

Development of a Prognostic Signature Based on Hypoxia-related Genes for Thyroid Cancer

HongYang Zhang

The First Affiliated Hospital of China Medical University <https://orcid.org/0000-0002-4494-1094>

Sijia Li

The First Affiliated Hospital of China Medical University

Wei Li (✉ wli@cmu.edu.cn)

Department of Otolaryngological Medicine, The First Hospital of China Medical University, Shenyang, Liaoning 110001, China <https://orcid.org/0000-0001-9368-3397>

Research

Keywords: Hypoxia, Thyroid cancer, Prognosis

Posted Date: April 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-438962/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background. We aimed to establish a model to predict the prognosis of patients with thyroid cancer based on differentially expressed hypoxia-related genes.

Methods. By comparing the genes in TCGA database and hypoxiaDB database, we obtained differentially expressed genes (DEGs) related to hypoxia in thyroid cancer. Gene function enrichment analysis was performed, and a protein-protein interaction network was constructed using the STRING database. Univariate Cox regression were used to screen hypoxia-related genes with prognostic value. Subsequently, multivariate Cox analysis was used to determine prognostic markers based on thyroid cancer, a prognosis model based on these genes was established. The Kaplan-Meier analysis, Receiver operating characteristic (ROC) analysis and The Harrell's concordance indexes in the training set and the validation set were used to evaluate the performance of the model. Finally, we conducted univariate analyses of the prognostic value of clinical data (including risk scores) of thyroid cancer patients.

Results. 326 hypoxia-related thyroid cancer genes were found. Functional enrichment analysis demonstrated they were mainly involved in regulating biological functions. 23 genes have been proved to be associated with the prognosis of thyroid cancer with univariate Cox regression, among them, 11 marker genes were used to construct a new prognosis model by multivariate Cox analysis. Accordingly, the system of risk scores was constructed, patients with high-risk scores ($P < 0.005$) had shorter overall survival than those with low-risk scores. The ROC curve indicated good performance of the eleven-gene signature at predicting overall survival. The Harrell's concordance indexes in the internally validated for the 11-gene prognostic signature was 0.881. Moreover, univariate analysis showed that the risk score and age were significantly associated with patient overall survival. The model we created was significantly associated with patient overall survival.

Conclusions. The model we established had excellent performance in the prognosis of thyroid cancer.

Introduction

The most common form of endocrine cancer is thyroid cancer, which occurs in all age groups, i.e., children to the elderly [1] Due to the rapid development of detection technology, more cases of thyroid cancer have been clinically diagnosed, which is more concerning. At present, the diagnosis of thyroid cancer mainly relies on the ultrasound and percutaneous fine-needle aspiration (pFNA), while the treatment mainly remains to be surgery and I^{131} radiotherapy. Such mature diagnosis and treatment programs are widely being recognized. Meanwhile, the prognosis of thyroid cancer has become a hot topic of discussion among scholars.

It is proven that the microenvironment in tumor cells is chaotic and complex, which is caused by insufficient oxygen supply from hypoxia. Under a hypoxia environment, gene and protein expression could undergo regulation, while hypoxia can also play a role in genetic instability, tumorigenesis, and

progression. For example, hypoxia causes changes in the tumor microenvironment, promoting inflammation, immune suppression, and treatment resistance, inducing lung cancer[2]

Recently, scientists have made great efforts to use hypoxia-related markers to evaluate the prognosis of tumors. For example, three hypoxia-related genes (PDSS1, CDCA8, and SLC7A11) were used to construct a prognosis, recurrence, and diagnosis model for hepatocellular carcinoma (HCC) [3] Based on hypoxia-related genes, a hypoxia risk model was developed to evaluate the prognosis of glioma patients[4]. The establishment of these hypoxia-related prognostic independent models has significantly contributed to the comprehensive treatment of cancer patients. Our study intended to provide novel and reliable prognostic markers for a comprehensive treatment of thyroid cancer patients.

Materials And Methods

Acquisition of hypoxia-related genes

According to FDR (false discovery rate) <0.05 and $|\log_2FC| >1$, we obtained 2215 differentially expressed genes, combined with the differentially expressed genes downloaded from hypoxia DB database, which provided a complete and up-to-date database for the study of human hypoxia-related regulatory proteins, we obtained 373 differentially expressed hypoxia-related genes.

All differentially expressed genes and differentially expressed hypoxia-related genes are shown in raw data.

Hypoxia-related gene database

A total of 373 differentially expressed genes were selected from the hypoxia-related gene database. The database, HypoxiaDB, was based on the literature, which provided a complete and up-to-date database for the study of human hypoxia-related regulatory proteins.

Thyroid cancer transcriptome data and clinical data

The transcriptome data was downloaded from the official website of the Cancer Genome Atlas of Thyroid Cancer, which also included 510 thyroid cancer specimen data and 58 adjacent to cancer specimen data for further analysis.

Functional Analyses

In this study, the clusterProfiler was used for GO and KEGG analyses to study the functional correlation between the differentially expressed hypoxia-related genes. The value of $P < 0.05$ was considered statistically significant.

Construction of protein interaction network

Hypoxia-related genes were mapped to the STRING database to create an interactive network to clarify the association between differentially expressed genes. Then, the Cytoscape software was adopted to visualize the protein-protein interaction network.

Construction and validation of a prognostic model of hypoxia-related markers

Univariate Cox proportional hazard regression analyses were used to identify differentially expressed hypoxia-related genes associated with overall survival time, where $P < 0.05$ was considered to be statistically significant. After combining the regression coefficient (β) of each specific gene, the patient's risk score formula was constructed by weighing the estimated regression coefficient. The risk score of each patient was found according to the following formula: Risk score = β gene(1) \times expression gene(1) + β gene(2) \times expression gene(2) + ... + β gene(n) \times expression gene(n). In the testing set and the validation set, according to the risk score formula, with the median risk score as the critical point, patients were divided into low-risk groups and high-risk groups. The Kaplan-Meier analysis was used to assess the survival difference between the two groups in two sets using the log-rank as the comparison. Receiver operating characteristic (ROC) analysis and The Harrell's concordance indexes was used to assess the accuracy of the model predictions. Univariate Cox regression was used to identified the independent prognostic factors among risk scores, age, tumor TNM, stage, Focus and gender.

Internal validation of the prediction model

To further verify the predictive power of this model, we used the 50% thyroid cancers samples randomly selected from the entire TCGA database as internal validation dataset (n = 255). The C-index was used to assess the performance of the established model.

Statistical Analyses

The TCGA database was analyzed using the R language. The survival curve was made using Kaplan-Meier and compared by the log-rank test. Multivariate Cox analysis of risk factors was used to establish a prognostic prediction model for hypoxia-related genes. All statistical analyses were conducted using the R language. All statistical tests were two-sided with $P < 0.05$ being statistically significant.

Results

Identification of differentially expressed hypoxia-related genes

In 373 differentially expressed hypoxia-related genes, we showed the top 30 genes (Figs. 1A,1B). One of the up-regulated genes identified was CCND1. while there were 29 down-regulated genes, including: AKAP8L, ANKZF1, AOX1, ASXL1, CXXC1, DES, EPB41L2, FGF7, GBA2, HIF3A, INHA, IRAK3, LIFR, LYVE1, NDRG2, PAK3, PITRM1, PLAGL1, RUNX1T1, SGSM3, SLC5A6, SLC7A6, SORBS1, SRPX, TBC1D8, TFPI, TPCN2, TWIST2, UCKL1. The expression patterns of 30 differentially expressed hypoxia-related genes between the thyroid cancer and adjacent tissues are summarized in the box plot (Fig.1C).

Functional enrichment analysis of differentially expressed hypoxia-related genes

The functional enrichment analysis of differentially expressed hypoxia-related genes helped in their biological understanding. Figure 2 summarizes the first ten biological processes and ten pathways of GO and KEGG enrichment, respectively. GO enrichment showed that the biological processes of differential genes mainly involved 'extracellular structure organization', 'collagen-containing extracellular matrix' and 'extracellular matrix structural constituent' (Fig. 2A). while the KEGG enrichment showed the following pathways of differential genes, such as: 'PI3K-Akt signaling pathway', 'Cytokine-cytokine receptor interaction' and 'MAPK signaling pathway' (Fig.2B).

PPI network construction and important gene modules

To fully understand the differentially expressed hypoxia-related genes, we utilized Cytoscape software to construct an interactive PPI network (Fig.3). As shown in the figure, the red color represents up-regulated hypoxia-related genes, while the green color indicates down-regulated ones. Molecular Complex Detection (MCODE) tool was used to identify significant gene modules, and three significant gene modules were screened out (Supplementary Fig.1).

Identification of hypoxia-related genes for prognosis

Univariate Cox regression analysis was performed to determine the hypoxia-related genes associated with the prognosis of thyroid cancer patients. The forest plot showed that there were 23 genes (PTX3, SLC6A8, STC1, STARD4, PTGIS, SLC29A4, SLC24A3, SEMA6D, COL6A1, FGF7, DPP4, HIF3A, PLAGL1, ANK2, APOE, TMEM45A, PIM1, NDRG1, CSGALNACT, PHKG1, TRIO, AKR1C3, KDM6B) with significant prognostic value in thyroid cancer patients ($P < 0.05$; Fig. 4A). A total of 11 genes were identified through multivariate Cox regression analysis, including: SLC6A8, STC1, PTGIS, SLC24A3, DPP4, ANK2, APOE, PIM1, PHKG1, AKR1C3, NDRG1 ($P < 0.05$; Fig.4B).

Establishment of Prognostic Formula by Multivariate Cox Regression Analysis

The risk score of genes was found established according to the following formula: Genetic risk score = $0.7926 \times \text{SLC6A8} + 0.8064 \times \text{STC1} + 0.7867 \times \text{PTGIS} - 1.057 \times \text{SLC24A3} - 0.3492 \times \text{DPP4} - 0.7492 \times \text{ANK2} - 1.2040 \times \text{APOE} + 0.5211 \times \text{PIM1} + 3.0850 \times \text{PHKG1} + 0.5473 \times \text{AKR1C3} + 1.1714 \times \text{NDRG1}$.

Based on the prognosis formula of hypoxia-related genes, we determined the distribution of these genes in different risk populations and the survival rate of patients in the testing set and validation set. To determine the role of 11 hypoxia-related genes in predicting the clinical prognosis of thyroid cancer patients, a K-M (Kaplan-Meier) survival curve was plotted to analyze different survival times between high-risk and low-risk groups in the two sets. The K-M analyses informed us that the survival rate of patients in the high-risk group was significantly lower than that of the low-risk group (Fig.5).

Later, 11 genes related to hypoxia were used to construct ROC curves for survival rates of thyroid cancer patients for 1 year, 3 years, and 5 years in the two sets to evaluate their predictive performance (Fig.6).

The areas under the curve (AUC) were 0.943 (1 year), 0.897 (3 years), and 0.831 (5 years) in the testing set. While the areas under the curve (AUC) were 0.944 (1 year), 0.964 (3 years), and 0.992 (5 years) in the validation set.

Subsequently, we conducted univariate (Table 1) analyses of the prognostic value of clinical data (including risk scores) of thyroid cancer patients, and it turned out that the risk score we constructed exhibited a high value in the prognostic evaluation.

Discussion

Thyroid cancer incidence has been showing an increasing trend in recent years, posing a major threat to human health [5]. In the United States, the detection rate of thyroid cancer has increased in both males and females, i.e., 4.9 cases per 100,000 people in 1975 to 14.3 cases per 100,000 people in 2014 [6]. Although most of the pathological types of thyroid cancer have a high survival rate, there are poorly differentiated thyroid cancer (PDTTC) with a five-year specific survival rate of 66% [7], which has attracted the attention of many scholars. In this study, we hope to find a new prognostic plan for thyroid cancer patients as well.

With the in-depth understanding of the hypoxia-induced changes in the tumor microenvironment along with the improvement of TCGA and HypoxiaDB database, scholars have paid more attention to the prediction of tumor prognosis based on hypoxia-related factors [8, 9]. Recently, emerging evidence has emphasized the importance of the immune and inflammatory microenvironment in thyroid cancer [10, 11]. However, none of the studies have analyzed the prognosis of thyroid cancer patients using hypoxia-related factors.

At present, most studies on the hypoxic microenvironment in thyroid cancer are still focused on a single gene. In this study, 23 differentially expressed genes were obtained using the thyroid cancer hypoxia-related data from the TCGA database and hypoxia-related genes data from the HypoxiaDB database. Besides being related to hypoxia, the enrichment analysis of GO and KEGG showed that the biological process of these hypoxia-related differential genes also involved 'extracellular structure organization', 'collagen-containing extracellular matrix' and 'extracellular matrix structural constituent' with the following pathways, such as: 'PI3K-Akt signaling pathway', 'Cytokine-cytokine receptor interaction' and 'MAPK signaling pathway'. After multivariate survival analyses, we identified 11 hypoxia-related genes, including SLC6A8, STC1, PTGIS, SLC24A3, DPP4, ANK2, APOE, PIM1, PHKG1, AKR1C3, NDR-G1, which were closely related to the prognosis. Moreover, a risk scoring system was constructed to provide clinicians with new tools for analyzing the prognosis of patients. This scoring tool showed greater advantages compared to the patients' clinical data (such as gender, age, TNM stage) during the prognostic analysis.

Among 11 hypoxia-related genes that we have identified, many of them were closely related to the prognosis of tumors in previous studies as well. In the previous animal experiments, it was confirmed that knockout of the CrT (SLC6A8) gene lead to the insufficient uptake of creatine while impairing the

immunity of anti-tumor T cells. *It was* also proved that the supplementation of creatine significantly inhibited tumor growth, which also synergized with PD-1/PD-L1 blockers to inhibit the tumors [12]. Apolipoprotein E (ApoE), the main cause of hyperlipidemia, was also shown to improve the hypoxic microenvironment of tumors in hyperlipidemic models through exercise while also slowing down the formation of primary and secondary E0771 breast tumors [13].

Further, PTGIS and DPP4 have also been confirmed in the bioinformatics research. In the previous study of novel biomarkers related to liver hepatocellular carcinoma (LIHC), PTGIS was one of the 21 hub genes identified by mRNA expression network analyses, which might have been a potential therapeutic target for inhibiting LIHC cells [14] while the dipeptidyl peptidase-4 (DPP4) was shown to be the target gene of NF- κ B [15].

Stanniocalcin-1 (STC1), PIM1, and NDRG1 are other target genes that are studied in emerging literature, where STC1 is thought to play a carcinogenic role in hypoxic gastric cancer by causing an imbalance of Bcl-2. Thus, it is regarded as a potential therapeutic target for gastric cancer [16]. Down-regulation of miR-124 and miR-144 in the hypoxic microenvironment can increase the risk of prostate cancer by weakening the inhibitory effect of PIM1, which eventually promotes hypoxia and radiation in the prostate cancer cells [17]. As a key gene in regulating lipid metabolism, NDRG1 may promote the aggressiveness of breast cancer, and the close relationship between NDRG1 and poor prognosis of breast cancer makes NDRG1 a promising therapeutic target for breast cancer [18].

In short, in this study, a variety of prognostic markers for thyroid cancer was identified based on the comprehensive analysis of expression profiles of differentially expressed hypoxia-related genes and corresponding clinical characteristics. By constructing a new risk scoring model, the prognosis of patients with thyroid cancer can be effectively evaluated. However, the limitation lies in that it is a retrospective study. More prospective studies should be conducted to verify the prognostic function of hypoxia-related genes. Multi-center data was required to confirm our findings. In short, we hope to contribute to the prognostic analysis of thyroid cancer and the determination of treatment targets.

Conclusions

A variety of prognostic markers for thyroid cancer were identified in this study based on the comprehensive analysis of expression profiles of differentially expressed hypoxia-related genes and corresponding clinical characteristics. By constructing a new risk scoring model, the prognosis of patients with thyroid cancer could be effectively evaluated. However, the limitation is that it is a retrospective study. Thus, more prospective studies should be conducted to verify the prognostic function of hypoxia-related genes. Also, multi-center data are required to confirm our findings. In short, we hope to contribute to the prognostic analysis of thyroid cancer.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The author declares that he has no competing interests.

Funding statement

This study was supported by educational commission of Liaoning Province in China (Numbers: QN2019009).

Authors' Contributions

Wei, Li designed the study protocol and analyzed the data. Sijia, Li performed statistical analysis. Zhang, Hong Yang completed the figures and wrote the article. Li, Wei and Hong yang, Zhang made revisions. All authors approved the final version of the article.

Acknowledgements

Not applicable.

References

1. Luzon-Toro B, Fernandez RM, Villalba-Benito L et al. Influencers on Thyroid Cancer Onset: Molecular Genetic Basis. *Genes (Basel)* 2019; 10.
2. Mittal V, El Rayes T, Narula N et al. The Microenvironment of Lung Cancer and Therapeutic Implications. *Adv Exp Med Biol* 2016; 890: 75-110.
3. Zhang B, Tang B, Gao J et al. A hypoxia-related signature for clinically predicting diagnosis, prognosis and immune microenvironment of hepatocellular carcinoma patients. *J Transl Med* 2020; 18: 342.
4. Lin W, Wu S, Chen X et al. Characterization of Hypoxia Signature to Evaluate the Tumor Immune Microenvironment and Predict Prognosis in Glioma Groups. *Front Oncol* 2020; 10: 796.

5. Kitahara CM, Sosa JA. The changing incidence of thyroid cancer. *Nat Rev Endocrinol* 2016; 12: 646-653.
6. Davies L, Welch HG. Current thyroid cancer trends in the United States. *JAMA Otolaryngol Head Neck Surg* 2014; 140: 317-322.
7. Ibrahimasic T, Ghossein R, Shah JP, Ganly I. Poorly Differentiated Carcinoma of the Thyroid Gland: Current Status and Future Prospects. *Thyroid* 2019; 29: 311-321.
8. Wang J, Wang Y, Xing P et al. Development and validation of a hypoxia-related prognostic signature for breast cancer. *Oncol Lett* 2020; 20: 1906-1914.
9. Yang L, Roberts D, Takhar M et al. Development and Validation of a 28-gene Hypoxia-related Prognostic Signature for Localized Prostate Cancer. *EBioMedicine* 2018; 31: 182-189.
10. Ferrari SM, Fallahi P, Galdiero MR et al. Immune and Inflammatory Cells in Thyroid Cancer Microenvironment. *Int J Mol Sci* 2019; 20.
11. Lin P, Guo YN, Shi L et al. Development of a prognostic index based on an immunogenomic landscape analysis of papillary thyroid cancer. *Aging (Albany NY)* 2019; 11: 480-500.
12. Di Biase S, Ma X, Wang X et al. Creatine uptake regulates CD8 T cell antitumor immunity. *J Exp Med* 2019; 216: 2869-2882.
13. Buss LA, Dachs GU. Voluntary exercise slows breast tumor establishment and reduces tumor hypoxia in ApoE(-/-) mice. *J Appl Physiol (1985)* 2018; 124: 938-949.
14. Bai KH, He SY, Shu LL et al. Identification of cancer stem cell characteristics in liver hepatocellular carcinoma by WGCNA analysis of transcriptome stemness index. *Cancer Med* 2020; 9: 4290-4298.
15. Li Y, Yang L, Dong L et al. Crosstalk between the Akt/mTORC1 and NF-kappaB signaling pathways promotes hypoxia-induced pulmonary hypertension by increasing DPP4 expression in PSMCs. *Acta Pharmacol Sin* 2019; 40: 1322-1333.
16. Wang Y, Qi Z, Zhou M et al. Stanniocalcin1 promotes cell proliferation, chemoresistance and metastasis in hypoxic gastric cancer cells via Bcl2. *Oncol Rep* 2019; 41: 1998-2008.
17. Gu H, Liu M, Ding C et al. Hypoxia-responsive miR-124 and miR-144 reduce hypoxia-induced autophagy and enhance radiosensitivity of prostate cancer cells via suppressing PIM1. *Cancer Med* 2016; 5: 1174-1182.
18. Sevinsky CJ, Khan F, Kokabee L et al. NDRG1 regulates neutral lipid metabolism in breast cancer cells. *Breast Cancer Res* 2018; 20: 55.

Tables

Due to technical limitations, table 1 is only available as a download in the Supplemental Files section.

Figures

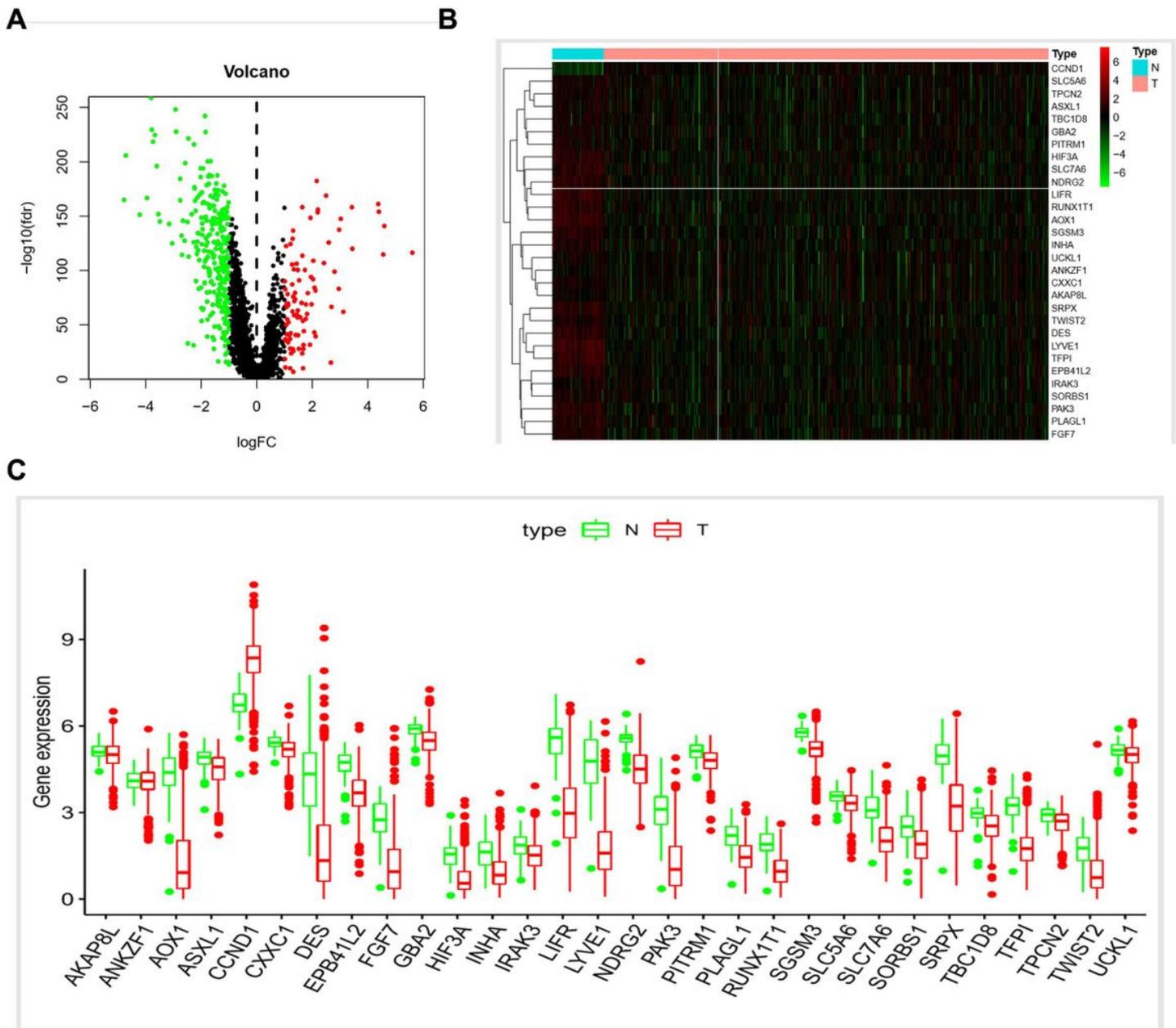


Figure 1

Differentially expressed hypoxia-related genes. (A) Volcano map showing differentially expressed hypoxia-related genes between thyroid cancers and paracancerous tissues. Red dots indicate significantly up-regulated genes, green dots represent markedly down-regulated genes. (B) Heatmaps showed cluster analysis of the top 30 differentially expressed hypoxia-related genes. (C) Box plot of 30 hypoxia-related genes in thyroid cancer tissues and paired paracancerous tissues. Each red box represents an hypoxia-related gene in thyroid cancer tissues, and each blue box represents an hypoxia-related gene in paracancerous tissues.

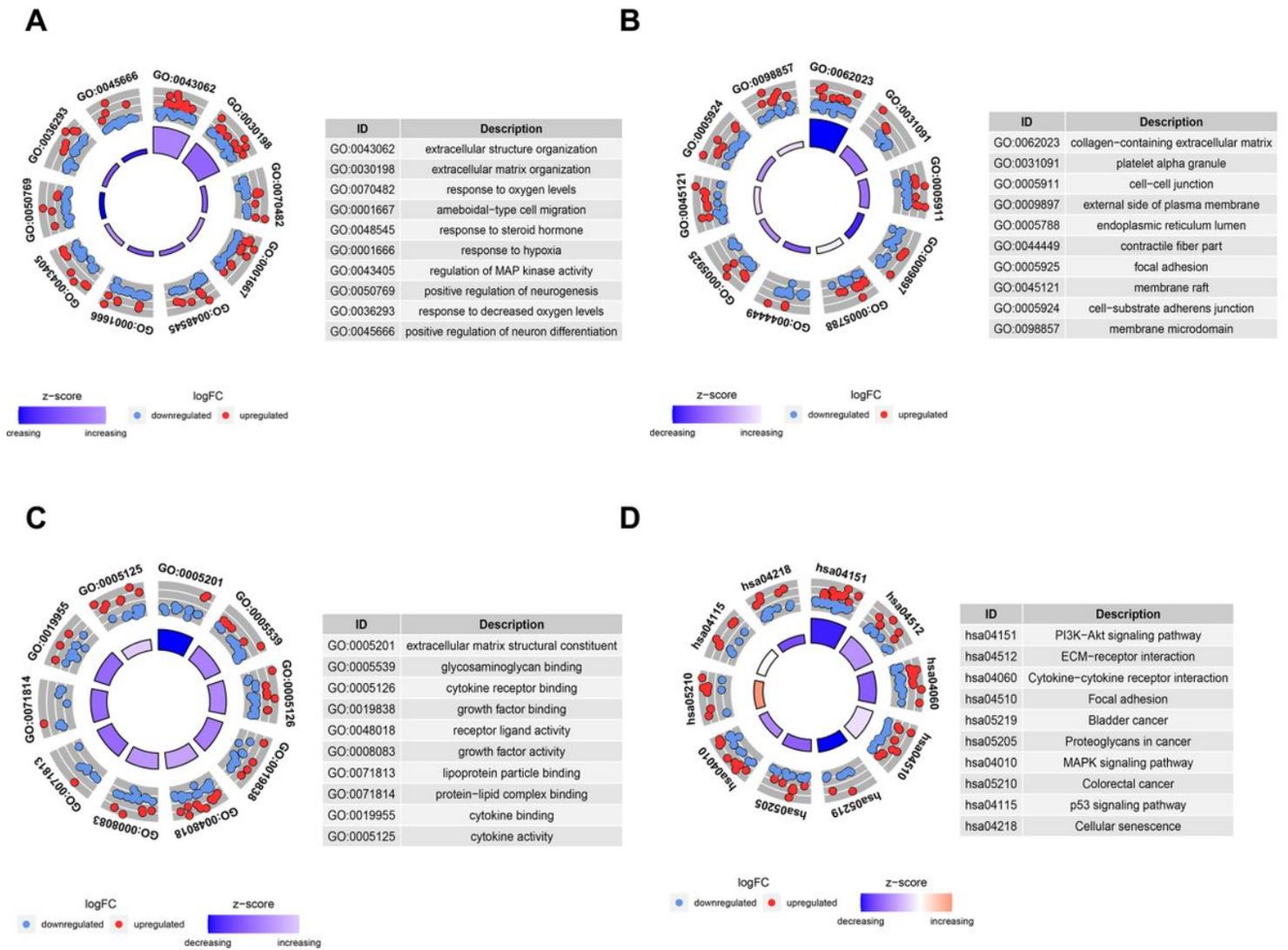


Figure 2

Functional enrichment of differentially expressed hypoxia-related genes. GO analysis to reveal the (A) Biological Processes, (B) Cell Components, (C) Molecular Functions involved in the differentially expressed hypoxia-related genes. (D) KEGG enrichment to show the signal pathways involved in differentially expressed hypoxia-related genes.

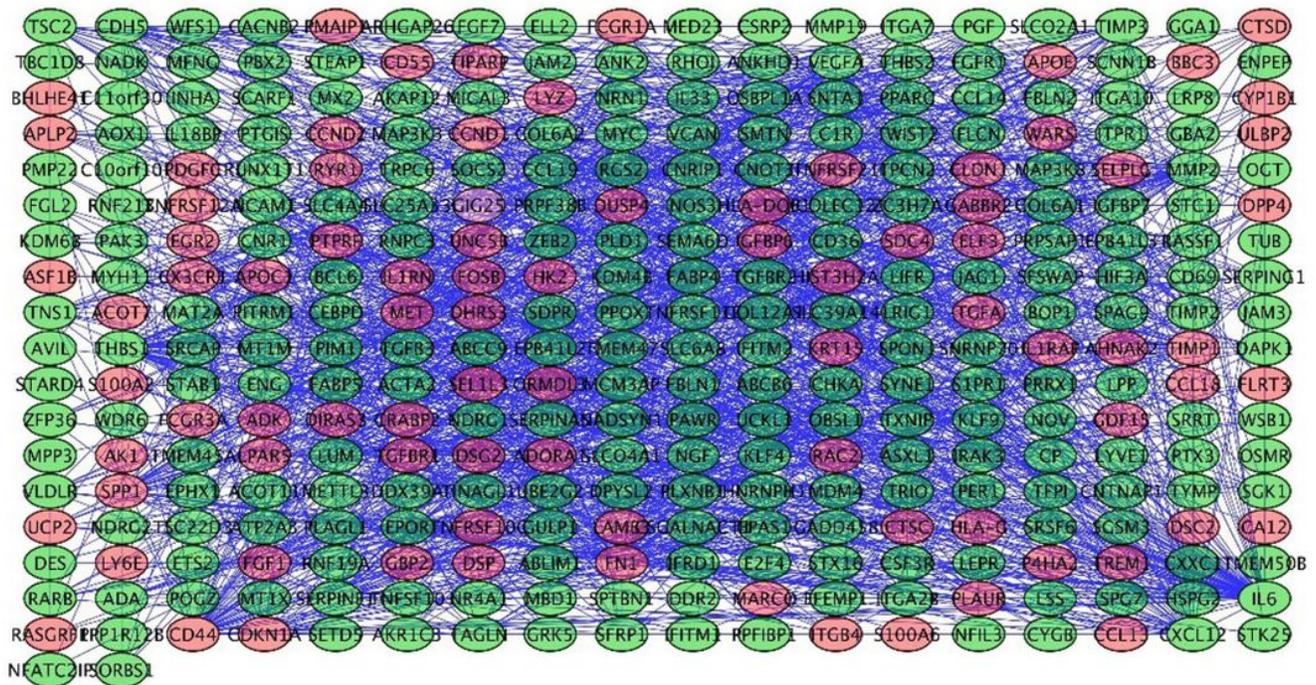


Figure 3

Construction of PPI network. A comprehensive PPI network of differentially expressed hypoxia-related genes was constructed. Red indicates up-regulated hypoxia related genes, and green indicates down-regulated hypoxia related genes. There are 326 nodes and 1486 edges.

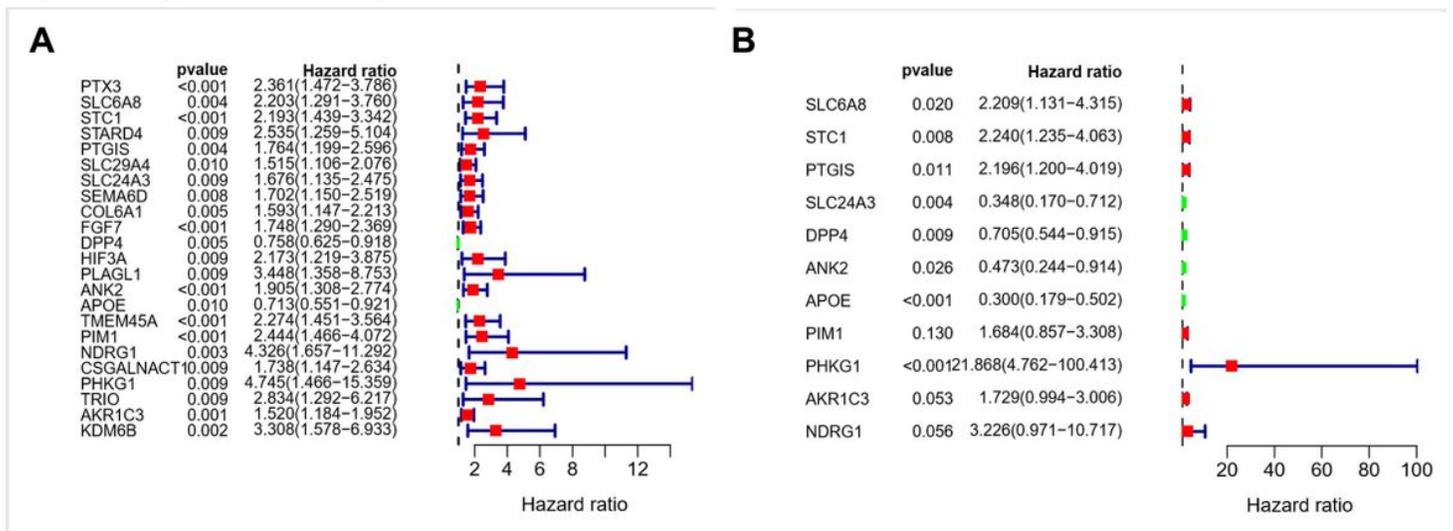


Figure 4

Univariate and multivariate analysis of differentially expressed genes. According to (A) univariate analysis, a total of 23 hypoxia-related genes were significantly associated with overall survival. Red

represents up-regulated genes and green represents down-regulated genes. According to (B) multivariate analysis, a total of 11 hypoxia-related genes were significantly associated with overall survival.

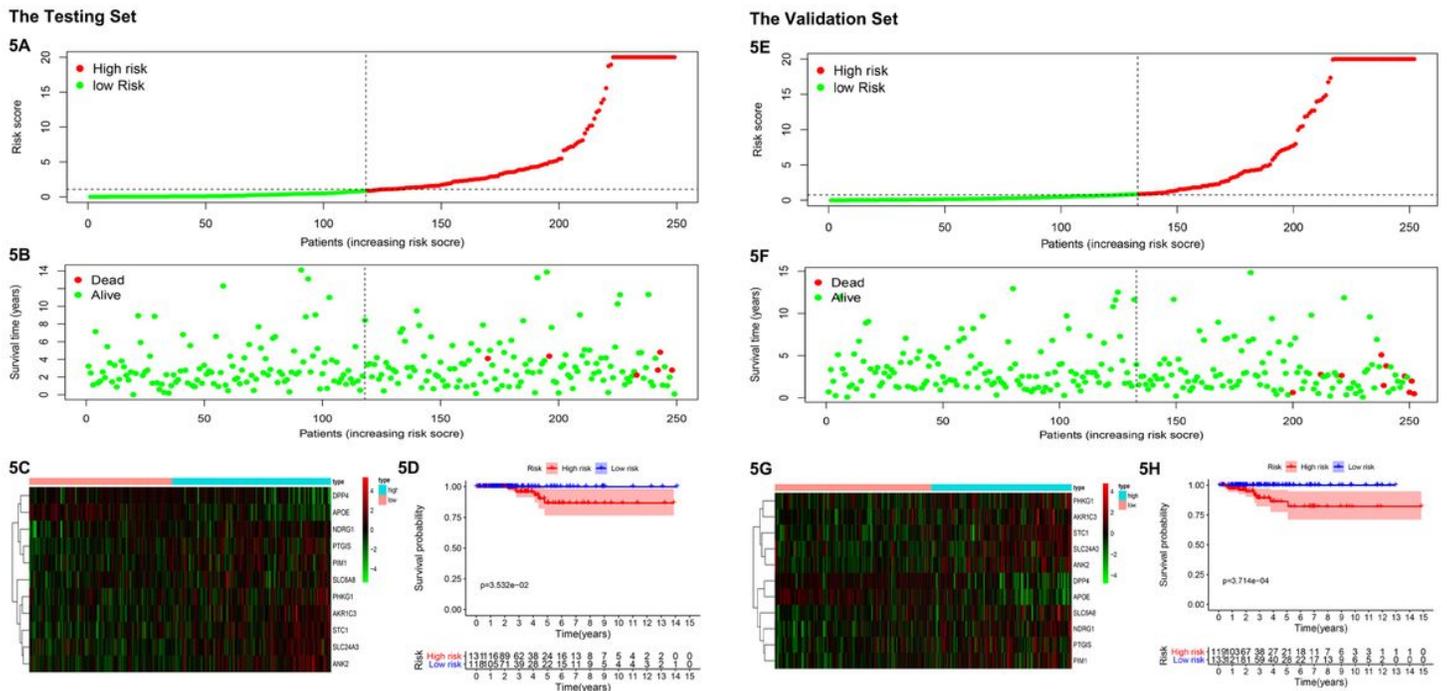


Figure 5

Verification of prognostic indicators for thyroid cancer based on hypoxia-related genes. (A) Distribution of different risk groups in the testing set. (B) Survival status of patients in different groups, wherein the red dots indicate dead patients and the blue dots represent surviving patients in the testing set. (C) Heat maps of hypoxia-related gene expression profiles from 11 multivariate analyses in the testing set. (D) The high-risk patients exhibited a shorter overall survival in the testing set. (E) Distribution of different risk groups in the validation set. (F) Survival status of patients in different groups, wherein the red dots indicate dead patients and the blue dots represent surviving patients in the validation set. (G) Heat maps of hypoxia-related gene expression profiles from 11 multivariate analyses in the validation set. (H) The high-risk patients exhibited a shorter overall survival in the validation set.

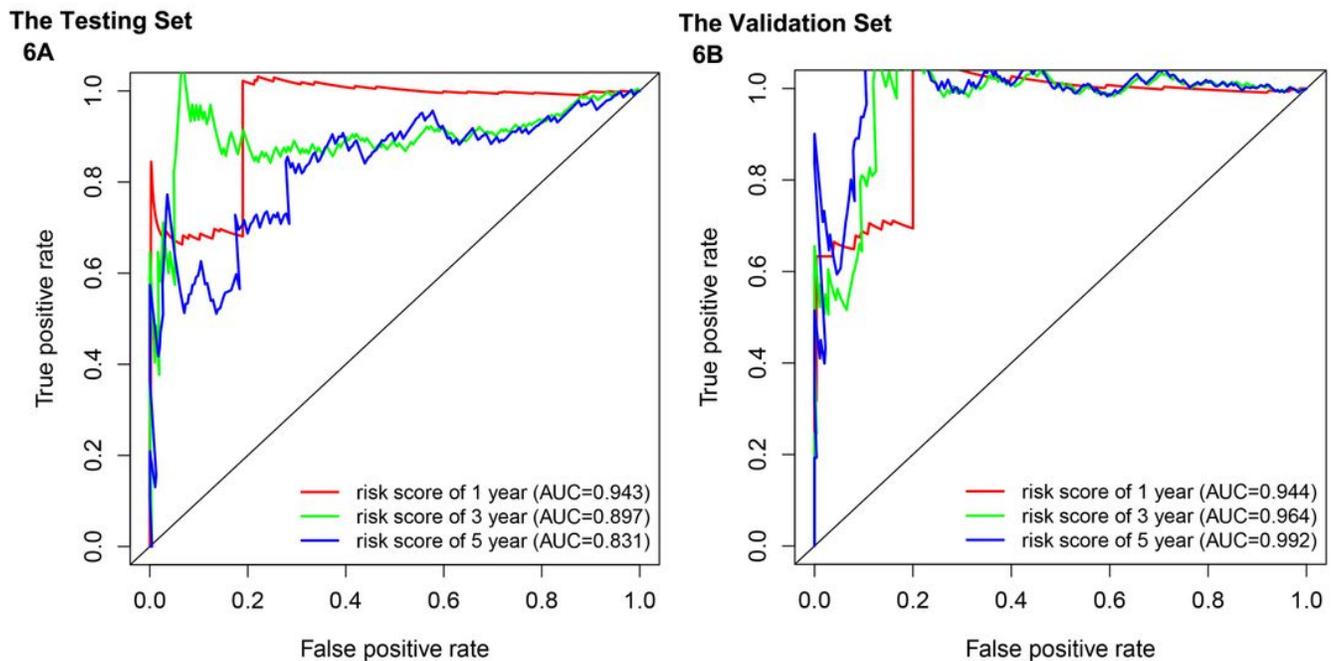


Figure 6

Prognostic indicators based on hypoxia-related genes showed good predictive performance. (A) ROC analysis of overall survival after 1-year, 3 years, 5-years for 11 gene markers in the testing set. (B) ROC analysis of overall survival after 1-year, 3 years, 5-years for 11 gene markers in the validation set. Table 1 Univariate analyses of potential markers for thyroid cancer patients. Univariate analysis was used to determine the assessment of patient outcomes with different clinical variables.

Supplementary Files

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

- [SupplementFig1.jpg](#)
- [Table1.docx](#)