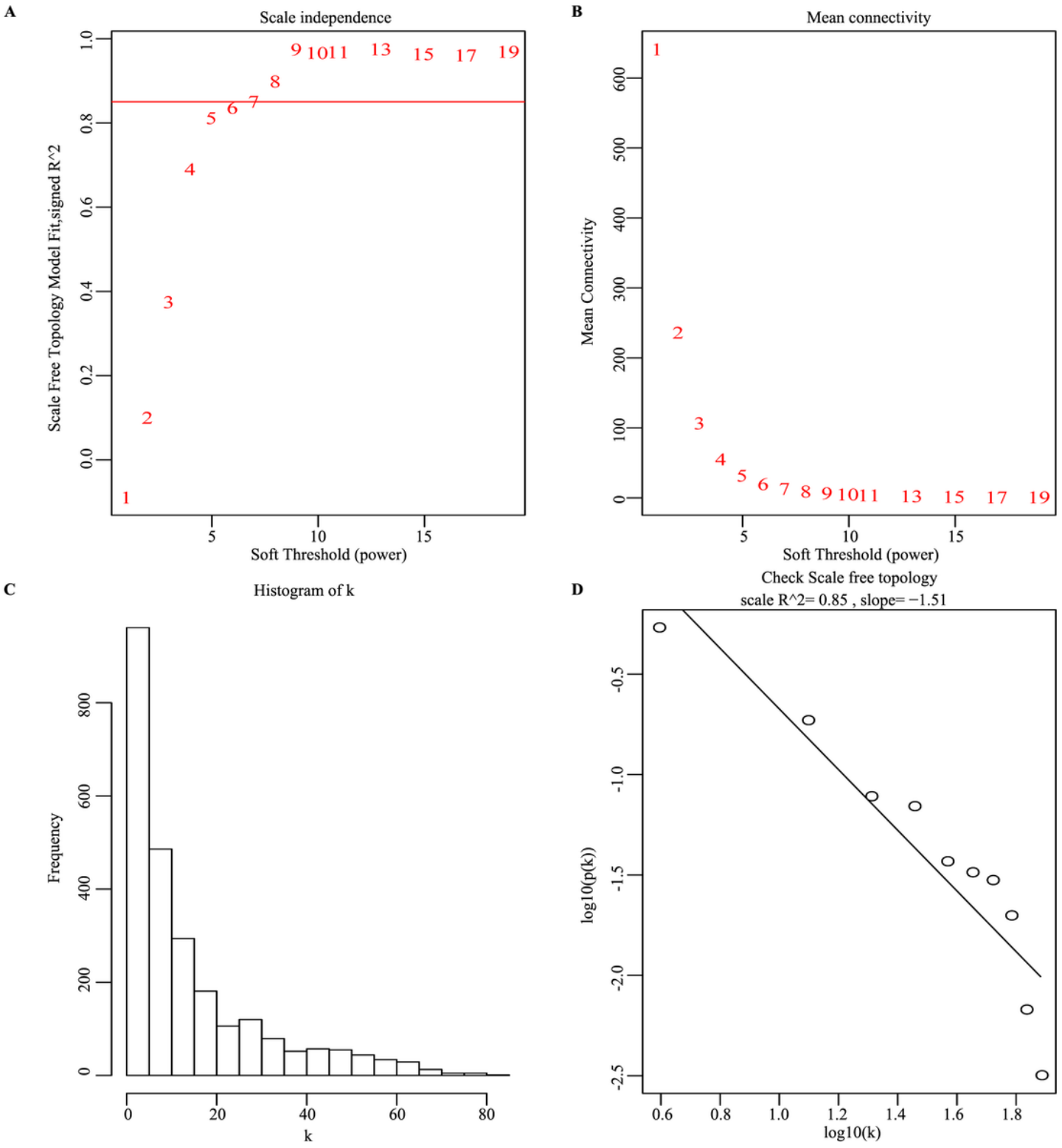

Figure 1

Determination of soft-thresholding power in the WGCNA. a The scale-free index analysis for soft-thresholding powers (β) in HNSCC. b The mean connectivity analysis for various soft-thresholding powers in HNSCC. c Histogram of connectivity distribution when β=7 in HNSCC. d Checking the scale-free topology when β=7 in HNSCC.

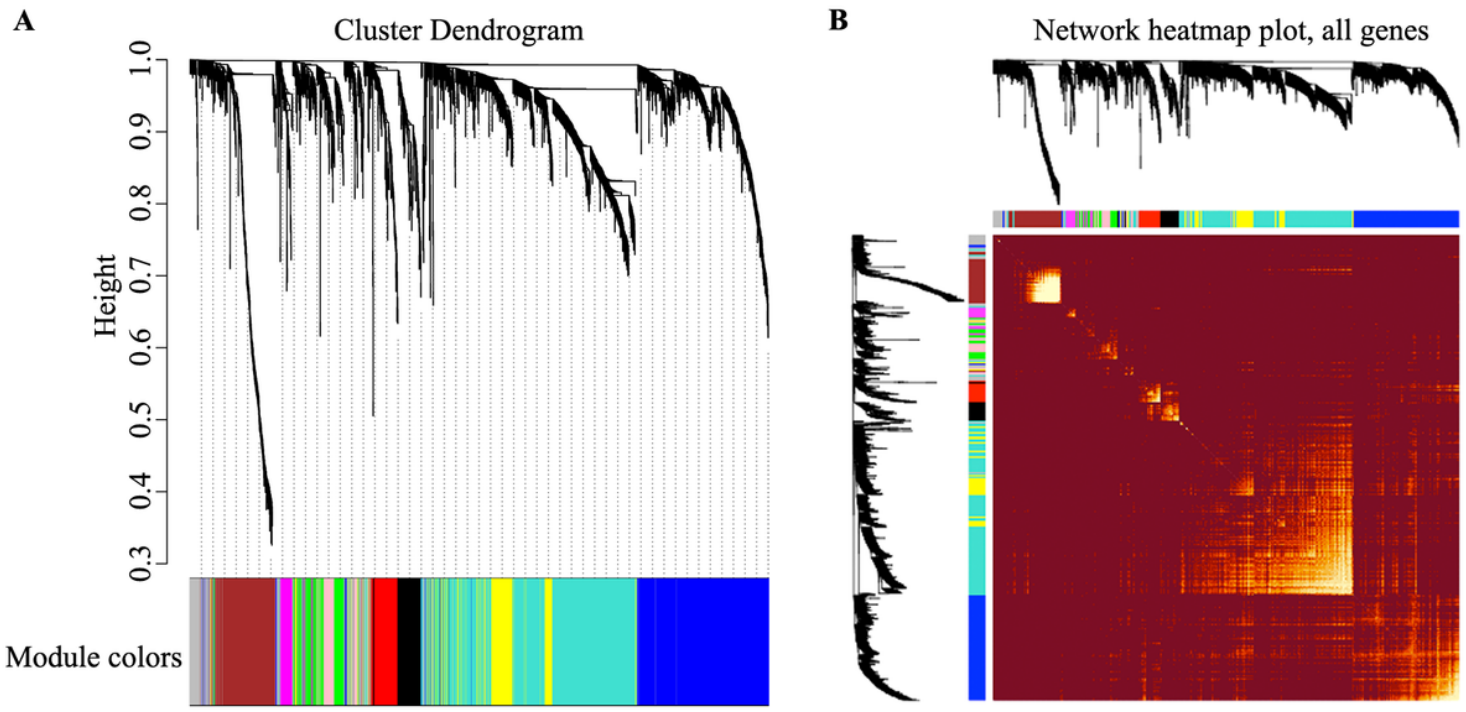

Figure 2

Visualization of WGCNA results. a mRNA clustering dendrogram obtained by hierarchical cluster analysis of of TOM-based dissimilarity with the corresponding module colors indicated by the color row. Each colored row represents color-coded module which contains a group of highly connected mRNAs. Each color represents a module in the constructed gene co-expression network. b The heatmap depicts the Topological Overlap Matrix among all genes in the WGCNA analysis. Light color represents low overlap and progressively darker red color represents higher overlap

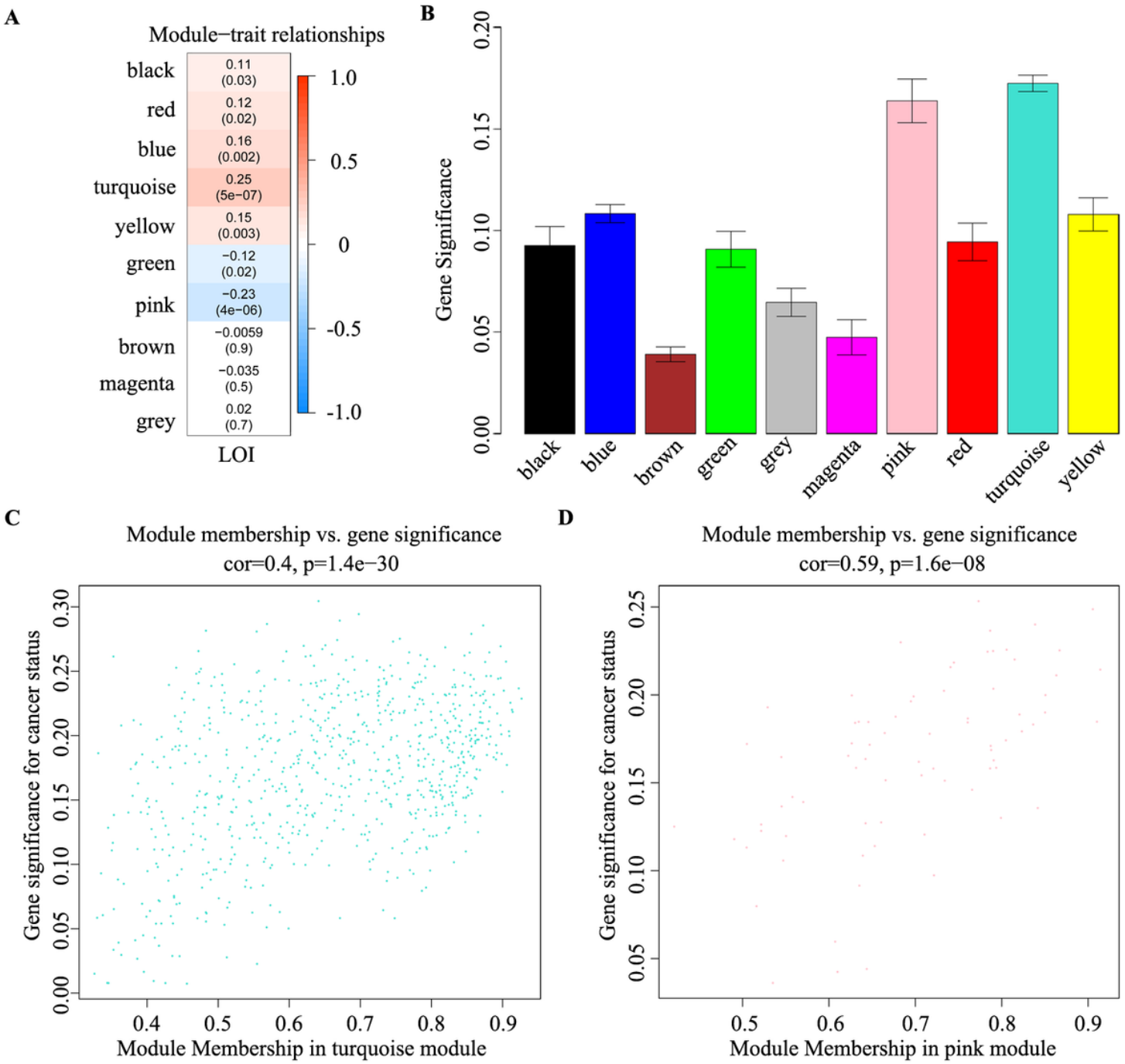

Figure 3

The correlation analysis of module–trait and clinical characteristics. a The column corresponds to LOI phenotypic trait, labeled below. Heatmap of each cell at the row-column contains the p-value between that module and LOI trait. The correlations between turquoise module with LOI phenotypic trait (cor=0.25; P=5e-07) and pink module with LOI phenotypic trait (cor=-0.23; P=4e-06) were significant. b Bar plot of the significance level of 10 co-expression modules associated with LOI status. (c and d) Correlation analysis between gene significance of LOI status and module membership in the turquoise (c) and pink (d) modules.

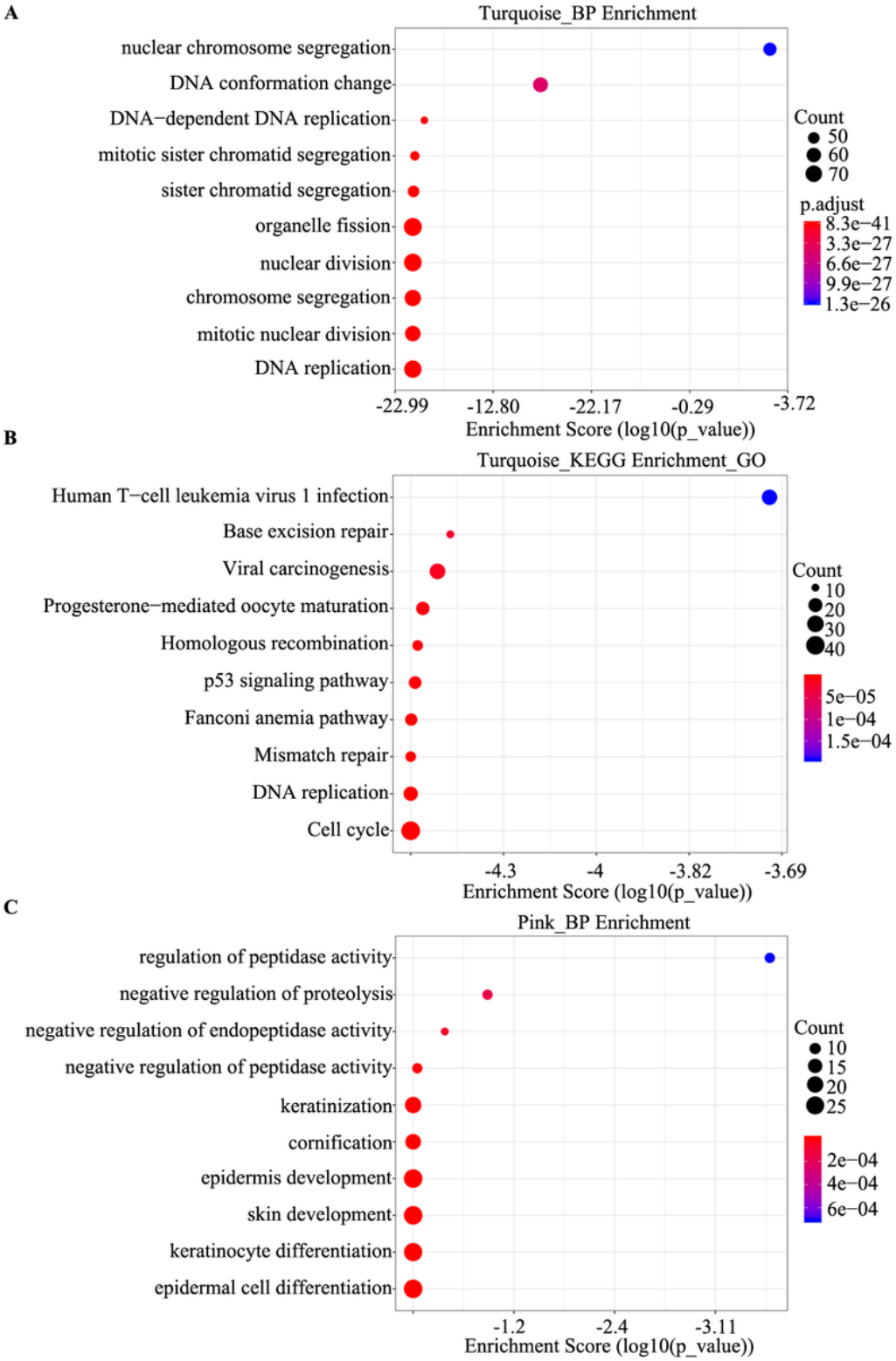

Figure 4

GO and KEGG analysis in the two key turquoise and pink modules. a Go enriched analysis of turquoise module in the biological process. b Go enriched analysis of turquoise module in the KEGG pathway. c Go enriched analysis of pink module in the biological process.

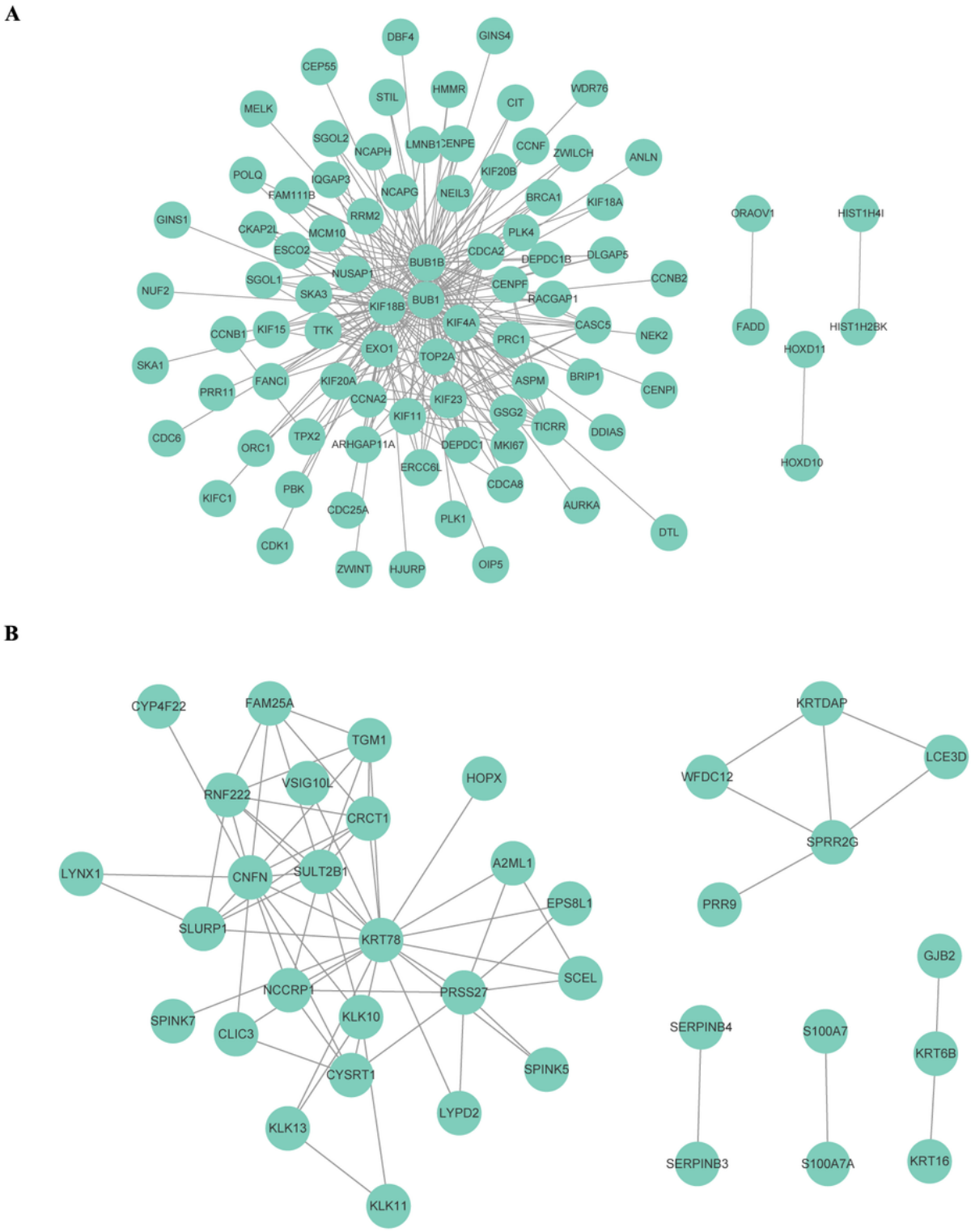

Figure 5

Hub genes identified by PPI network in the modules. (a and b) PPI interaction network of the DEGs in the turquoise module (a) and pink module (b).

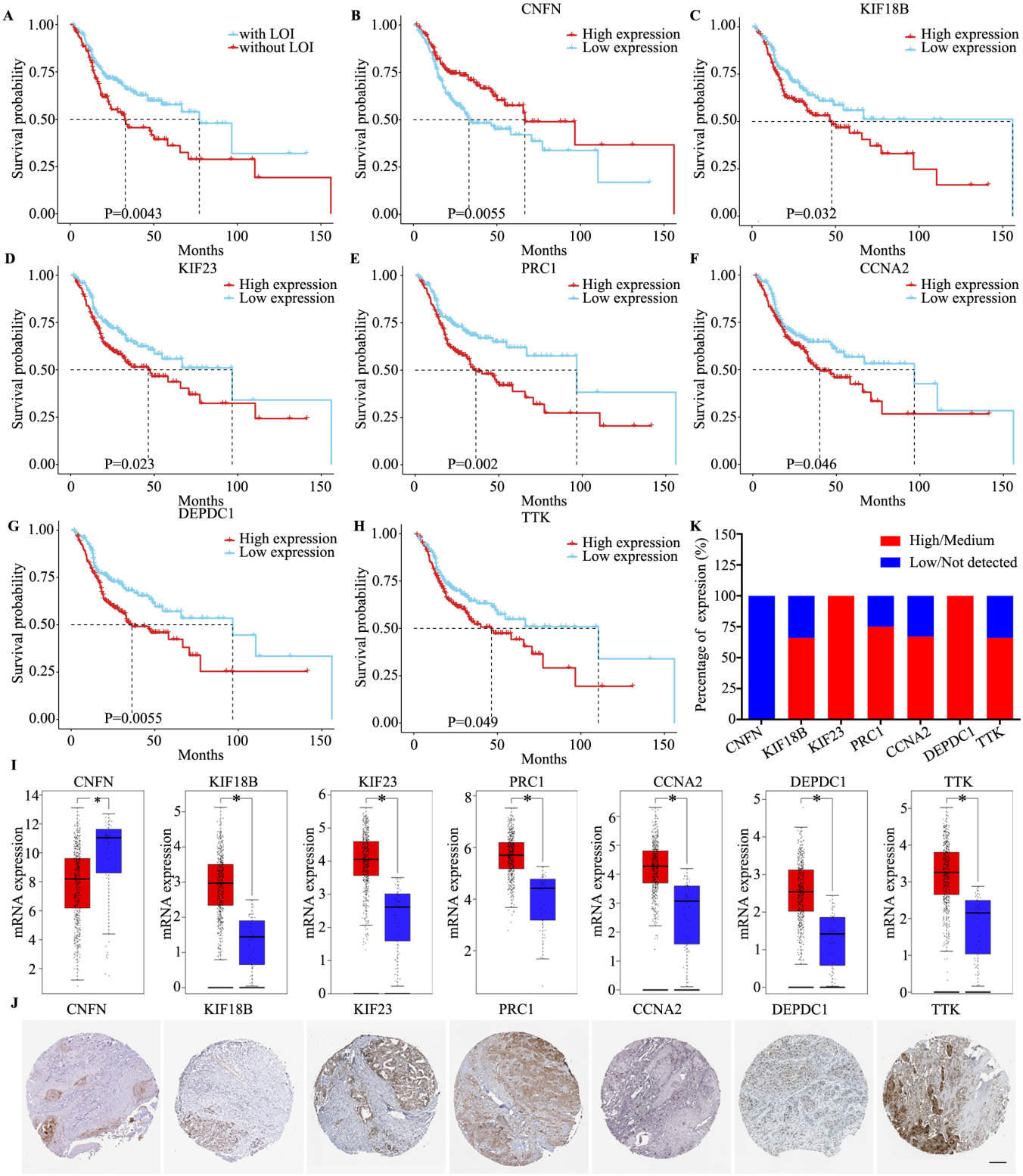

Figure 6

Prognostic value and expression analysis of seven hub genes in HNSCC. a Ten years cumulative survival of HNSCC patients with or without LOI. (b-h) Ten years survival analysis of CNFN (b), KIF18B (c), KIF23 (d), PRC1 (e), CCNA2 (f), DEPDC1 (g) and TTK (h). i mRNA expression of seven hub genes (CNFN, KIF18B, KIF23, PRC1, CCNA2, DEPDC1 and TTK) in HNSCC (n=519; Red box) and normal tissues (n=44; blue box) based on GEPIA database. j Immunohistochemistry images of seven hub genes (CNFN, KIF18B, KIF23,

PRC1, CCNA2, DEPDC1 and TTK) in head and neck cancer based on Human Protein Atlas database. k Protein expression levels in head and neck cancer analyzed by IHC in the Human Pathology Atlas database. **P<0.01 and *P<0.05.

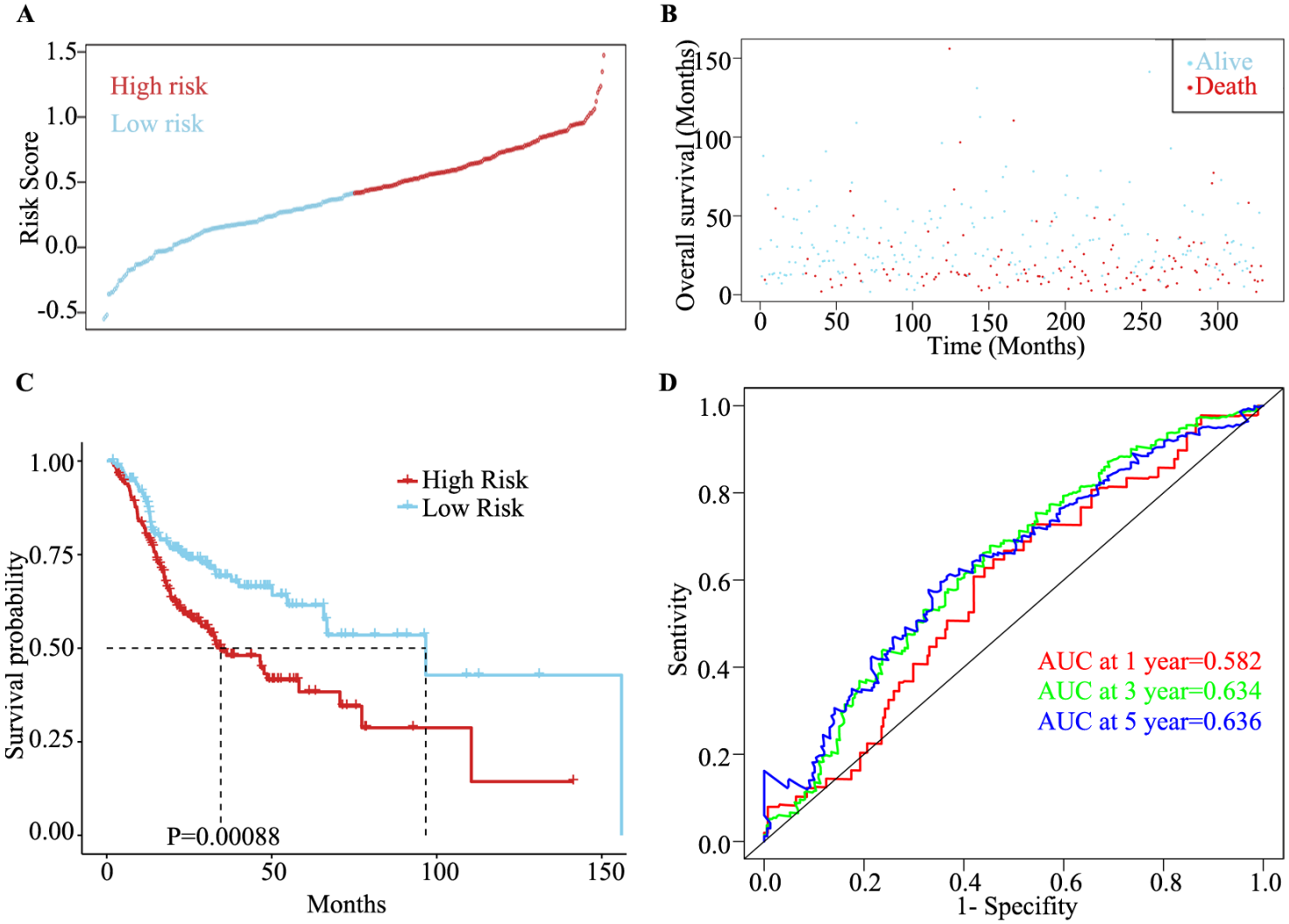

Figure 7

The distribution of risk score, survival status and time-dependent ROC analysis of integrated two mRNAs. a the risk score distribution b the overall survival status of 330 patients. c the Kaplan–Meier curve of the overall survival (OS) between the low-risk and high-risk groups split by median risk score. d Time-dependent ROC analysis for the 1-, 3- and 5-year OS probability.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS2.xlsx
- TableS1.xlsx
- TableS7.xlsx

- TableS6.xlsx
- TableS3.xlsx
- TableS5.xlsx
- TableS4.xlsx