

Mining of gene modules and identification of key genes in head and neck squamous cell carcinoma based on gene co-expression network analysis

Qian Zhao, PhD^a, Yan Zhang, MD^a, Xue Zhang, MD^a, Yeqing Sun, PhD^{b,*}, Zhengkui Lin, PhD^{a,*}

Abstract

To explore the gene modules and key genes of head and neck squamous cell carcinoma (HNSCC), a bioinformatics algorithm based on the gene co-expression network analysis was proposed in this study.

Firstly, differentially expressed genes (DEGs) were identified and a gene co-expression network (i-GCN) was constructed with Pearson correlation analysis. Then, the gene modules were identified with 5 different community detection algorithms, and the correlation analysis between gene modules and clinical indicators was performed. Gene Ontology (GO) analysis was used to annotate the biological pathways of the gene modules. Then, the key genes were identified with 2 methods, gene significance (GS) and PageRank algorithm. Moreover, we used the Disgenet database to search the related diseases of the key genes. Lastly, the online software onclnc was used to perform the survival analysis on the key genes and draw survival curves.

There were 2600 up-regulated and 1547 down-regulated genes identified in HNSCC. An i-GCN was constructed with Pearson correlation analysis. Then, the i-GCN was divided into 9 gene modules. The result of association analysis showed that, sex was mainly related to mitosis and meiosis processes, event was mainly related to responding to interferons, viruses and T cell differentiation processes, T stage was mainly related to muscle development and contraction, regulation of protein transport activity processes, N stage was mainly related to mitosis and meiosis processes, while M stage was mainly related to responding to interferons and immune response processes. Lastly, 34 key genes were identified, such as CDKN2A, HOXA1, CDC7, PPL, EVPL, PXN, PDGFRB, CALD1, and NUSAP1. Among them, HOXA1, PXN, and NUSAP1 were negatively correlated with the survival prognosis.

HOXA1, PXN, and NUSAP1 might play important roles in the progression of HNSCC and served as potential biomarkers for future diagnosis.

Abbreviations: DEG = differentially expressed gene, GS = gene significance, HNSCC = head and neck squamous cell carcinoma, HOXA1 = Homeobox A1, i-GCN = gene co-expression network, ME = module eigengene, NuSAP1 = Nucleolar spindle associated protein 1, PCA = principal component analysis, TCGA = the cancer genome atlas, WGCNA = Weighted Gene Co-expression Network Analysis.

Keywords: gene co-expression network, HNSCC, key genes

Editor: Lu Wang.

This work was supported by the National Science Foundation of China [Grant No. 31770918], the Strategic Priority Research Program of the Chinese Academy of Sciences [Grant No. XDA04020202-12 and XDA04020412].

The authors declare that we don't have any financial or associative interest that represents a conflict of interest correlated with the work submitted.

Supplemental Digital Content is available for this article.

The datasets generated during and/or analyzed during the current study are publicly available.

^a College of Information Science and Technology, ^b Institute of Environmental System Biology, College of Environmental Science and Engineering, Dalian Maritime University, Dalian, China.

* Correspondence: Zhengkui Lin, Linghai Road, Dalian 116026, Liaoning, China (e-mail: zhengkui@dlmu.edu.cn).

Copyright © 2020 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-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

How to cite this article: Zhao Q, Zhang Y, Zhang X, Sun Y, Lin Z. Mining of gene modules and identification of key genes in head and neck squamous cell carcinoma based on gene co-expression network analysis. *Medicine* 2020;99:49 (e22655).

Received: 15 April 2020 / Received in final form: 1 September 2020 / Accepted: 10 September 2020

<http://dx.doi.org/10.1097/MD.00000000000022655>

Highlights

1. A gene co-expression network of HNSCC was constructed
2. The gene modules which were highly related to five clinical indicators sex, event, T, N, and M was identified
3. The key genes which played important roles in HNSCC were identified, and HOXA1, PXN and NUSAP1 were negatively correlated with the survival prognosis.

1. Introduction

Head and neck squamous cell carcinoma (HNSCC) is one of the top 10 common cancer diseases in the world.^[1] In recent years, there have been many studies on HNSCC. Zou et al analyzed the expression profiles of lncRNA and miRNA in 422 cases of HNSCC, and found that 307 differential genes were related to the survival of the patients and were related to the mutations of gene CDKN2A, TP53, CASP8, etc.^[2] Besides, Yan et al found the hub genes CPBP, NF-AT1, and miR-1 from the TF-miRNA-gene network which was constructed by 2594 differentially expressed

genes (DEGs) and 25 miRNAs of HNSCC.^[3] The previous studies focused on identifying DEGs of HNSCC, but there is a lack of research on gene co-expression networks (i-GCNs) and identifying its functional gene modules and key genes.

I-GCN was first proposed by Butte and Kohane in 1999.^[4] They measured gene co-expression relationship with Pearson correlation coefficient as an index, and constructed an i-GCN with gene expression profile data.^[4,5] At present, the commonly used method in gene co-expression network analysis is Weighted Gene Co-expression Network Analysis (WGCNA).^[6] Xia et al identified the modules which were closely related to the stage and the hub genes TOP2A, TTK, CHEK1, and CENPA with WGCNA on the transcriptome data of adrenocortical carcinoma.^[7] Tang et al found 13 hub genes associated with breast cancer prognosis, such as SSPN, NELL2, AGTR1 with WGCNA.^[8] Ao et al constructed a differential i-GCN between normal tissues and tumor tissues of papillary thyroid carcinoma and identified the potential functional genes with WGCNA.^[9] As a result, WGCNA has made important contributions to cancer research by constructing an i-GCN to identify modules which related to clinical indicators and hub genes. Based on this, we think there are 2 parts of this analysis that can be optimized. On the one hand, multiple algorithms can be used to divide the i-GCN into modules, and then the optimal division result can be selected. On the other hand, two methods can be used to find key genes. One is identifying key genes with the correlation of gene expression levels and clinical indicators. The other was calculating the importance of nodes in i-GCN based on the topological structure with the PageRank algorithm to identify the key genes.

To explore the gene modules and key genes of HNSCC, a bioinformatics algorithm based on the gene co-expression network analysis was proposed in this study. Firstly, gene expression data of HNSCC and its adjacent tissues was obtained from TCGA. The original data was preprocessed for further analysis. Then, the FC-t algorithm was used to identify DEGs. Secondly, we constructed the i-GCN with the Pearson correlation analysis. After that, five different community detection algorithms were used to divide the i-GCN. Then the algorithm with the highest modularity was selected to divide the i-GCN into gene modules for further analysis. Furthermore, we found the gene modules which were highly related to 5 clinical indicators sex, event, T, N, and M. The biological significance of each gene module was explored using GO enrichment analysis. Finally, the PageRank algorithm and gene significance (GS) both were used to identify key genes, which next to be search for related diseases in the Disgenet database. In the end, the survival analysis was performed on the key genes.

2. Methods and materials

To explore the gene modules and identify the key genes of HNSCC, a bioinformatics algorithm based on the gene co-expression network analysis was proposed in this study. And the transcriptome data of HNSCC and its adjacent tissues were used. Flow-chart of data analysis in this paper was shown in Fig. 1.

2.1. Data collection and pre-processing

The transcriptome data of HNSCC and its adjacent tissues which were used in this paper were downloading from the cancer

genome atlas (TCGA)^[10] (<https://cancergenome.nih.gov/>), which included 498 HNSCC samples and 44 adjacent tissues samples, each of which contains 60,483 genes. Of the 60,483 genes in each sample, we removed genes which expression below 1. Then, the filtered genes were used for hierarchical clustering of HNSCC. This study based on public sources data, which contains its ethnic approval. Thus, we do not need any further ethnic approval.

2.2. Identification of DEGs

We identified the DEGs between HNSCC and its adjacent tissues with FC-t algorithm^[11] in this study. $FC \geq 2$ or $FC \leq 0.5$ and $P < .05$ was set as the cut-off criteria.

2.3. Construction of i-GCN

The Pearson correlation coefficient and its P value between pairwise DEGs were calculated. Then the conditions $|r| \geq 0.6$ and $P < .05$ (r represents the Pearson correlation coefficient) was set as the cut-off criteria. An i-GCN was constructed based on this.

2.4. Community division of i-GCN

Five different community detection algorithms, multilevel,^[12] eigenvector,^[13] label-propagation,^[14] map-equation^[15,16] and edge-betweenness^[17] were used to divide the i-GCN to obtain communities (gene modules). Therefore, we installed igraph package^[18] in R (v1.2.4) for community detection using multilevel.community, leading.eigenvector.community, label.propagation.community, infomap.community, edge.betweenness.community. The modularity was carried out to determine the results of different algorithms. Then the algorithm with the highest modularity was selected to divide the i-GCN for further analysis.

2.5. Association analysis between gene modules and clinical indicators

To measure the association between the gene modules and clinical indicators, the principal component analysis (PCA)^[19] was carried out to perform the gene expression profiles in each module. Then the first principal component was used as the module eigengene (ME). Pearson correlation analysis was used to find the association matrix between the MEs of each module and five clinical indicators sex, event, T, N, and M, respectively.

2.6. GO enrichment analysis

To explore the biological significance of the gene modules, the genes contained in gene modules were enriched with the biological processes provided by the GO database (<http://geneontology.org/>). The 10 GO Terms with the smallest P value were selected for further research.

2.7. Identification of key genes

We used two methods to identify key genes. One was identifying key genes with the Pearson correlation analysis of gene expression levels and clinical indicators event. The Pearson correlation coefficient was defined as the GS, and the DEGs with $|GS| > 0.16$ were selected as key genes. The other was calculating

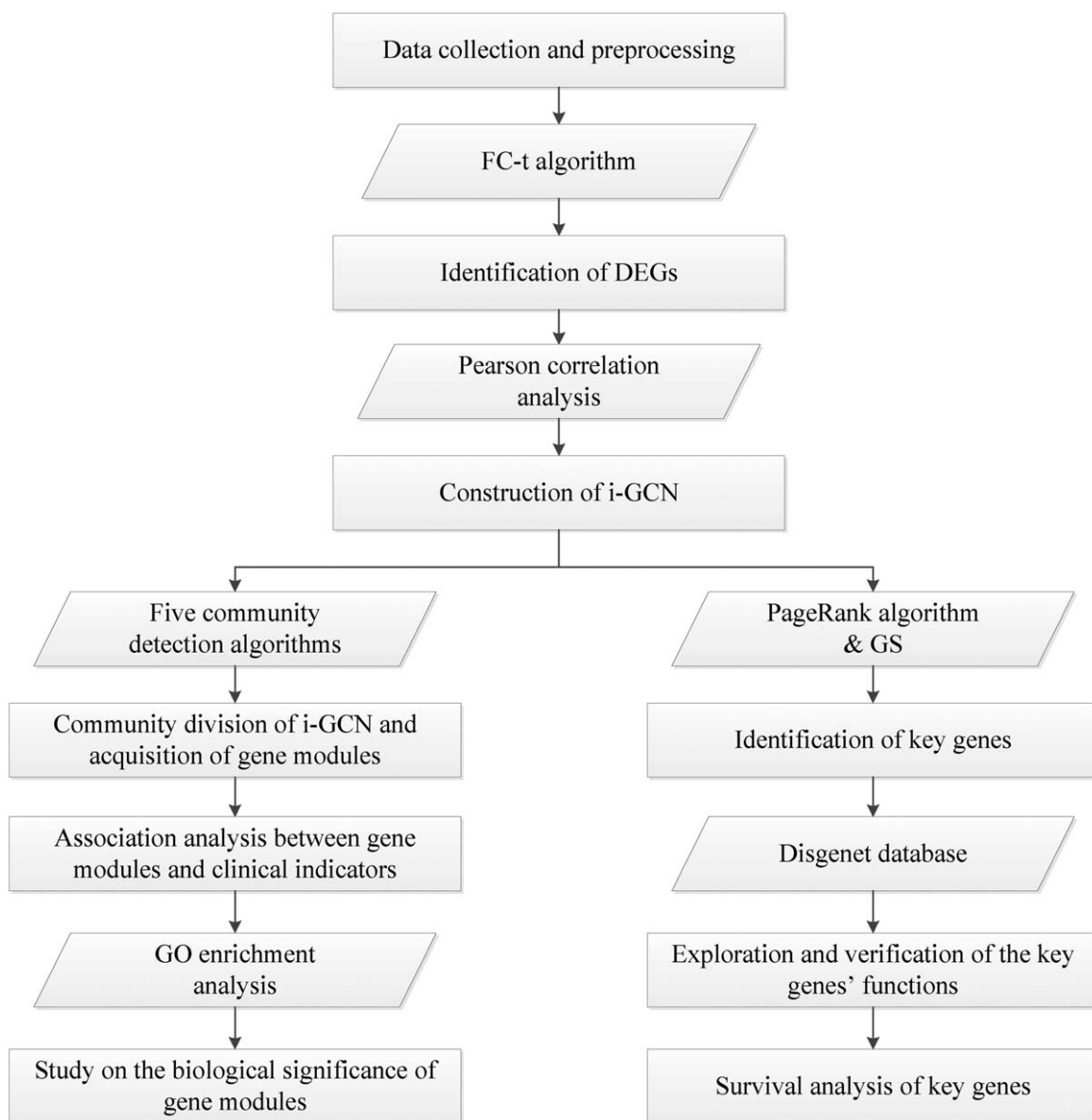


Figure 1. Flow-chart of data analysis in this paper. The rectangular boxes represent the processing steps, and the parallelogram boxes represent the method or database.

the importance of nodes in i-GCN based on the topological structure with the PageRank algorithm^[20] to identify the key genes. And the top 20 genes with the highest scores were selected as key genes.

2.8. Use of Disgenet database

Disgenet Database^[21] (<http://www.disgenet.org/>) contains the information between diseases and genes. We used the Disgenet database to search the related diseases of the key genes.

2.9. Survival analysis of key genes

The online software oncolnc^[22] (<http://www.oncolnc.org/>) was used to perform the survival analysis on the key genes and draw survival curves. Cancer was set to HNSCC and lower and upper percentiles were both set as 20.

3. Results

3.1. Data pre-processing

We removed genes which expression below 1. Then, the remained 18,510 genes were used for hierarchical clustering of HNSCC (Fig. 2A). From the Figure 2A, there were 2 obvious outlier samples in the original data: TCGA-D6-A6ES and TCGA-IQ-7631, which were removed to obtain a data set for further analysis. Details of the remaining genes and samples were presented in the Supplementary Table 1, <http://links.lww.com/MD/F274>, which contained the pre-processed data.

3.2. Identification of DEGs with FC-t algorithm

After data pre-processing, the remaining 18,510 genes were used for identifying the DEGs between HNSCC and its adjacent tissues with FC-t algorithm (Fig. 2B and Supplementary Table 2,

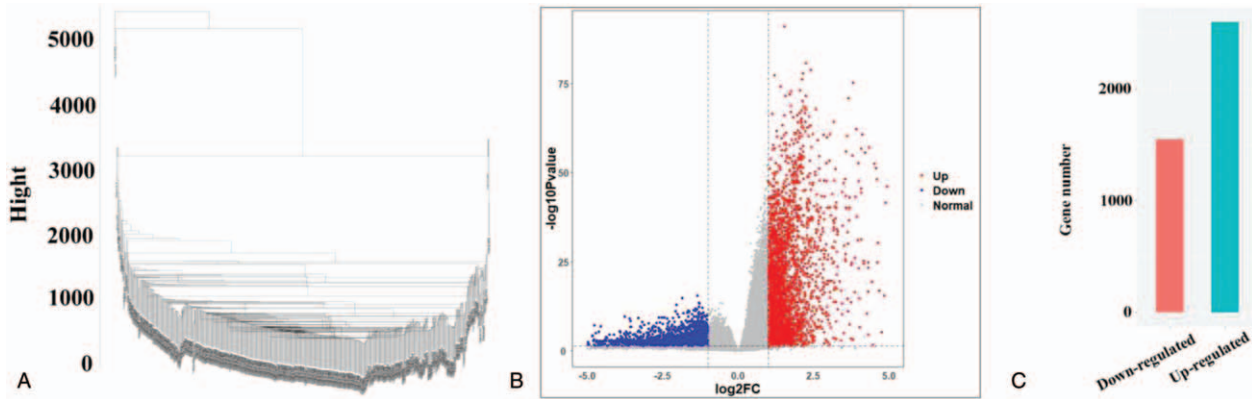


Figure 2. Data preprocessing and identification of DEG. (A) Sample clustering was conducted to detect outliers, while TCGA-D6-A6ES and TCGA-IQ-7631 were removed. (B) X-axis represents log₂ fold-changes and Y-axis represents negative logarithm to the base 10 of the P-values. Black vertical and horizontal dashed lines reflect filtering criteria (FC=±1 and P value=.05). (C) Red and blue bars are number of significantly down-regulated (n=1547) and up-regulated genes (n=2600) in HNSCC compared with its adjacent tissues.

<http://links.lww.com/MD/F275>, which showed the fold changes and P values of the DEGs). In total, 4147 genes had differential expression, including 2600 up-regulated genes and 1547 down-regulated genes (Fig. 2C).

3.3. Construction of i-GCN with Pearson correlation analysis

The Pearson correlation coefficient and its P value between pairwise DEGs were calculated. The number of preliminary relationships was 17,197,609. Then the conditions $|r| \geq 0.6$ and $P < .05$ was set as the cut-off criteria. There were 129,220 relationships remained with 2,526 genes. An i-GCN was constructed based on this. The remained relationships were imported into Cytoscape software^[2,3] for visualization (Fig. 3A). There were a large net and several small nets, and the number of genes in each small net was less than 10. The large net (i-GCN) containing 2241 genes were remained for further research after small nets removed.

3.4. Divide i-GCN with community detection algorithms

Five different community detection algorithms, multilevel, eigenvector, label-propagation, map-equation and edge-betweenness were

used to identify the community of i-GCN. The modularity of each algorithm was shown in Table 1. It could be seen that the multilevel algorithm with the highest modularity was selected for further analysis.

The i-GCN was divided into 13 communities by multilevel algorithm (Fig. 3B and Supplementary Table 3, <http://links.lww.com/MD/F276>, which demonstrated the details of each module). The communities with genes less than 20 were removed and 9 communities corresponding to 9 gene modules were left. The network densities of these 9 communities were shown in Table 2. While the network density of i-GCN without community detection was 0.05131319. Each network density of these 9 communities was higher than the i-GCN.

3.5. Association analysis between gene modules and clinical indicators

The ME of each module was obtained by PCA, and the details were showed in Supplementary Table 4, <http://links.lww.com/MD/F277>. The association matrix was obtained from the result of correlation analysis between the modules and clinical indicators (Fig. 3C). The module was considered to be highly correlated with the clinical indicator when the absolute value of

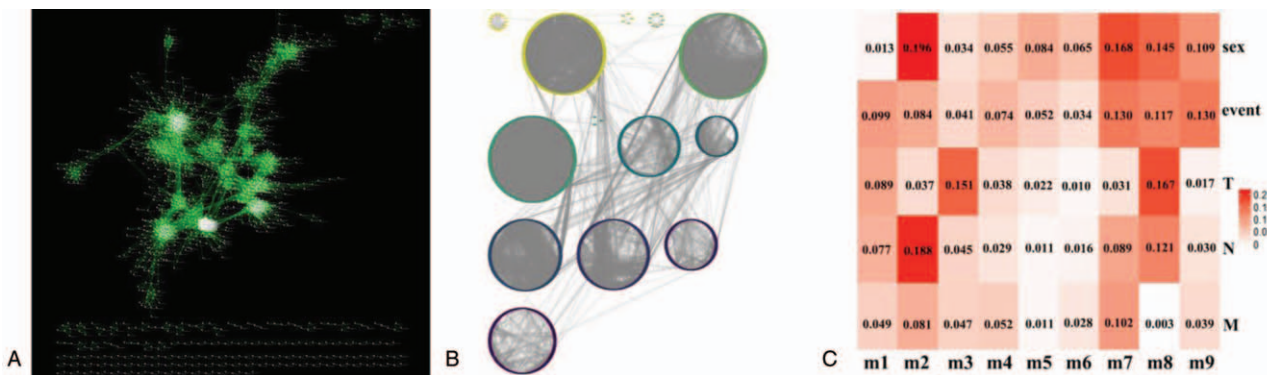


Figure 3. Construction of GCN and mining of gene modules. (A) The i-GCN was constructed by Pearson correlation analysis. (B) Division result obtained by the multilevel algorithm. The multilevel algorithm divided i-GCN into 13 communities. (C) The heat map of the correlation between modules and clinical indicators. The row corresponds to module, the column corresponds to clinical indicator. The modules m2, m3, m7, m8, and m9 were highly correlated with clinical indicators.

Table 1**The modularities of five algorithms.**

Algorithm	Modularity
multilevel	0.5256592
eigenvector	0.5137625
label-propagation	0.5155226
map-equation	0.01748193
edge-betweenness	0.5223523

the correlation coefficient was over 0.1. The results showed that m2, m7, m8, and m9 were highly correlated with sex, m7, m8 and m9 were highly correlated with event, m3 and m8 were highly correlated with T, m2 and m8 were highly correlated with N, m7 were highly correlated with M.

3.6. The biological significance of gene modules with GO enrichment analysis

The above results showed that modules m2, m3, m7, m8, and m9 were highly related to the clinical indicators. Then we found the biological functions of these modules with GO enrichment analysis (Table 3). Association analysis between the MEs of each module and clinical indicators showed that (Fig. 3C), sex was mainly related to mitosis and meiosis processes, event was mainly related to responding to interferons, viruses and T cell differentiation processes, T was mainly related to muscle development and contraction, regulation of protein transport activity processes, N was mainly related to mitosis and meiosis processes, while M was mainly related to responding to interferons and immune response processes.

3.7. Identification of key genes and exploration of their functions

In this paper, 2 methods were used to find key genes. One method was identifying key genes with the correlation of gene expression levels and clinical indicators. The GS values of DEGs were showed in Supplementary Table 5, <http://links.lww.com/MD/F278>. Among the results, 14 genes with $|GS| > 0.16$ were identified as the key genes. Then we searched these key genes for related diseases in the Disgenet database. The result showed that 9 key genes have a strong correlation with tumor diseases, i.e., HOXA1, ZAP70, XPR1, DONSON, CDC7, HENMT1, TNFRSF25, GNMT, CDKN2A. These key genes were related

Table 2**Network density of nine communities containing more than 20 genes.**

Module	Densities
m1	0.220859773
m2	0.221835782
m3	1.219824805
m4	0.141935484
m5	0.367359229
m6	0.158419958
m7	0.106810852
m8	0.10614192
m9	0.063424239

to a variety of tumor diseases, i.e., Lymphoma, Glioma, Lip and Oral Cavity Carcinoma, Malignant neoplasm of mouth, Colorectal Cancer, Prostate carcinoma, Central neuroblastoma, Precursor T-Cell Lymphoblastic Leukemia-Lymphoma, Prostate carcinoma, Breast Carcinoma, Liver carcinoma, Pancreatic Ductal Adenocarcinoma, etc. (Supplementary Table 5 <http://links.lww.com/MD/F278>, which demonstrated the key genes related to the diseases). It worth noted that 3 key genes were related to HNSCC, CDKN2A, HOXA1, and CDC7.

The other method was scoring the importance of all nodes in the i-GCN with the PageRank algorithm, the scores of each node were showed in the Supplementary Table 6, <http://links.lww.com/MD/F279> and the top 20 genes with the highest score were identified as the key genes. The results of the Disgenet database showed that 15 of the 20 key genes have a strong correlation with tumor diseases, i.e., PPL, PRSS27, AHSA2P, CD27, SULT2B1, FCRL5, IKZF2, EVPL, PXN, ASF1B, PDGFRB, LAMA4, CALD1, CD79A, and NUSAP1. These key genes were related to a variety of tumor diseases, i.e., Prostatic Neoplasms, Thyroid carcinoma, Lymphoma, Carcinoma of lung, Malignant Glioma, Breast Carcinoma, Ovarian Carcinoma, Dermatofibrosarcoma, Liver carcinoma, Urothelial Carcinoma, etc. (Supplementary Table 6, <http://links.lww.com/MD/F279> which demonstrated the key genes related to the diseases). It worth noted that 6 key genes were related to HNSCC, PPL, EVPL, PXN, PDGFRB, CALD1, and NUSAP1.

3.8. Survival analysis of key genes

To assess the utility of i-GCN at identifying key genes indicative of HNSCC, we conducted survival analysis with onclnc (Fig. 4). The survival curves results showed that the expression of HOXA1, PXN, and NUSAP1 were negatively correlated with the survival prognosis. While EVPL showed the opposite result. These were consistent with the results of DEGs identifying.

4. Discussion

To explore the gene modules and key genes of HNSCC, a bioinformatics algorithm based on the gene co-expression network analysis was proposed in this study. Related studies showed that there was a co-expression relationship between two genes if their absolute values of Pearson correlation coefficient over a certain threshold.^[24] Chang et al applied Pearson correlation coefficient to construct an i-GCN to compare transcriptomes from maize leaf and identified regulators of maize C4 enzyme genes.^[24] Based on the Pearson correlation coefficient between 2 gene expression data, an i-GCN was constructed in our study.

Then, comparing the results of 5 different community detection algorithms (multilevel, eigenvector, label-propagation, map-equation and edge-betweenness), the multilevel algorithm which had the highest modularity divided the i-GCN into 9 gene modules. Association analysis between the MEs of each module and clinical indicators showed that, sex was mainly related to mitosis and meiosis processes, event was mainly related to responding to interferons, viruses and T cell differentiation processes, T was mainly related to muscle development and contraction, regulation of protein transport activity processes, N was mainly related to mitosis and meiosis processes, while M was mainly related to responding to interferons and immune response processes.

Table 3**GO analysis of DEGs in highly correlative module.**

ID	Description	P value	Count
Sex-related biological processes			
GO:0007059	chromosome segregation	1.23E-71	86
GO:0000819	sister chromatid segregation	9.62E-70	73
GO:0006260	DNA replication	4.52E-58	70
GO:0000280	nuclear division	1.05E-57	80
GO:0140014	mitotic nuclear division	3.93E-56	67
GO:0034340	response to type I interferon	1.07E-42	33
GO:0051607	defense response to virus	3.92E-38	36
GO:0000082	G1/S transition of mitotic cell cycle	7.00E-35	50
GO:0009615	response to virus	2.41E-34	37
GO:0071103	DNA conformation change	4.06E-33	50
Event-related biological processes			
GO:0034340	response to type I interferon	1.07E-42	33
GO:0051607	defense response to virus	3.92E-38	36
GO:0009615	response to virus	2.41E-34	37
GO:0045071	negative regulation of viral genome replication	2.49E-23	17
GO:0070268	cornification	5.43E-22	21
GO:0019079	viral genome replication	5.97E-18	17
GO:0043900	regulation of multi-organism process	6.39E-17	26
GO:0035455	response to interferon-alpha	3.10E-13	8
GO:0018149	peptide cross-linking	1.45E-12	11
GO:0032480	negative regulation of type I interferon production	2.87E-09	8
T-related biological processes			
GO:0006936	muscle contraction	3.56E-56	68
GO:0055001	muscle cell development	2.25E-40	44
GO:0030239	myofibril assembly	2.50E-40	32
GO:0030049	muscle filament sliding	6.02E-35	24
GO:0033275	actin-myosin filament sliding	6.02E-35	24
GO:0010927	cellular component assembly involved in morphogenesis	2.24E-34	33
GO:0007517	muscle organ development	4.19E-34	53
GO:0031032	actomyosin structure organization	5.69E-28	35
GO:0014706	striated muscle tissue development	7.28E-25	43
GO:0050879	multicellular organismal movement	8.75E-21	18
N-related biological processes			
GO:0007059	chromosome segregation	1.23E-71	86
GO:0000819	sister chromatid segregation	9.62E-70	73
GO:0006260	DNA replication	4.52E-58	70
GO:0000280	nuclear division	1.05E-57	80
GO:0140014	mitotic nuclear division	3.93E-56	67
GO:0000082	G1/S transition of mitotic cell cycle	7.00E-35	50
GO:0071103	DNA conformation change	4.06E-33	50
GO:1901990	regulation of mitotic cell cycle phase transition	1.99E-31	57
GO:0051983	regulation of chromosome segregation	2.04E-30	31
GO:0007088	regulation of mitotic nuclear division	1.56E-27	35
M-related biological processes			
GO:0034340	response to type I interferon	1.07E-42	33
GO:0051607	defense response to virus	3.92E-38	36
GO:0009615	response to virus	2.41E-34	37
GO:0045071	negative regulation of viral genome replication	2.49E-23	17
GO:0019079	viral genome replication	5.97E-18	17
GO:0043900	regulation of multi-organism process	6.39E-17	26
GO:0035455	response to interferon-alpha	3.10E-13	8
GO:0032480	negative regulation of type I interferon production	2.87E-09	8
GO:0032606	type I interferon production	3.53E-09	11
GO:0050688	regulation of defense response to virus	9.98E-08	8

Lastly, combine with the GS values of all the DEGs and the PageRank algorithm to find the key genes. It was worth noting that the key genes were related to many skin disease, such as Dry skin, Vesicular Stomatitis, Skin Erosion, Eczema, Dermatologic disorders, Dermatitis, Atopic, Hyperextensible skin, Thin skin, etc. And there were skin-related tumor diseases, such as

Squamous cell carcinoma of skin, Skin Neoplasms, etc. We speculated that the pathogenesis of HNSCC might be similar with skin diseases. In addition, the key genes PRSS27, AHSA2P, CD27, SULT2B1, FCRL5, IKZF2, ASF1B, LAMA4, CD79A, ZAP70, XPR1, DONSON, HENMT1, TNFRSF25, GNMT were related to a variety of tumor diseases without HNSCC.

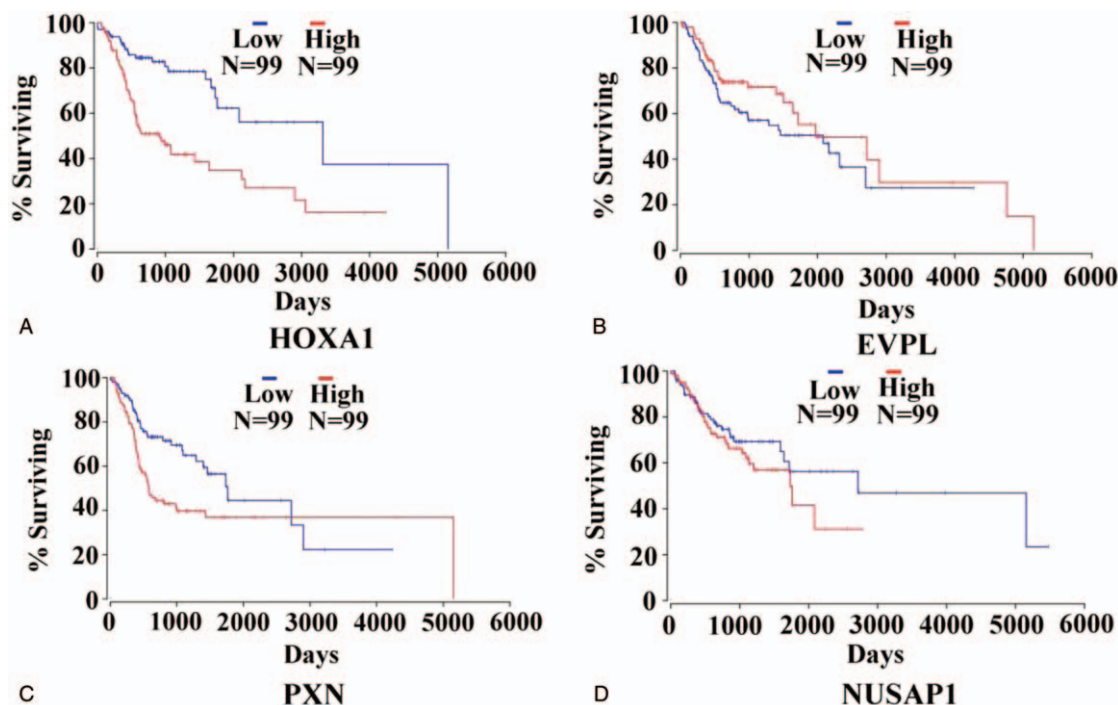


Figure 4. Significant correlation between key genes expression and survival. Survival curves of genes HOXA1, EVPL, PXN, and NUSAP1, X-axis represented survival time and Y-axis represented survival rate.

Therefore, the roles of these genes in HNSCC should be further study.

It was noted that HOXA1, PXN and NUSAP1 were negatively correlated with the survival prognosis. The survival curves results showed that the expression of HOXA1, PXN, and NUSAP1 were negatively correlated with the survival prognosis, the gene expression lower, the survival prognosis better. While EVPL showed the opposite result. Homeobox A1 (HOXA1) was a member of HOX gene family, which was a part of a cluster on chromosome 7 and encoded a DNA-binding transcription factor that might regulate gene expression, morphogenesis and differentiation. Previous results showed that HOXA1 was abnormally expressed in leukemia, cervical cancer, and breast cancer and it was associated with prognosis.^[25,26] PXN participated in cell signal transmission and played a role in organ development, damage repair and cell movement. PXN was abnormally expressed in a large number of digestive system tumors, but its function of suppressing or promoting cancer was remained unclear.^[27,28] Nucleolar spindle associated protein 1 (NuSAP1) mainly participated in the assembly process of mitotic spindle. It was an important regulatory molecule to ensure the normal cell cycle. NuSAP1 was overexpressed in a variety of tumors, which was significantly associated with invasion and metastasis and poor prognosis.^[29,30] However, the specific functions of these genes that contribute to HNSCC cell proliferation, differentiation, and metastasis needed further study.

5. Conclusions

An i-GCN was constructed with Pearson correlation analysis. Association analysis between the MEs of each module and clinical indicators showed that, sex was mainly related to mitosis

and meiosis processes, event was mainly related to responding to interferons, viruses and T cell differentiation processes, T was mainly related to muscle development and contraction, regulation of protein transport activity processes, N was mainly related to mitosis and meiosis processes, while M was mainly related to responding to interferons and immune response processes. Lastly, HOXA1, PXN, and NUSAP1 might play important roles in the progression of HNSCC and severed as potential biomarkers for future diagnosis.

Author contributions

Conceptualization: Qian Zhao.

Data curation: Yan Zhang, Xue Zhang.

Formal analysis: Qian Zhao, Yan Zhang.

Investigation: Xue Zhang.

Methodology: Yan Zhang.

Project administration: Yeqing Sun, Zhengkui Lin.

Software: Yan Zhang.

Supervision: Yeqing Sun, Zhengkui Lin.

Writing – original draft: Qian Zhao, Yan Zhang.

Writing – review & editing: Yeqing Sun, Zhengkui Lin.

References

- [1] Pablo , Ramos-García , Miguel , et al. Prognostic and clinicopathological significance of CTTN/cortactin alterations in head and neck squamous cell carcinoma: Systematic review and meta-analysis. *Head Neck* 2019;41:1963–78.
- [2] Zou AE, Zheng H, Saad MA, et al. The non-coding landscape of head and neck squamous cell carcinoma. *Oncotarget* 2016;7: 51211–22.
- [3] Yan L, Zhan C, Wu J, et al. Expression profile analysis of head and neck squamous cell carcinomas using data from The Cancer Genome Atlas. *Mol Med Report* 2016;13:4259–65.

- [4] Butte AJ, Kohane IS. Unsupervised knowledge discovery in medical databases using relevance networks. *Proc AMIA Symp* 1999;711–5.
- [5] Butte AJ. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing* 2000;418–29.
- [6] Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005;Article 17.
- [7] Wang-Xiao , Xia , Qin , et al. Identification of four hub genes associated with adrenocortical carcinoma progression by WGCNA. *PeerJ* 2019;7:e6555.
- [8] Tang W, Guo X, Niu L, et al. Identification of key molecular targets correlation with breast cancer through bioinformatic methods. *J Gene Med* 2019;22:e3141.
- [9] Ao ZX, Chen YC, Lu JM, et al. Identification of potential functional genes in papillary thyroid cancer by co-expression network analysis. *Oncol Lett* 2018;16:4871–8.
- [10] Hutter C, Zenklusen JC. The cancer genome atlas: creating lasting value beyond its data. *Cell* 2018;173:283–5.
- [11] Chen J, Wang X, Hu B, et al. Candidate genes in gastric cancer identified by constructing a weighted gene co-expression network. *PeerJ* 2018; e4692.
- [12] Blondel VD, Guillaume JL, Lambiotte R, et al. Fast unfolding of communities in large networks. *J Statistical Mechanics* 2008;0.
- [13] Newman MEJ. Finding community structure in networks using the eigenvectors of matrices. *Phys Rev E Stat Nonlin Soft Matter Phys* 2006;74:036104.
- [14] Raghavan UN, Albert R, Kumara S. Near linear time algorithm to detect community structures in large-scale networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 2007;76:036106.
- [15] Rosvall M, Bergstrom CT. Maps of information flow reveal community structure in complex networks. *Proceedings National Academy Sci USA* 2007;1118–23.
- [16] Rosvall M, Axelsson D, Bergstrom CT. The map equation. *Eur Phy J Special Topics* 2009;13–23.
- [17] Newman MEJ, Girvan M. Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 2004;026113.
- [18] Csardi G, Nepusz T. The igraph software package for complex network research. *Int J Complex Syst* 2006;1–9.
- [19] Wold S, Esbensen K, Geladi P. Principal component analysis. *Chemometrics Intellig Lab Syst* 1987;37–52.
- [20] Brin S, L, Page. The anatomy of a large-scale hypertextual web search engine. 1998;Computer Networks and ISDN Systems,
- [21] Pinero J, Ramirez-Anguita JM, Sauch-Pitarch J, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res* 2020;D845–55.
- [22] Anaya J. OncoLnc: linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. *PeerJ Computer Sci* 2016;e67.
- [23] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–504.
- [24] Chang YM, Lin HH, Liu WY, et al. Comparative transcriptomics method to infer gene coexpression networks and its applications to maize and rice leaf transcriptomes. *Proc Natl Acad Sci U S A* 2019;116:3091–9.
- [25] Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. *CA Cancer J Clin* 2015.
- [26] John S, MA. Inflammatory mediators drive metastasis and drug resistance in head and neck squamous cell carcinoma. *Laryngoscope* 2015;125(Suppl 3):S1–1.
- [27] Li D, Li Z, Xiong J, et al. MicroRNA-212 functions as an epigenetic-silenced tumor suppressor involving in tumor metastasis and invasion of gastric cancer through down-regulating PXN expression. *Am J Cancer Res* 2015;5:2980–97.
- [28] Minyoung L, Jung-Jin P, L, Yun-Sil. Adhesion of ST6Gal I-mediated human colon cancer cells to fibronectin contributes to cell survival by integrin beta1-mediated paxillin and AKT activation. *Oncol Rep* 2010;23:757–61.
- [29] Iyer J, Moghe S, Furukawa M, et al. What's Nu(SAP) in mitosis and cancer? *Cell Signal* 2011;23:991–8.
- [30] Kotian S, Banerjee T, Lockhart A, et al. NUSAP1 influences the DNA damage response by controlling BRCA1 protein levels. *Cancer Biol Ther* 2014;15:533–43.