


Identification of an 8 HPV-related RNA signature as a novel prognostic biomarker for squamous cell carcinoma of the head and neck

Zhang Mei, MD^a, Luo Zhengdong, MM^b, Liu Shupeng, MM^c, Zhang Xin, MD^d, Wang Lei, MD^d, Chuanxin Wang, MD^{e,*} 

Abstract

Squamous cell carcinoma of the head and neck (SCCHN) is a commonly detected cancer worldwide. Human papillomavirus (HPV) is emerging as an important risk factor affecting SCCHN prognosis. Therefore, identification of HPV status is essential for effective therapies in SCCHN. The aim of this study was to investigate the prognostic value of HPV-associated RNA biomarkers for SCCHN. The clinical data, survival data, and RNA-seq data of SCCHN were downloaded from The Cancer Genome Atlas database. Before the differential expression analysis, the heterogeneity between the 2 groups (HPV+ vs HPV-) of samples was analyzed using principal component analysis. The differentially expressed genes (DEGs) between HPV+ and HPV- SCCHN samples were analyzed using the R edgeR package. The Gene Ontology functional annotations, including biological process, molecular function and cellular component (CC), and Kyoto Encyclopedia of Genes and Genomes pathways enriched by the DEGs were analyzed using DAVID. The obtained matrix was analyzed by weighed gene coexpression network analysis. A total of 350 significant DEGs were identified through differential analysis, and these DEGs were significantly enriched in functions associated with keratinization, and the pathway of neuroactive ligand-receptor interaction. Moreover, 72 hub genes were identified through weighed gene coexpression network analysis. After the hub genes and DEGs were combined, we obtained 422 union genes, including 65 survival-associated genes. After regression analysis, a HPV-related prognostic model was established, which consisted of 8 genes, including *Clorf105*, *CGA*, *CHRNA2*, *CRIP3*, *CTAG2*, *ENPP6*, *NEFH*, and *RNF212*. The obtained regression model could be expressed by an equation as follows: risk score = 0.065 × *Clorf105* + 0.012 × *CGA* + 0.01 × *CHRNA2* + 0.047 × *CRIP3* + 0.043 × *CTAG2* – 0.034 × *ENPP6* – 0.003 × *NEFH* – 0.068 × *RNF212*. *CGA* interacted with 3 drugs, and *CHRNA2* interacted with 11 drugs. We have identified an 8 HPV-RNA signature associated with the prognosis of SCCHN patients. Such prognostic model might serve as possible candidate biomarker and therapeutic target for SCCHN.

Abbreviations: BP = biological process, CC = cellular component, *CGA* = alpha polypeptide, *CHRNA2* = cholinergic receptor nicotinic alpha 2 subunit, *CRIP3* = cysteine rich protein 3, *CTAG2* = cancer/testis antigen 2, DEGs = differentially expressed genes, DGIdb = drug-gene interaction database, *ENPP6* = ectonucleotide pyrophosphatase/phosphodiesterase 6, ERα = estrogen receptor alpha, FC = fold change, FDR = false discovery rate, GO = Gene Ontology, HPV = Human papillomavirus, KEGG = Kyoto Encyclopedia of Genes and Genomes, MAD = median absolute deviation, MF = molecular function, miRNA = microRNA, *NEFH* = neurofilament heavy, PCA = principal component analysis, *RNF212* = ring finger protein 212, ROC = receiver operating characteristic, SCCHN = Squamous cell carcinoma of the head and neck, TCGA = The Cancer Genome Atlas, TMM = trimmed mean of M values, WGCNA = weighed gene coexpression network analysis.

Keywords: biomarkers, head and neck neoplasms, prognosis

This work was supported by Medical and Health Technology Development Plan of Shandong Province (2013WS0228), and Shandong Technological Development Project (2016CYJS01A02).

The authors have no conflicts of interest to disclose.

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Supplemental Digital Content is available for this article.

^a Department of Dental, Shandong Medical College, Jinan, China, ^b Department of Clinical Laboratory, Qilu Hospital of Shandong University, Jinan, China, ^c Department of Outpatient, Yidu Central Hospital of Weifang, Weifang, China, ^d Department of Orthodontics, Qilu Hospital of Shandong University, Jinan, China, ^e Department of Clinical Laboratory, the Second Hospital of Shandong University, Jinan, China.

* Correspondence: Chuanxin Wang, Department of Clinical Laboratory, the Second Hospital of Shandong University, No. 247 Beiyuan Street, 250033 Jinan, China. (e-mail: cwxwang@sdu.edu.cn).

Copyright © 2024 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Mei Z, Zhengdong L, Shupeng L, Xin Z, Lei W, Wang C. Identification of an 8 HPV-related RNA signature as a novel prognostic biomarker for squamous cell carcinoma of the head and neck. *Medicine* 2024;103:6(e36448).

Received: 11 October 2021 / Received in final form: 7 October 2023 / Accepted: 13 November 2023

<http://dx.doi.org/10.1097/MD.00000000000036448>

1. Introduction

Squamous cell carcinoma of the head and neck (SCCHN), characterized by phenotypic, biological, etiological, and clinical heterogeneity, is a commonly detected cancer worldwide, with >600,000 newly diagnosed cases every year.^[1,2] Most SCCHN patients present with locoregionally advanced disease, with >50% of patients relapsing within 3 years.^[3,4] Previous investigations have suggested that tobacco use and alcohol consumption play a fundamental role in the pathogenesis of SCCHN in developing countries. Recent studies have revealed that human papillomavirus (HPV) is emerging as an important risk factor affecting nonsmokers in developed countries.^[5,6] It is well known that HPV-associated oropharyngeal cancer is a carcinogenic subtype that differs from HPV-negative tumors in terms of genetic changes and better prognosis.^[7] Therefore, the HPV status may have significant prognostic importance in SCCHN patients.

The recent identification of prognostic signatures has significantly improved our understanding about cancer progression and survival rate.^[8] Many prognostic gene-expression signatures associated with human cancers have been developed to predict the prognostic outcomes of various cancers.^[9–11] The development of prognostic signatures needs in-depth analysis of the genetic profiles. Nowadays, the high-throughput gene-expression profiles have been made available by The Cancer Genome Atlas (TCGA), providing a comprehensive set of patient genetic profiles across multiple cancer types,^[12] including SCCHN. A total of 529 SCCHN samples involving RNA-seq and miRNA-seq profiles are available in TCGA database.^[12] Shi et al^[13] have identified a 6-microRNA (miRNA) signature as a novel potential prognostic biomarker for SCCHN based on the TCGA dataset. Wong et al^[14] have also systematically identified biomarkers associated with the outcomes of SCCHN patients using TCGA database. However, to the best of our knowledge, no studies have reported the prognostic biomarkers associated with SCCHN with different HPV status.

In the present study, we aimed to investigate the prognostic value of HPV-associated RNA biomarkers for SCCHN using the profiling data obtained from TCGA. Therefore, we performed a series of analyses on the data of SCCHN with clinical information of HPV infection (HPV+ or HPV-). Our findings provided valuable insights into the best course of treatment for SCCHN patients.

2. Materials and methods

2.1. Data collection

TCGA database^[15] in UCSC Xena (<https://xenabrowser.net/datapages/>) includes the clinical data, survival data, and RNA-seq data of SCCHN. In the present study, the SCCHN samples with the diagnostic information of HPV were included in the analysis, and the samples lacking the diagnosis information of HPV and normal tissue samples were excluded. The sample selection criteria were set as follows: (1) sample type was primary tumor; and (2) HPV status was determined by “hpv_status_by_p16_testing” information. For those without “hpv_status_by_p16_testing” information, HPV status was obtained by filling with “hpv_status_by_ish_testing” information.

2.2. Differential expression analysis

Prior to the differential expression analysis, the heterogeneity between the 2 groups (HPV+ vs HPV-) of samples was analyzed using principal component analysis (PCA). The differentially expressed genes (DEGs) between HPV+ and HPV- SCCHN samples were analyzed using the R edgeR package.^[16] The original count expression data were normalized using trimmed mean of M values (TMM). The screening thresholds for DEGs were false discovery rate < 0.05 and \log_2 fold change > 1.5, and the results were visualized through volcano plot.

2.3. Function and pathway enrichment analyses of DEGs

The Gene Ontology (GO) functional annotations, including biological process (BP), molecular function (MF) and cellular component (CC), and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched by the DEGs were analyzed using DAVID^[17] (version 6.8, <https://david-d.ncicrf.gov/>). Results were visualized using R language. false discovery rate < 0.05 was used as the threshold.

2.4. Weighed gene coexpression network analysis

Weighed gene coexpression network analysis (WGCNA) is a type of bioinformatics algorithm for coexpression network construction, which is used to identify modules associated with disease, and to select important pathogenic mechanisms and potential therapeutic targets.^[18]

Briefly, the original count expression matrix was normalized with $\log_2(x + 1)$, and the genes with median absolute deviation (MAD) of the top 75% and MAD > 0.01 were screened. The missing value was detected and processed. After pretreatment, the obtained matrix was analyzed by WGCNA.

Soft threshold is a very important part of WGCNA algorithm. The selection principle of soft threshold is to make the constructed network more consistent with scale-free network characteristics. Before screening of soft threshold, the outliers were detected and removed. The “type” of soft threshold screening was “signed,” and “corType” was “bicolor.”

The adjacent matrix and topology matrix were obtained based on the soft threshold. After whether the memory network under the selected soft threshold approximated scale free was tested, the dissimilarity between genes was used to cluster genes in the obtained topology matrix. Then the tree was cut into different modules (minimum number of genes in module was 30) using the dynamic shear method, and the similar modules (module heterogeneity > 0.8) were combined to obtain the final clustering module.

After WGCNA, the HPV status (HPV+ or HPV-) was used as the external biological parameter to conduct correlation study with each module, and the module most relevant to the external biological parameter (HPV+ or HPV-) was selected as the key module for further analysis. Subsequently, the hub genes were further screened using the 3 criteria as follows: significance of gene and specified module > 0.2, significant module membership value > 0.8, and q-weighted < 0.01. The hub genes were further subjected to function and pathway enrichment analyses using DAVID.^[19]

2.5. Prognostic gene selection by survival analysis

The union genes of the significant DEGs and the hub genes of WGCNA were selected for the prognostic correlation analysis of patients with HPV+. For each gene in the union set, the expression value after $\log_2(\text{count} + 1)$ conversion was calculated, and R survminer was used to calculate the optimal cutoff (>optimal cutoff as high expression, and <optimal cutoff as low expression). Survival analysis of survival time and survival status was conducted based on the survival package of R language. Genes with $P < .05$ in survival analysis were considered to be significantly associated with prognosis.

2.6. Univariate regression analysis and lasso cox regression analysis

The survival package coxph function was used for univariate regression analysis of the prognostic genes, and $P < .05$ was considered as the significant result. Then based on the obtained univariate regression analysis results, lasso cox regression analysis was carried out using the glmnet package to obtain the optimal

model that could be used for predicting the survival status of HPV + SCCHN. After obtaining the model, the risk score was calculated based on the model coefficients, and the reliability of the model was verified in the original data by using boxplot, receiver operating characteristic (ROC) efficacy analysis and survival analysis.

2.7. Model validation through external biological data

To further verify the prognostic guidance of the obtained model for HPV + SCCHN patients, we further verified the efficacy of the model using GSE65858 dataset from GEO database. After preprocessing such as reannotating and normalization of the original data, HPV + SCCHN patients were extracted from clinical information, and corresponding gene expression was obtained from the expression profile. The model obtained was put into the obtained risk score, and the boxplot, ROC efficacy verification, and survival analysis were used for verification.

2.8. Expression of model genes in different groups

To further investigate the expression significance of model genes, the expression of the model genes in 2 groups (HPV+ and HPV-) was analyzed, and the rank sum test was performed for the different groups. This analysis was performed in TCGA data and GEO data, respectively.

2.9. Drug interactions

The drug-gene interaction database (DGIdb) presents drug-gene interactions and gene druggability information from web resources, papers, and databases. In this study, DGIdb 2.0^[20] (<http://www.dgiddb.org/>) was used to predict the drug-gene interactions, and the parameter setting was consistent with the database default values. The results supported by literature were selected as reliable results in this analysis.

3. Results

3.1. DEG analysis

A total of 112 eligible SCCHN samples (33 HPV+ and 79 HPV-) were enrolled in this study. Then, PCA analysis was performed before differential expression analysis to determine the heterogeneity between HPV+ and HPV- SCCHN samples. Figure 1A shows that data could be well interpreted from 2 dimensions for the 112 SCCHN samples. Further 2-dimensional and 3-dimensional PCA revealed that the HPV+ and HPV- samples could be separated in the lower dimensions (Fig. 1B,C), indicating that heterogeneity existed between the 2 groups, which could be used for differential expression analysis. After differential analysis, a total of 350 significant DEGs were identified, including 199 significantly upregulated DEGs and 151 significantly downregulated DEGs (Fig. 1D).

3.2. Function and pathway enrichment analyses of DEGs

To investigate the potential functions of these DEGs, function and pathway enrichment analyses were carried out for these DEGs. Figure 2 presents that DEGs were significantly enriched in BP terms associated with keratinization, peptide cross-linking, and keratinocyte differentiation, CC terms associated with cornified envelope, and MF terms associated with structural molecule activity, and sequence-specific DNA binding. In addition, the pathway of neuroactive ligand-receptor interaction was significantly enriched.

3.3. WGCNA

To explore the interactions between these DEGs, we applied a system biology approach using WGCNA that could convert coexpression measure into connection weights and topology overlap measures.^[21] This analysis could screen the genes with MAD of the top 75% and MAD >0.01 from the original data. After the missing values were processed, an outlier occurred in 1 sample, and such outlier was removed from the following analysis (Fig. 3A).

Following data preprocessing, soft threshold was selected, and “12” was selected as the soft threshold for topology network construction (Fig. 3B). Then the soft threshold was verified to determine whether the memory network approximated scale free using the selected soft threshold. Figure 3C shows that k was negatively correlated with $p(k)$ (correlation coefficient: 0.91), indicating that the selected soft threshold could be used to establish gene scale-free network.

According to the eigenvector gene of the module, the module heterogeneity was calculated, and the module similarity of >0.8 was combined to obtain the scale-free network for subsequent analysis (Fig. 3D). Subsequently, HPV status (HPV+ or HPV-) was used as the external biological parameter to conduct correlation study with each module, and the result showed that there existed clusters between samples of the same character (HPV+ or HPV-; see Figure S1, Supplemental Digital Content, <http://links.lww.com/MD/L18>, showed the sample clustering diagram [top] and sample trait heat map [bottom]). It could also explain the heterogeneity of samples in different groups shown by PCA results. According to the correlation between trait and module feature vector gene as well as P value, the trait-related modules were mined. Figure 3E indicates that the light green module was the module most associated with HPV+ and HPV- status. Finally, we identified 72 hub genes associated with the external trait (HPV+/HPV-) from the light green module (see Figure S2, Supplemental Digital Content, <http://links.lww.com/MD/L19>, showed Bar chart and heat map of correlation between modules and samples). The connectivity of genes was measured by absolute value of the Pearson correlation. Genes with high within-module connectivity were considered as hub genes of the modules ($\text{cor.geneModuleMembership} = 0.92$; see Figure S3, Supplemental Digital Content, <http://links.lww.com/MD/L20>, showed The direct relationship between connectivity within the light green module and module membership [left] and hub gene screening [right]).

3.4. Functional and pathway enrichment analyses of hub genes

We further explored the functions of these hub genes through GO and pathway analysis. These hub genes were significantly associated with BP terms of DNA replication, DNA repair, and DNA replication, CC terms of nucleoplasm, nucleus, and condensed chromosome kinetochore as well as MF terms of protein binding, protein binding, and protein binding. Additionally, these hub genes participated in pathways of protein binding, mismatch repair, and protein binding. See Figure S4, Supplemental Digital Content, <http://links.lww.com/MD/L21>, showed GO functional annotations including BP, MF, and CC, as well as KEGG pathways enriched by the hub genes.

3.5. Survival analysis and regression model construction

The obtained hub genes were combined with DEGs to perform the survival analysis, and 422 union genes were obtained (Fig. 4), among which 65 genes were significantly associated with survival of HPV + SCCHN (see Table S1, Supplemental Digital Content, <http://links.lww.com/MD/L16>, showed the 65 genes significantly associated with survival of

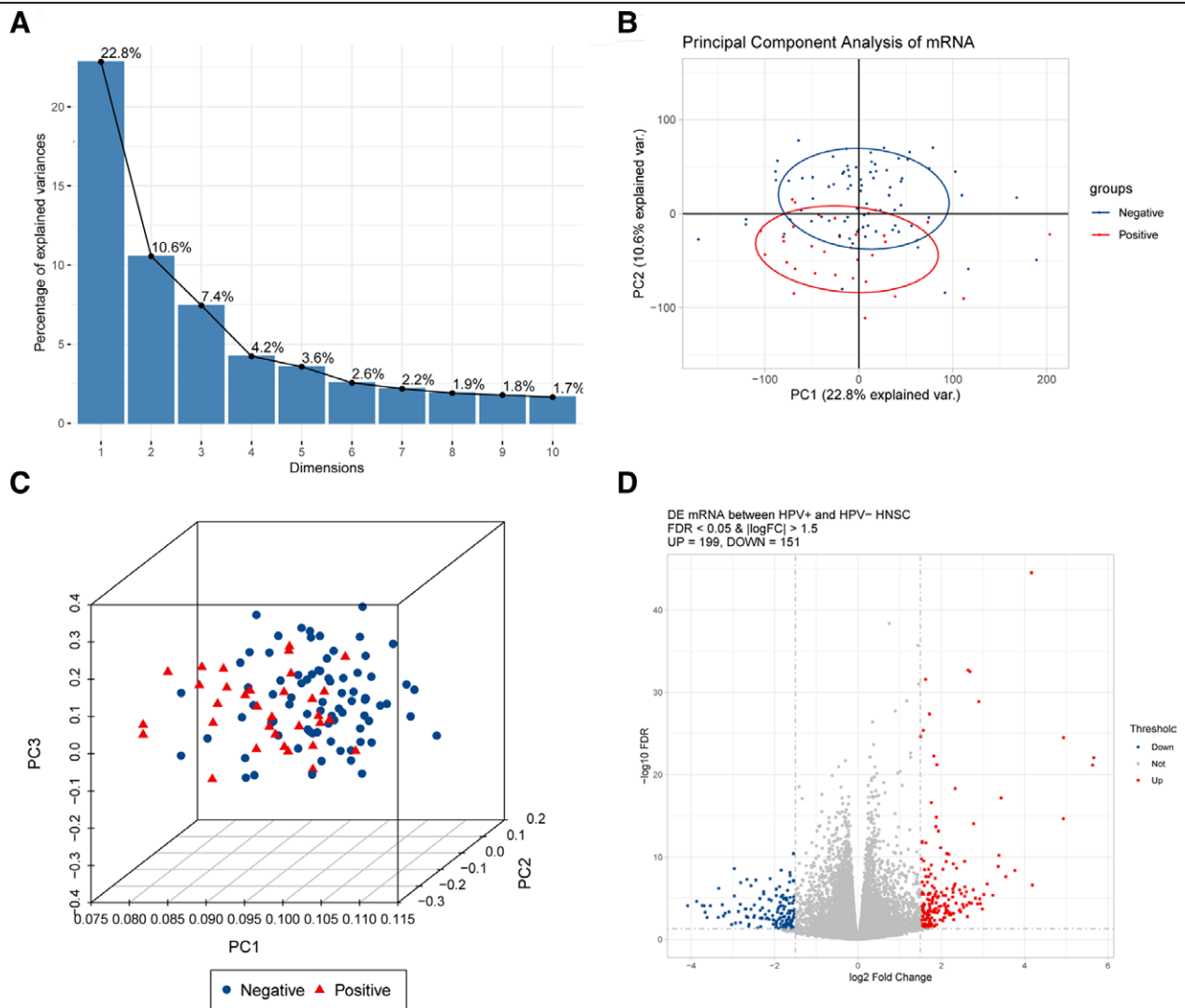


Figure 1. (A–C) PCA of samples of HPV+ and HPV– squamous cell carcinoma of the head and neck; (D) Volcano plot of differentially expressed genes. Red is the significantly upregulated differentially expressed genes, and blue is the significantly downregulated differentially expressed genes. PCA = principal component analysis.

DAVID 6.8 Enrichment results of DE mRNAs

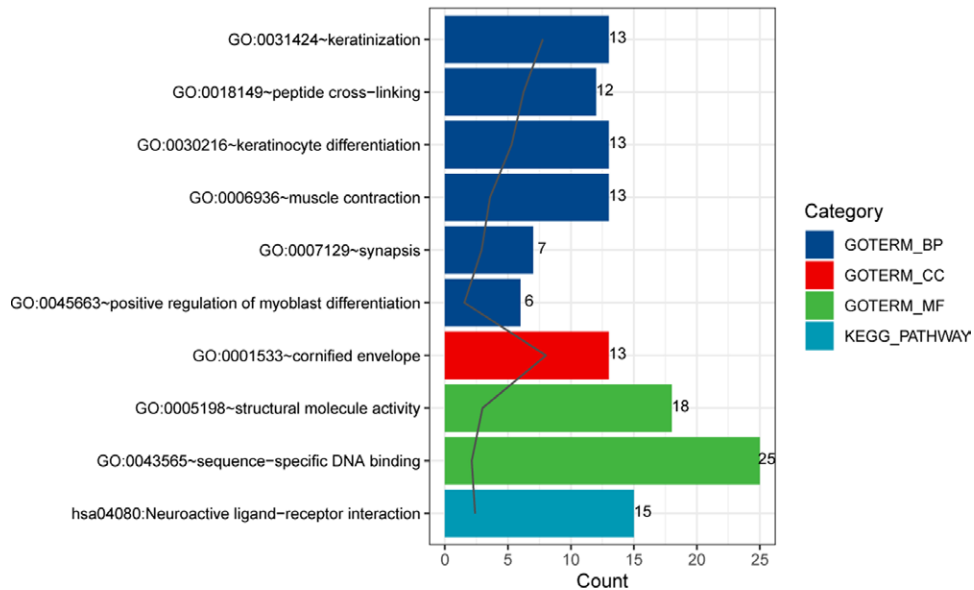


Figure 2. GO functional annotations including BP, MF, and CC, as well as KEGG pathways enriched by the differentially expressed genes. BP = biological process, CC = cellular component, GO = Gene Ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes, MF = molecular function.

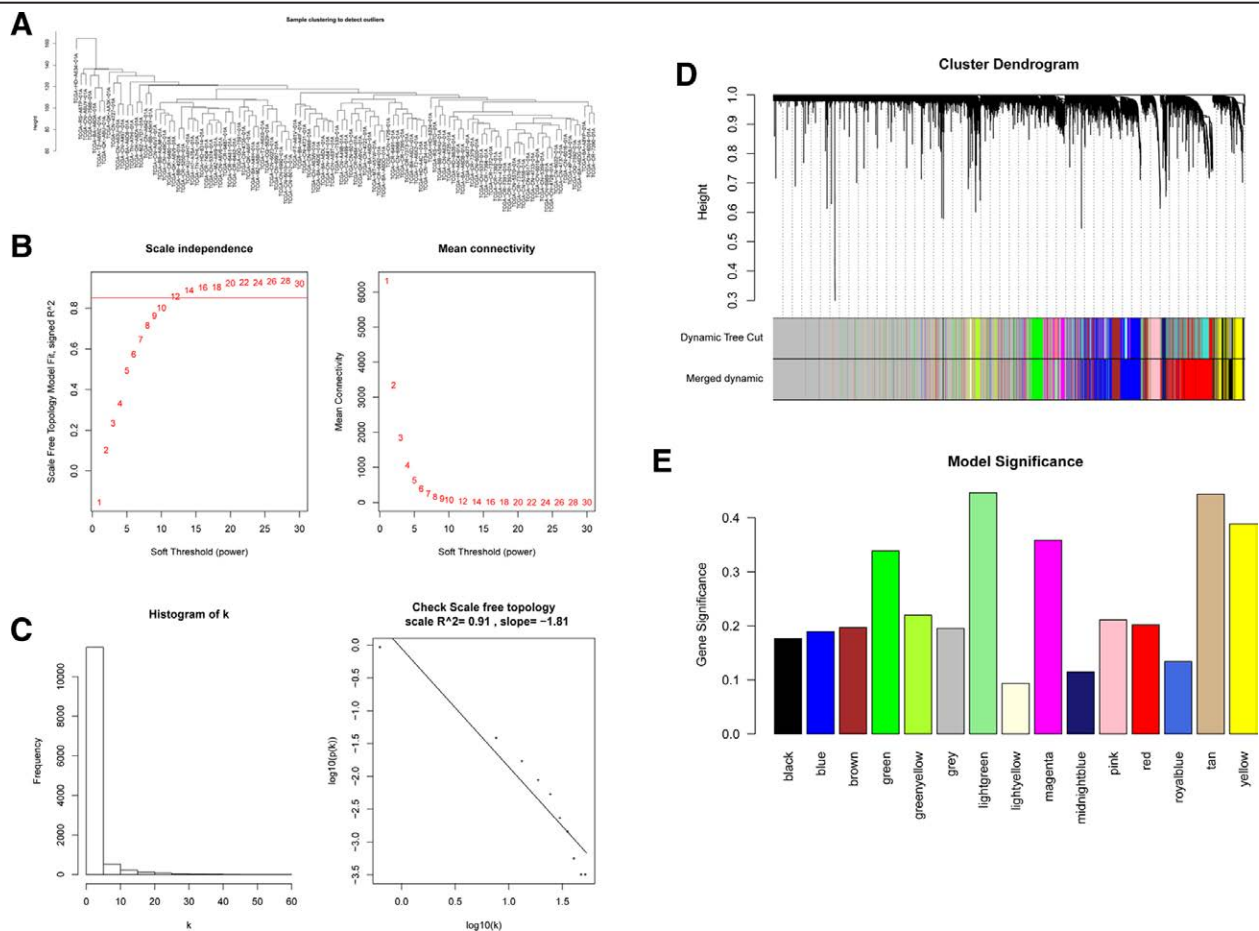


Figure 3. (A) Sample outlier detection; (B) Soft threshold selection. The horizontal axis of the figure on the left is soft threshold (power), and the vertical axis is the evaluation parameter of scale-free network (the higher the value, the more scale-free the network). On the right, the soft threshold and average connectivity approach 0. (C) Soft threshold testing (k is negatively correlated with $p(k)$ (correlation coefficient 0.91)). (D) A scale-free network combining modules of similarity > 0.8 . (E) Model significance. The module with the greatest significance is most associated with the trait.

HPV + SCCHN). Based on these prognosis-associated genes, regression analysis was conducted using the prognostic risk model. A total of 11 significant genes were identified through univariate regression analysis of the 65 genes (see Table S2, Supplemental Digital Content, <http://links.lww.com/MD/L17>, showed significant results of univariate regression analysis). Then lasso Cox regression analysis was performed to determine the optimal model (Fig. 5A). Finally, an HPV+-related prognostic model was obtained, which consisted of 8 genes: *Clorf105*, glycoprotein hormones, alpha polypeptide (*CGA*), cholinergic receptor nicotinic alpha 2 subunit (*CHRNA2*), cysteine rich protein 3 (*CRIP3*), cancer/testis antigen 2 (*CTAG2*), ectonucleotide pyrophosphatase/phosphodiesterase 6 (*ENPP6*), neurofilament heavy (*NEFH*), and ring finger protein 212 (*RNF212*). The obtained regression model could be described using the equation as follows: risk score = $0.065 \times Clorf105 + 0.012 \times CGA + 0.01 \times CHRNA2 + 0.047 \times CRIP3 + 0.043 \times CTAG2 - 0.034 \times ENPP6 - 0.003 \times NEFH - 0.068 \times RNF212$.

3.6. Model validation

To verify the prognostic effect of regression model in SCCHN, survival analysis and ROC efficacy were verified based on the risk score in the original dataset. Figure 5B shows that the prognostic model of the 8 HPV-related genes could provide good prognostic guidance for SCCHN. The survival analysis showed that the risk score based on 8 genes was highly significant

(Fig. 5C). Additionally, grouped boxplot and ROC efficacy analyses showed that the prognostic model composed of 8 genes could significantly predict the prognosis for SCCHN patients in the dataset of GSE65858 (Fig. 5D). The survival analysis indicated that the risk score based on 8 genes was significant in the dataset of GSE65858 (Fig. 5E).

3.7. Expression of model genes in different groups

After the predictive effect of prognostic model was confirmed, we investigated the expressions of the 4 model genes in HPV+ and HPV- groups. Figure 6A shows that the expressions of *Clorf105* ($P = .006$), *CHRNA2* ($P < .001$), *CRIP3* ($P < .001$), *ENPP6* ($P < .001$), *NEFH* ($P < .001$), and *RNF212* ($P < .001$) were significantly different between the HPV+ and HPV- groups in TCGA dataset. The expressions of the other 2 genes were not significantly different between the 2 groups. Moreover, the expressions of *CTAG2* ($P = .048$), *ENPP6* ($P = .013$), *NEFH* ($P = .028$), and *RNF212* ($P < .001$) were significantly different between the HPV+ and HPV- groups in the GSE65858 database (Fig. 6B).

3.8. Drug interaction analysis

Among the 4 models of genes, *CGA* interacted with 3 drugs, including staurosporine, testosterone, and vinblastine. Staurosporine is a broad-specific protein kinase inhibitor, which

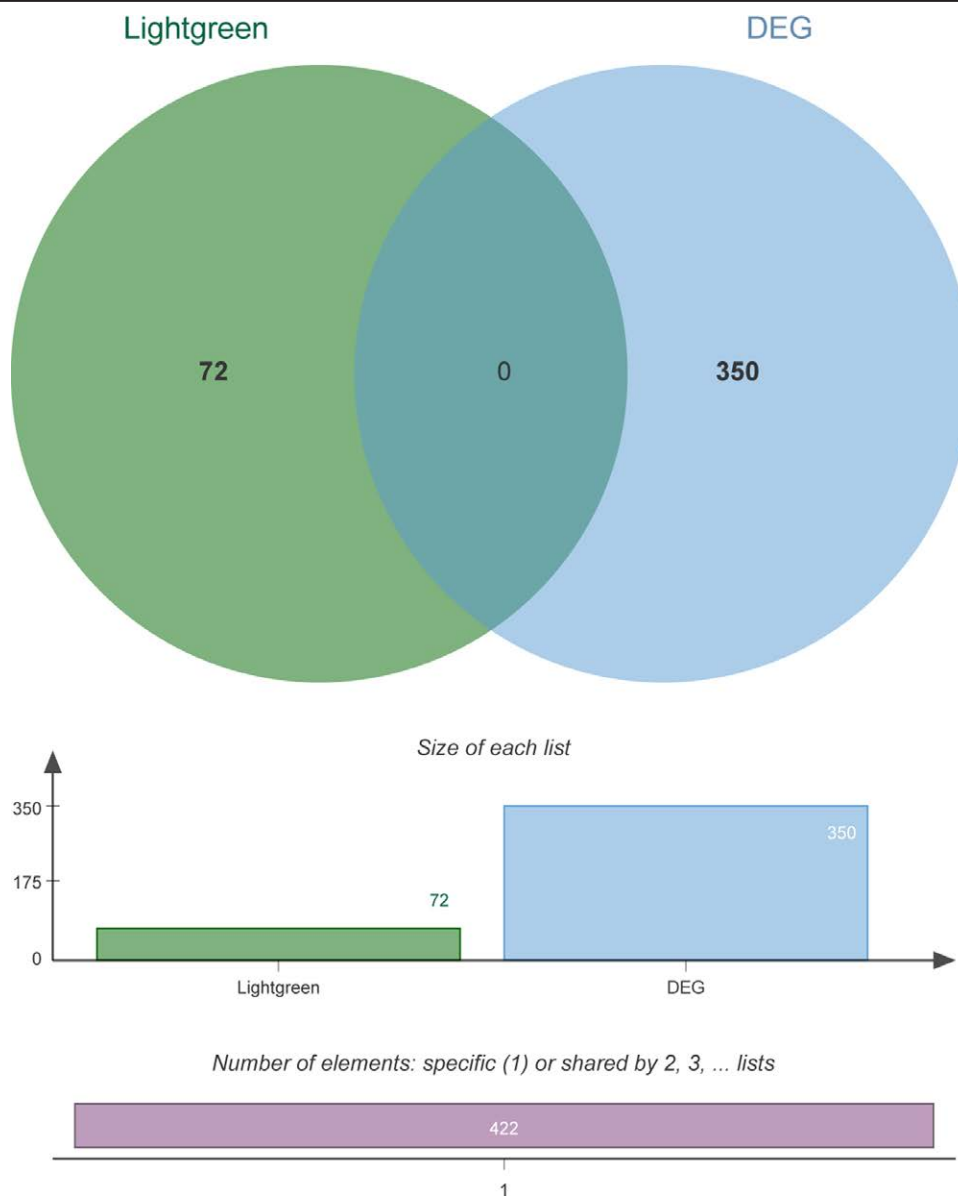


Figure 4. Venn diagram of differentially expressed genes and hub genes.

possesses strong antitumor activity.^[22] Testosterone is the principal male sex hormone, which is widely used in the treatment of prostate cancer.^[23] Vinblastine can inhibit the aggregation of tubulin in cancer cells, interfere with the formation of spindle in proliferating cells, block the mitosis in the middle stage, and have immunosuppressive effects.^[24] *CHRNA2* interacted with 11 drugs, such as procaine, doxacurium chloride, and mecamlamine hydrochloride (Fig. 7).

4. Discussion

SCCHN ranks the 6th most common malignancy worldwide.^[7] Despite the advances in various treatments, such as surgery, chemotherapy, radiotherapy, and immune checkpoint inhibitors, the overall survival rates of SCCHN patients remain unsatisfactory.^[25] Therefore, identification of new promising prognostic biomarkers is essential for effective therapies in SCCHN. In the present study, we took advantage of the high-throughput data from TCGA for biomarker analysis of SCCHN. A prognosis model consisting of 8 HPV-associated

RNAs was identified, which was proved to effectively and reliably predict the prognosis of SCCHN. Among the 8 genes in the prognostic model, *ENPP6*, *NEFH*, and *RNF212* were significantly differentially expressed between the HPV+ and HPV- SCCHN groups in both TCGA and GSE65858 datasets. Furthermore, *CGA* and *CHRNA2* were found to interact with several drugs.

In this study, the DEGs between HPV+ and HPV- SCCHN groups were the most significantly associated with function of keratinization. Keratinization is a histologic feature of hematoxylin-eosin staining, which is associated with the poor prognosis of head and neck cancer.^[26] Some recent studies have suggested that nonkeratinizing tumors are strongly associated with HPV positivity in patients with oropharyngeal squamous cell carcinoma.^[27,28] On the contrary, HPV-negative oropharyngeal squamous cell carcinoma is associated with keratinization.^[29,30] Therefore, the functional enrichment of keratinization could reflect the reliability of our analysis.

Among *ENPP6*, *NEFH*, and *RNF212*, the roles of *ENPP6* and *RNF212* have been rarely reported to be associated with

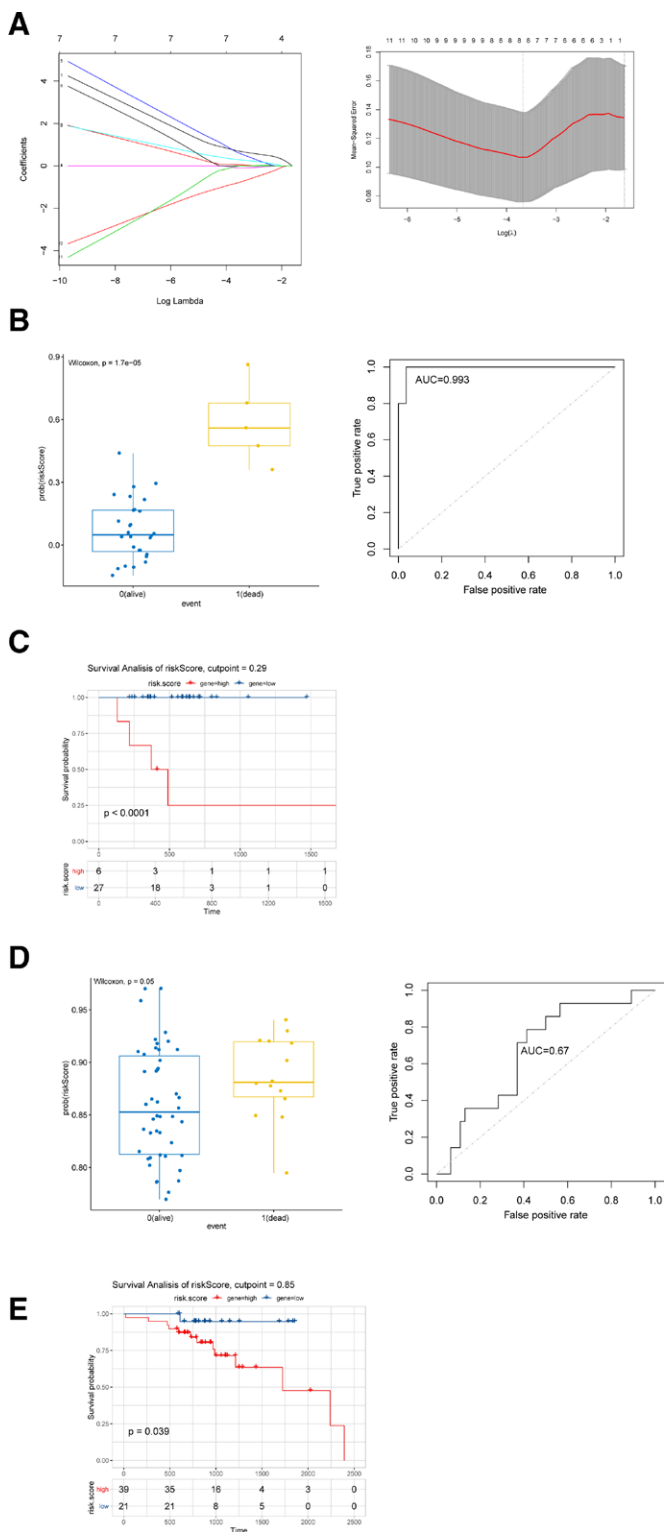


Figure 5. (A) The results of lasso cox regression analysis of 11 genes were calculated by R language (lambda.min = 0.02547615; lambda.1se = 0.19725256). (B) Grouping calculations and ROC efficacy analysis based on prognostic model calculations in TCGA dataset. (C) Survival analysis of HPV+ patients with squamous cell carcinoma of the head and neck by risk score based on prognostic model calculations in TCGA dataset. (D) Grouping calculations and ROC efficacy analysis based on prognostic model calculations with risk score in GSE65858 dataset. (E) Survival analysis of HPV+ patients with squamous cell carcinoma of the head and neck by risk score based on prognostic model calculations in GSE65858 dataset. HPV = human papillomavirus, ROC = receiver operating characteristic, TCGA = The Cancer Genome Atlas.

human cancers. *NEFH* encodes neurofilament heavy chain, which is one of the major components of the neuronal cytoskeleton neurofilaments.^[31] Loss of *NEFH* leads to active repression of mitochondrial metabolism and biogenesis in esophageal squamous cell carcinoma cells. Therefore, *NEFH* may act as a tumor suppressor gene in esophageal squamous cell carcinoma.^[32] Recently, promoter CpG island methylation of *NEFH* has been reported in clear cell renal cell cancer, and it is related to the poor progression-free survival.^[33] Given their significant differential expressions between the HPV+ and HPV- SCCHN groups, we speculated that these genes served as prognostic molecular targets of HPV-associated SCCHN.

CGA and *CHRNA2* were found to interact with several drugs. *CGA* encodes the common alpha subunit of 4 glycoprotein hormones, including luteinizing hormone, human chorionic gonadotropin, thyroid-stimulating hormone, and follicle-stimulating hormone. Previous study has noted that the increased expression of the alpha subunit is present in the sera of half of the patients with malignant islet cell tumors, suggesting that the alpha subunit is a specific marker for malignancy.^[34] *CGA* is identified as a new estrogen receptor alpha (ER α)-responsive gene in human breast cancer cell.^[35] Its role in SCCHN has not been reported. Interestingly, ER is considered to be involved in tumorigenesis of squamous cell carcinoma of the esophagus.^[36] Importantly, *CGA* and *CHRNA2* were significantly enriched in the pathway of neuroactive ligand-receptor interaction. Neuroactive ligand-receptor interaction has been involved in various types of squamous carcinoma, such as tongue squamous cell carcinoma, lung squamous cell carcinoma, and oral squamous cell carcinoma.^[37-39] Given their involvement in neuroactive ligand-receptor interaction, we speculated that *CGA* and *CHRNA2* served as prognostic biomarkers in HPV-associated SCCHN.

CGA was predicted to interact with 3 drugs of staurosporine, testosterone, and vinblastine. Staurosporine is a broad-specific protein kinase inhibitor, which has been reported to induce apoptosis in oral squamous cell carcinoma.^[22] Study has reported that testosterone has some therapeutic effects on squamous cell carcinoma.^[40] Vinblastine is a tubulin-targeting *Vinca* alkaloids, which is responsible for many chemotherapeutic successes as antitumor drugs.^[24] *CHRNA2* interacted with 11 drugs, such as procaine. Procaine is a DNA-demethylating agent that has growth-inhibitory effects on cancer cells, causing mitotic arrest.^[41] These drugs might be candidate agents for the treatment of HPV-associated SCCHN.

Although the obtained prognostic model was verified in this study, its exact role in clinical practice still remains largely unexplored. Therefore, it is necessary to further confirm our results through animal and clinical experiments.

5. Conclusion

Our study revealed an 8 HPV-related RNA signature for prognosis of patients with SCCHN. Such prognostic model might serve as a possible candidate biomarker and therapeutic target for HPV-associated SCCHN.

Author contributions

- Conceptualization:** Chuanxin Wang.
- Data curation:** Zhengdong Luo, Shupeng Liu.
- Formal analysis:** Mei Zhang.
- Funding acquisition:** Lei Wang.
- Methodology:** Mei Zhang, Lei Wang.
- Writing – original draft:** Mei Zhang, Lei Wang.
- Writing – review & editing:** Xin Zhang, Chuanxin Wang.

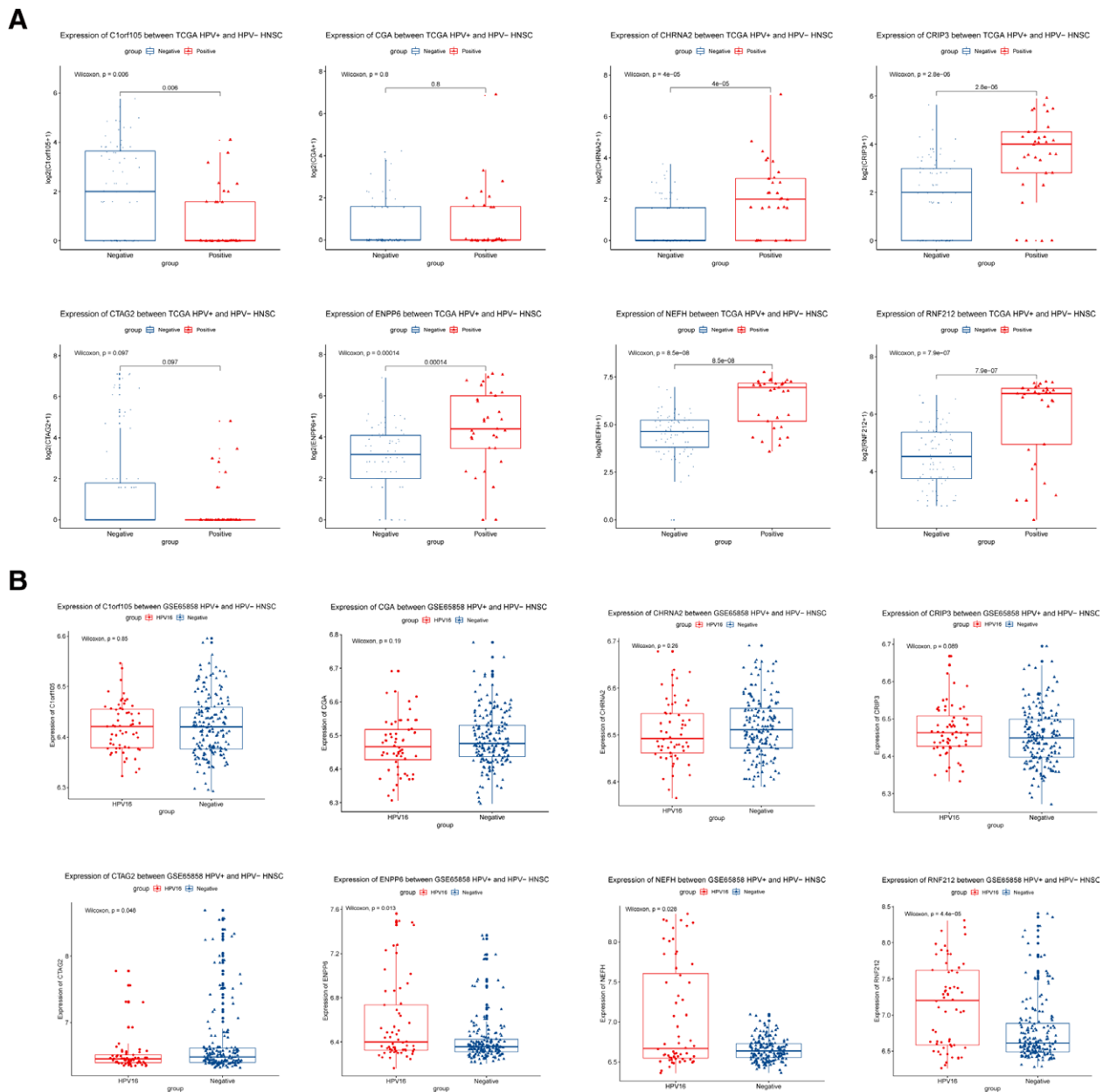


Figure 6. (A) Expression of 8 model genes in different subgroups (HPV+ and HPV-) of squamous cell carcinoma of the head and neck ([A] TCGA, [B] GSE65858). HPV = human papillomavirus, TCGA = The Cancer Genome Atlas.

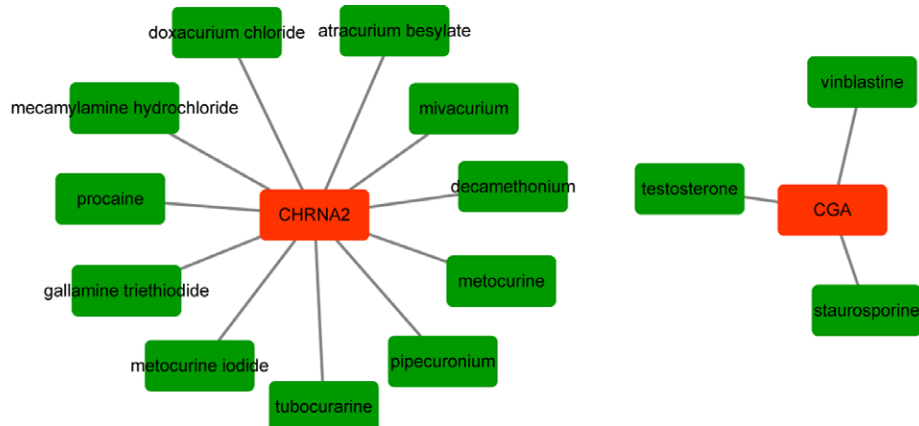


Figure 7. The constructed drug-gene interaction network. Red represents gene and green represents drug.

References

- [1] Ferlay J, et al. Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon (France): International Agency for Research on Cancer. 2013. GLOBOCAN 2012. Version 1.0.[cited 2016 July 11].
- [2] Network, C.G.A. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*. 2015;517:576.
- [3] Pignon J-P, le Maître A, Maillard E, et al. MACH-NC Collaborative Group. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): an update on 93 randomised trials and 17,346 patients. *Radiother Oncol*. 2009;92:4–14.
- [4] Bernier J, Dörmagen C, Ozsahin M, et al. European Organization for Research and Treatment of Cancer Trial 22931. Postoperative irradiation with or without concomitant chemotherapy for locally advanced head and neck cancer. *N Engl J Med*. 2004;350:1945–52.
- [5] Marur S, D'Souza G, Westra WH, et al. HPV-associated head and neck cancer: a virus-related cancer epidemic. *Lancet Oncol*. 2010;11:781–9.
- [6] Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med*. 2010;363:24–35.
- [7] Qian X, Nguyen DT, Dong Y, et al. Prognostic score predicts survival in HPV-negative head and neck squamous cell cancer patients. *Int J Biol Sci*. 2019;15:1336–44.
- [8] Davis-Dusenbery BN, Wu C, Hata A. Micromanaging vascular smooth muscle cell differentiation and phenotypic modulation. *Arterioscler Thromb Vasc Biol*. 2011;31:2370–7.
- [9] Rensen S, Doevendans P, Van Eys G. Regulation and characteristics of vascular smooth muscle cell phenotypic diversity. *Neth Heart J*. 2007;15:100–8.
- [10] Owens GK, Kumar MS, Wamhoff BR. Molecular regulation of vascular smooth muscle cell differentiation in development and disease. *Physiol Rev*. 2004;84:767–801.
- [11] McDonald RA, Hata A, MacLean MR, et al. MicroRNA and vascular remodelling in acute vascular injury and pulmonary vascular remodelling. *Cardiovasc Res*. 2011;93:594–604.
- [12] Network, C.G.A.R. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008;455:1061.
- [13] Shi H, Chen J, Li Y, et al. Identification of a six microRNA signature as a novel potential prognostic biomarker in patients with head and neck squamous cell carcinoma. *Oncotarget*. 2016;7:21579–90.
- [14] Wong N, Khwaja SS, Baker CM, et al. Prognostic micro RNA signatures derived from The Cancer Genome Atlas for head and neck squamous cell carcinomas. *Cancer Med*. 2016;5:1619–28.
- [15] Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)*. 2015;19:A68–77.
- [16] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–40.
- [17] Dennis G, Sherman BT, Hosack DA, et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol*. 2003;4:R60.
- [18] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf*. 2008;9:559.
- [19] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4:44–57.
- [20] Wagner AH, Coffman AC, Ainscough BJ, et al. DGIdb 20: mining clinically relevant drug–gene interactions. *Nucleic Acids Res*. 2015;44:D1036–44.
- [21] Horvath S, Dong J. Geometric interpretation of gene coexpression network analysis. *PLoS Comput Biol*. 2008;4:e1000117.
- [22] Abiko Y, Arai J, Mitamura J, et al. Alteration of proto-oncogenes during apoptosis in the oral squamous cell carcinoma cell line, SAS, induced by staurosporine. *Cancer Lett*. 1997;118:101–7.
- [23] Kaplan AL, Hu JC, Morgentaler A, et al. Testosterone therapy in men with prostate cancer. *Eur Urol*. 2016;69:894–903.
- [24] Rongying H, et al. Vinblastine enhances the expression of tumor resistance gene *abcb4* in zebrafish. *Basic Clin Med*. 2017;37:758–62.
- [25] Leemans CR, Snijders PJ, Brakenhoff RH. The molecular landscape of head and neck cancer. *Nat Rev Cancer*. 2018;18:269.
- [26] Cooper T, Biron VL, Adam B, et al. Association of keratinization with 5-year disease-specific survival in oropharyngeal squamous cell carcinoma. *JAMA Otolaryngol Head Neck Surg*. 2015;141:250–6.
- [27] Chernock RD. Morphologic features of conventional squamous cell carcinoma of the oropharynx: 'keratinizing' and 'nonkeratinizing' histologic types as the basis for a consistent classification system. *Head Neck Pathol*. 2012;6:41–7.
- [28] Chernock RD, El-Mofty SK, Thorstad WL, et al. HPV-related non-keratinizing squamous cell carcinoma of the oropharynx: utility of microscopic features in predicting patient outcome. *Head Neck Pathol*. 2009;3:186–94.
- [29] Cardesa A, Nadal A. Carcinoma of the head and neck in the HPV era. *Acta Dermatovenerol Alp Pannonica et Adriat*. 2011;20:161–73.
- [30] Wilczynski SP, Lin BT, Xie Y, et al. Detection of human papillomavirus DNA and oncoprotein overexpression are associated with distinct morphological patterns of tonsillar squamous cell carcinoma. *Am J Pathol*. 1998;152:145–56.
- [31] Lee MK, Cleveland DW. Neuronal intermediate filaments. *Annu Rev Neurosci*. 1996;19:187–217.
- [32] Kim MS, Chang X, LeBron C, et al. Neurofilament heavy polypeptide regulates the Akt- β -catenin pathway in human esophageal squamous cell carcinoma. *PLoS One*. 2010;5:e9003.
- [33] Dubrowskaja N, Gebauer K, Peters I, et al. Neurofilament heavy polypeptide CpG island methylation associates with prognosis of renal cell carcinoma and prediction of antivascular endothelial growth factor therapy response. *Cancer Med*. 2014;3:300–9.
- [34] Kahn CR, Rosen SW, Weintraub BD, et al. Ectopic production of chorionic gonadotropin and its subunits by islet-cell tumors: a specific marker for malignancy. *N Engl J Med*. 1977;297:565–9.
- [35] Bièche I, Parfait B, Le Doussal V, et al. Identification of CGA as a novel estrogen receptor-responsive gene in breast cancer: an outstanding candidate marker to predict the response to endocrine therapy. *Cancer Res*. 2001;61:1652–8.
- [36] Utsumi Y, Nakamura T, Nagasue N, et al. Role of estrogen receptors in the growth of human esophageal carcinoma. *Cancer*. 1989;64:88–93.
- [37] Zhang H, Liu J, Fu X, et al. Identification of key genes and pathways in tongue squamous cell carcinoma using bioinformatics analysis. *Med Sci Monit*. 2017;23:5924.
- [38] Zhang G, Bi M, Li S, et al. Determination of core pathways for oral squamous cell carcinoma via the method of attract. *J Cancer Res Ther*. 2018;14:1029.
- [39] Chen WJ, Gan TQ, Qin H, et al. Implication of downregulation and prospective pathway signaling of microRNA-375 in lung squamous cell carcinoma. *Pathol Res Pract*. 2017;213:364–72.
- [40] Kato Y, Ozono S, Koshika S. Testosterone metabolism in new squamous cell carcinoma cell line (RSS18) from 7, 12-dimethylbenz [a] anthracene-induced submandibular gland of female rat. *J Steroid Biochem Mol Biol*. 1996;57:349–55.
- [41] Villar-Garea A, Fraga MF, Espada J, et al. Procaine is a DNA-demethylating agent with growth-inhibitory effects in human cancer cells. *Cancer Res*. 2003;63:4984–9.