Medicine®

OPEN

Model establishment of prognostic-related immune genes in laryngeal squamous cell carcinoma

Ming Sun, MS[a], Sihan Chen, MS[b], Min Fu, PhD[a],*

Abstract

Background: Laryngeal squamous cell carcinoma (LSCC) is one of the most common malignant tumors of the head and neck in the world. At present, the treatment methods include surgery, radiotherapy, and chemotherapy, but the 5-year survival rate is still not ideal and the quality of life of the patients is low. Due to the relative lack of immunotherapy methods, this study aims to build a risk prediction model of related immune genes, which can be used to effectively predict the prognosis of laryngeal cancer patients, and provide targets for subsequent immunotherapy.

Methods: We collected the 111 cases of laryngeal squamous cell carcinoma and 12 matched normal samples in the The Cancer Genome Atlas Database (TCGA) gene expression quantification database. The differentially expressed related immune genes were screened by R software version 3.5.2. The COX regression model of immune related genes was constructed, and the sensitivity and specificity of the model were evaluated. The risk value was calculated according to the model, and the risk curve was drawn to verify the correlation between related immune genes, risk score, and clinical traits.

Results: We selected 8 immune-related genes that can predict the prognosis of LSCC in a COX regression model and plotted the Kaplan–Meier survival curve. The 5-year survival rate of the high-risk group was 16.5% (95% CI: 0.059–0.459), and that of the low-risk group was 72.9% (95% CI: 0.555–0.956). The area under the receiver operating characteristic (ROC) curve was used to confirm the accuracy of the model (AUG = 0.887). After univariate and multivariate regression analysis, the risk score can be used as an independent risk factor for predicting prognosis. The risk score ($P = .021$) was positively correlated with the clinical Stage classification.

Conclusion: We screened out 8 immune genes related to prognosis: RBP1, TLR2, AQP9, BTC, EPO, STC2, ZAP70, and PLCG1 to construct risk value models, which can be used to speculate the prognosis of the disease and provide new targets for future immunotherapy.

Abbreviations: HR = hazard ratio, LSCC = laryngeal squamous cell carcinoma, ROC = receiver operating characteristic, TCGA = The Cancer Genome Atlas Database, TF = transcription factor.

Keywords: Cox regression model, laryngeal squamous cell carcinoma, prognostic-related immune genes introduction, risk score, The Cancer Genome Atlas Database

1. Introduction

Laryngeal squamous cell carcinoma (LSCC) is the most common cancer of the head and neck. The main risk factors for laryngeal cancer include tobacco, human papillomavirus infection, laryngopharyngeal reflux, environmental and occupational exposure, and alcohol.[1] Despite tremendous progress in the treatment and research of the disease, the disease's survival rate and quality of life are still not optimistic.[2,3] With

the development of tumor molecular genetics, many scholars predict the prognosis of tumors using genetics.[4] Studies have shown that the up-regulation and down-regulation of the expression of immune-related genes in tumor cells may be correlated with tumor prognosis. Identifying patients with high risk scores enables more targeted clinical treatments. The availability of large amounts of genomic data makes statistical modeling a promising approach to predicting patient prognosis.

This paper aims to find a differential immune gene model related to prognosis to predict the survival rate of LSCC. The interaction network between differential immune-genes and TF, as well as the correlation between risk models and clinical characteristics, are explored. Directions for follow-up research are provided.

2. Materials and methods

2.1. Work flow

First, the differentially expressed immune genes were screened out after sorting out the database. The COX regression model of immune-related genes was constructed. Next, the survival curve and the interaction network with transcription factors were drawn, and finally, the risk value of the research samples was evaluated and the correlation of clinical indicators was analyzed (Fig. 1).

2.2. Download and organize the database

We downloaded and collected gene expression quantification data and clinical data of LSCC from the TCGA database[5] (http://portal.gdc.cancer.gov). We got the mRNA data set from the GENECODE[6] (https://www.gencodegenes.org), perform integration processing of transcriptome data and extract corresponding clinically relevant information. Data sets of immune related genes were obtained from the Immport[7] (Https://www.immport.org) and TF data set from the Cistrome database[8] (Https://cistrome.org).

2.3. Difference analysis of transcriptome genes and immune-related genes

A Wilcox test was performed on the obtained database using R software. Differentially expressed genes were then screened out. The Pheatmap package in R software was used to draw differential gene volcano maps and heatmaps. Among the screened differentially expressed genes, the above method was used to screen out differentially expressed immune-related genes.

2.4. Screening of immune genes related to prognosis

COX regression analysis was performed on the differential immune-related genes and clinical prognostic information to screen out the immune genes related to prognosis. The forest map is drawn using the R software's survival package.

2.5. Transcription factors difference analysis

A total of 318 immune gene transcription factors were downloaded from Cistrome. The TF data set was analyzed via Wilcox test using R software. The differentially expressed TF were screened out among the different genes. The volcano map and heat map of differential transcription factors were drawn.

2.6. Construction of regulatory networks of TF and prognosis-related immune genes

The correlation between differentially expressed transcription factors and prognosis-related immune genes was analyzed, and their interaction regulatory network was drawn by Cytoscape software.[9]

2.7. Construction of a prognostic-related immune gene model

We selected the immune genes associated with prognosis and calculated the patient's risk value as $\sum$(ExpmRNA1-n×coefmRNA1-n). "Exp" represents the expression of the gene, and "coef" represents the correlation coefficient of the gene.

2.8. Drawing Kaplan–Meier survival curve and ROC curve

The survival curve stakes the high-risk group with the low-risk group using the "survival" and "survminer" packages of the R software, and the accuracy of the model is evaluated using the ROC curve.

2.9. Drawing risk curve

The risk value curve shows the growth trend of the risk value, and we draw a scatter plot with the increase of the patient's risk value as the x axis and the survival time as the y axis. The expression of immune-related genes in the model is indicated by a heatmap.

2.10. Independent prognosis analysis

An independent prognosis analysis using single and multi-factors to determine whether risk score can be used as an independent factor for predicting prognosis, including sex, age, stage, grade, T, N stage, and risk score.

2.11. Clinical correlation analysis

The clinical data were transformed into binary variables to find out the correlation between individual immune genes and risk values involved in the construction of the model and the corresponding clinical traits.

3. Results

3.1. Visualization of differential gene data

The obtained database contains 111 tumor tissues and 12 normal tissues, and 5494 differentially expressed genes were screened ($P < .05$, FC value ≥ 2, | logFC | > 1). Of these, 4624 genes were upregulated and 872 genes were down-regulated (Fig. 2A and B). Among the screened differentially expressed genes, 432 differentially expressed immune-related genes were intersected, including 371 upregulated genes and 61 down-regulated genes (Fig. 2C and D). Using the same method, 65 differentially expressed TF were intersected among the screened differentially expressed genes, including 10 upregulated TF and 55 down-regulated TF (Fig. 2E and F). By COX regression analysis between differential
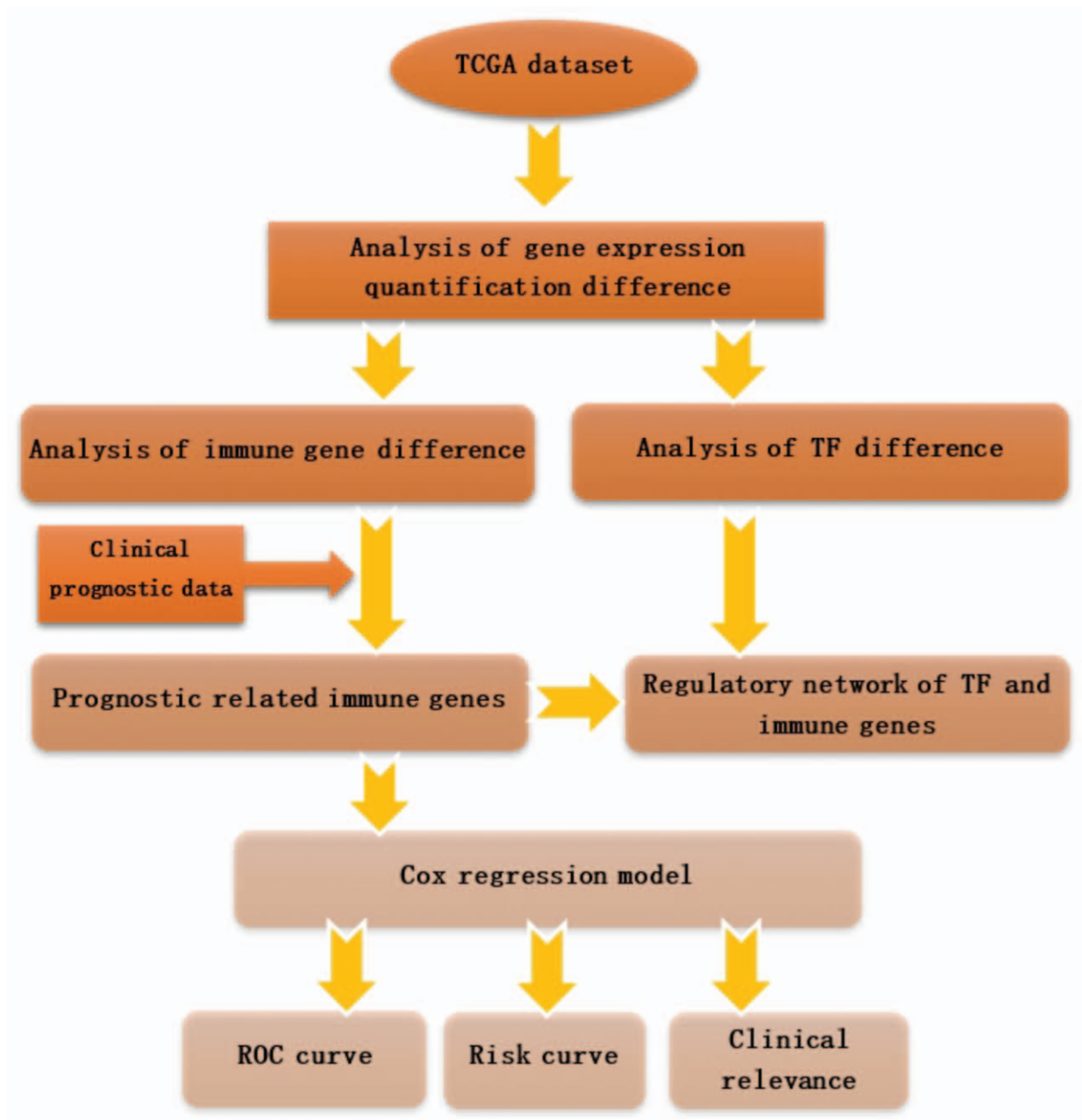
Figure 1. Main workflow for the study.

immune genes and clinical prognostic data ($P < .05$, FC≥2), 13 immune genes related to prognosis were screened out ($P < .015$). Hazard ratio (HR) > 1 is the prognostic risk factor, $0 < HR < 1$ is the prognostic protection factor, among them 5 protection factors and 8 risk factors were used to draw the forest map (Fig. 3A).

3.2. Interaction network of TF and prognosis-related immune genes

The correlation between differential TF and 13 prognostic immune genes was studied. All of the cor >0, indicating that the TF and immune genes are all positively regulated. The interaction network is shown in Fig. 3B

3.3. Prognosis related immune gene model

We selected 8 of 13 immune genes related to prognosis ($P < .015$) for model construction (Table 1), among which risk genes were RBP1, TLR2, AQP9, BTC, and STC2; protective genes were EPO, ZAP70, PLCG1. We calculated the sample risk value and drew the Kaplan–Meier survival curve (Fig. 4A). Patients were grouped according to the median of the risk value, with patients above the median classified as high-risk and patients below the median classified as low-risk. The survival rate of the low-risk group decreased ($P < .001$), and the area under the ROC curve was AUG=0.889 (Fig. 4B), which proved that the modified model had accurately evaluated the prognosis.
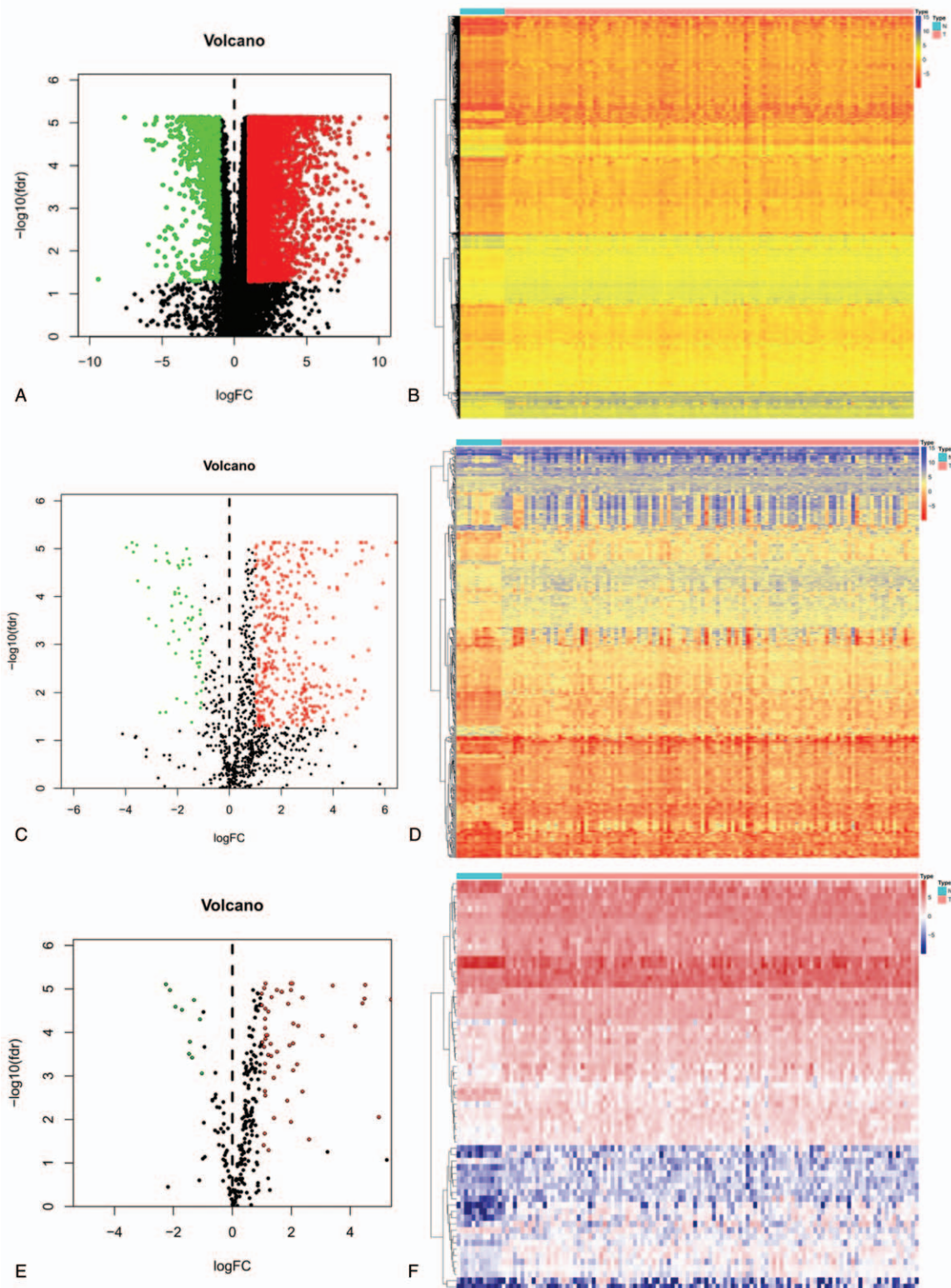
Figure 2. Volcano map and heatmap of differential genes (A, B). Differential immune genes (C, D) and differential TF (E, F): red represents upregulated genes, green represents down-regulated genes. (Heatmap: The expression level of each differential gene in tissues). TF = transcription factor.
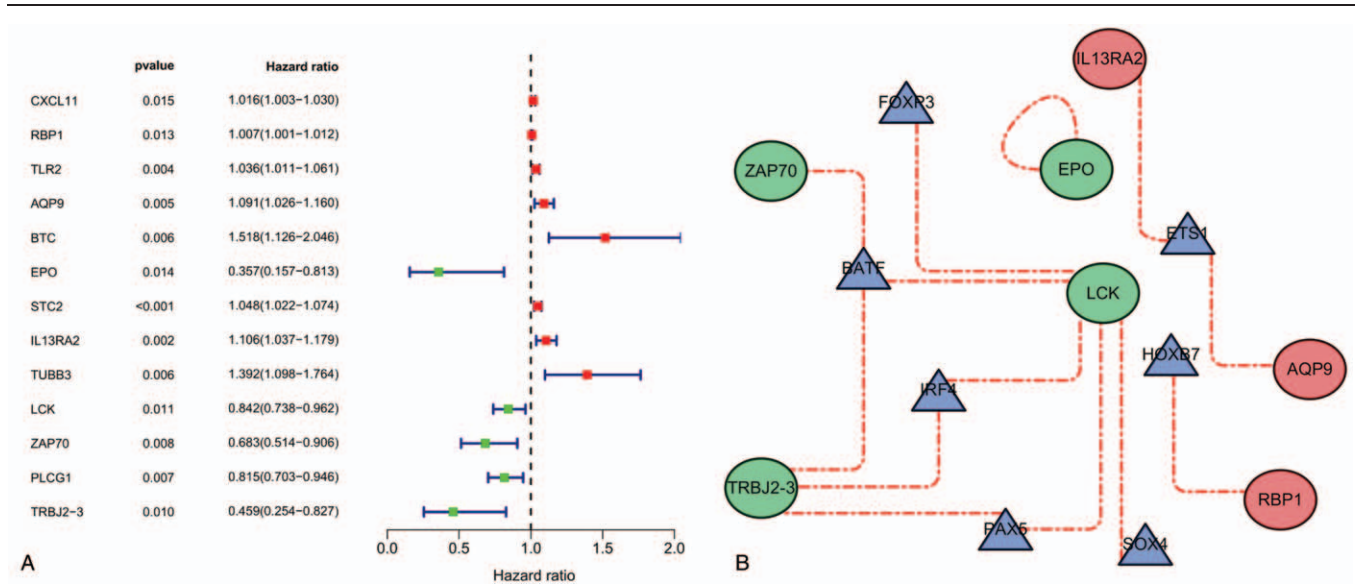
Figure 3. (A) Forest map of prognosis-related immune genes: red represents prognostic risk genes, green represents prognostic protection genes. (B) Interaction network of TF and prognosis-related immune genes: triangles represent TF, ellipse represents immune gene, red represents risk genes, green represents protection genes. TF = transcription factor.

Table 1

Prognosis-related immune genes in the model.

ID	Coef	HR	HR.95L	HR.95H	P value
RBP1	0.012	1.012	1.007	1.017	.000
TLR2	0.045	1.046	1.017	1.076	.002
AQP9	0.075	1.078	1.015	1.146	.014
BTC	0.580	1.787	1.245	2.563	.001
EPO	−0.931	0.394	0.169	0.919	.031
STC2	0.038	1.039	1.005	1.074	.024
ZAP70	−0.228	0.796	0.574	1.102	.169
PLCG1	−0.256	0.774	0.661	0.906	.001

HR = hazard ratio.
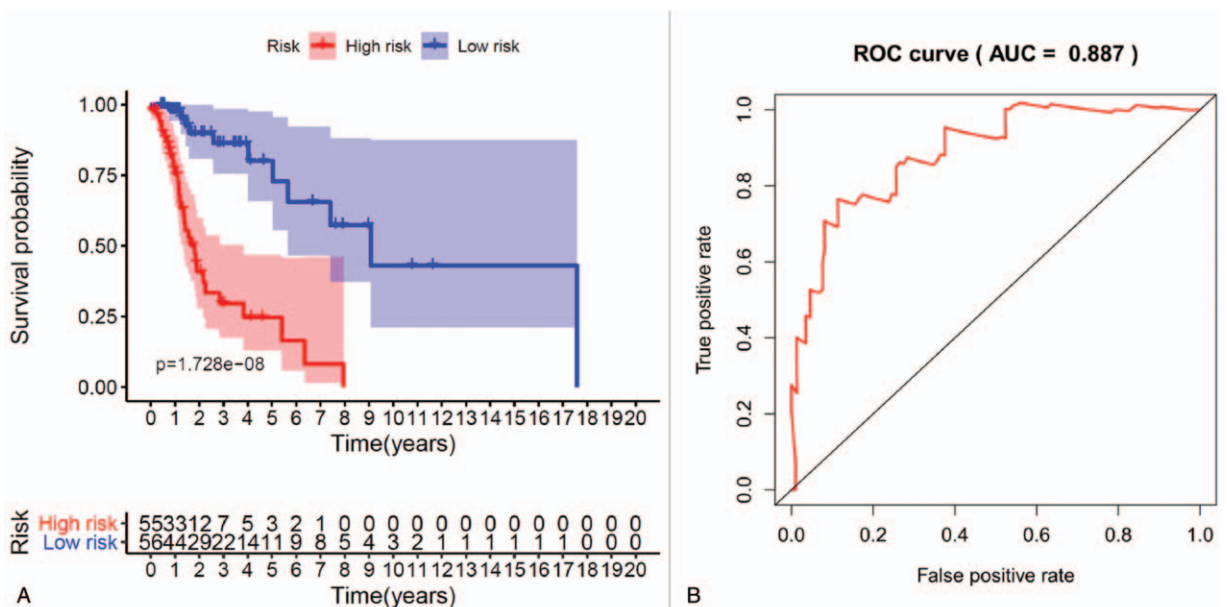
Figure 4. (A) Kaplan–Meier survival curve: it indicates the relationship between risk grade and survival rate. (Red represents high risk, blue represents low risk). (B) ROC curve: AUC = 0.887. ROC = receiver operating characteristic.
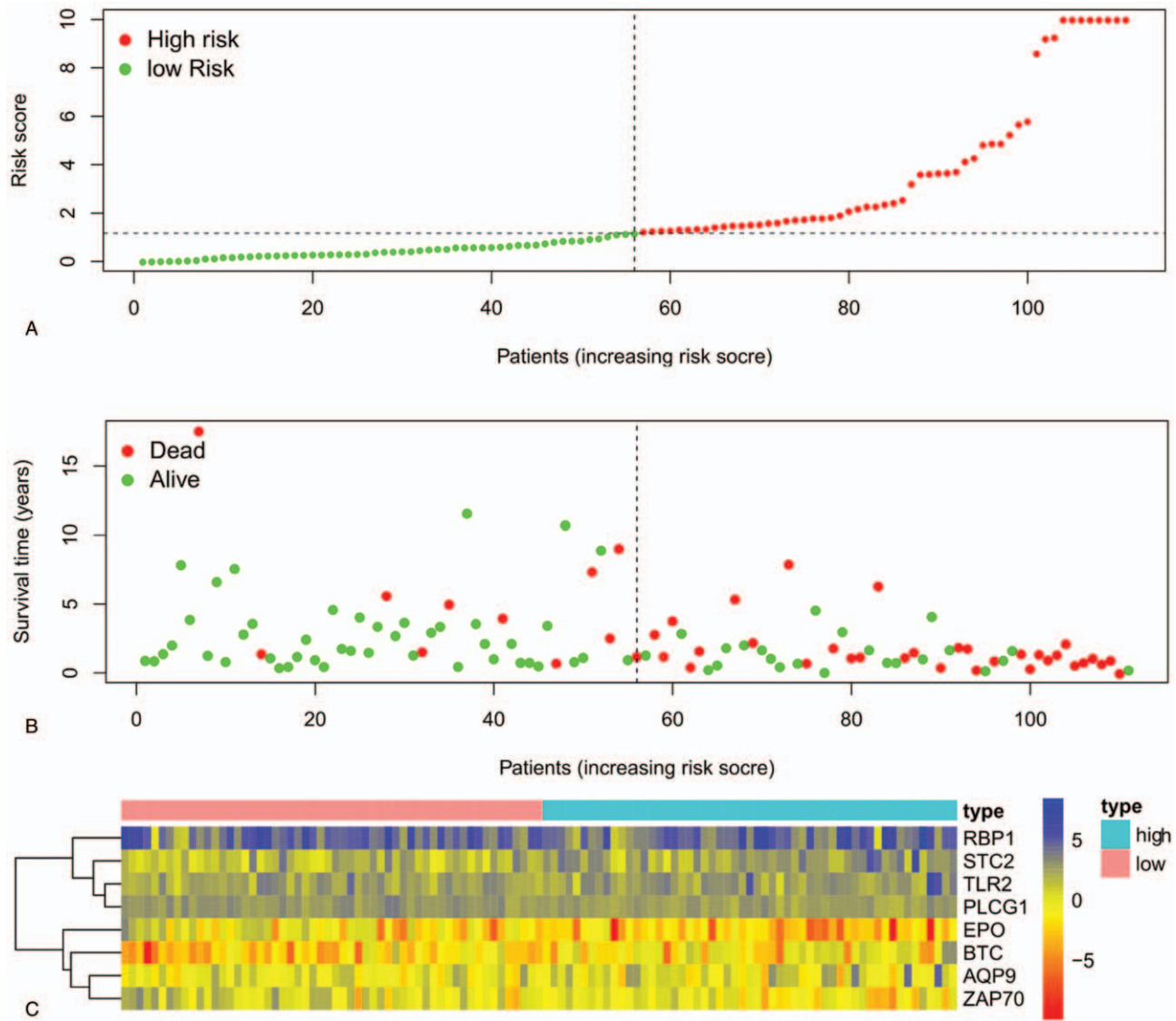
Figure 5. (A) Risk curve, (B) scatterplot of survival status, (C) differential expression of genes in high and low risk groups in the model.
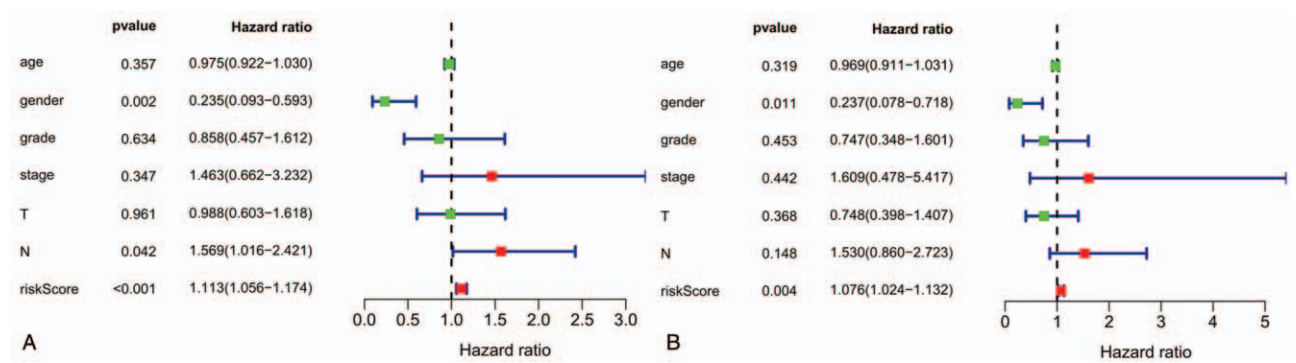

Figure 6. (A) Univariate independent prognostic analysis. (B) Multivariate independent prognostic analysis. Gender: female=0, male=1.
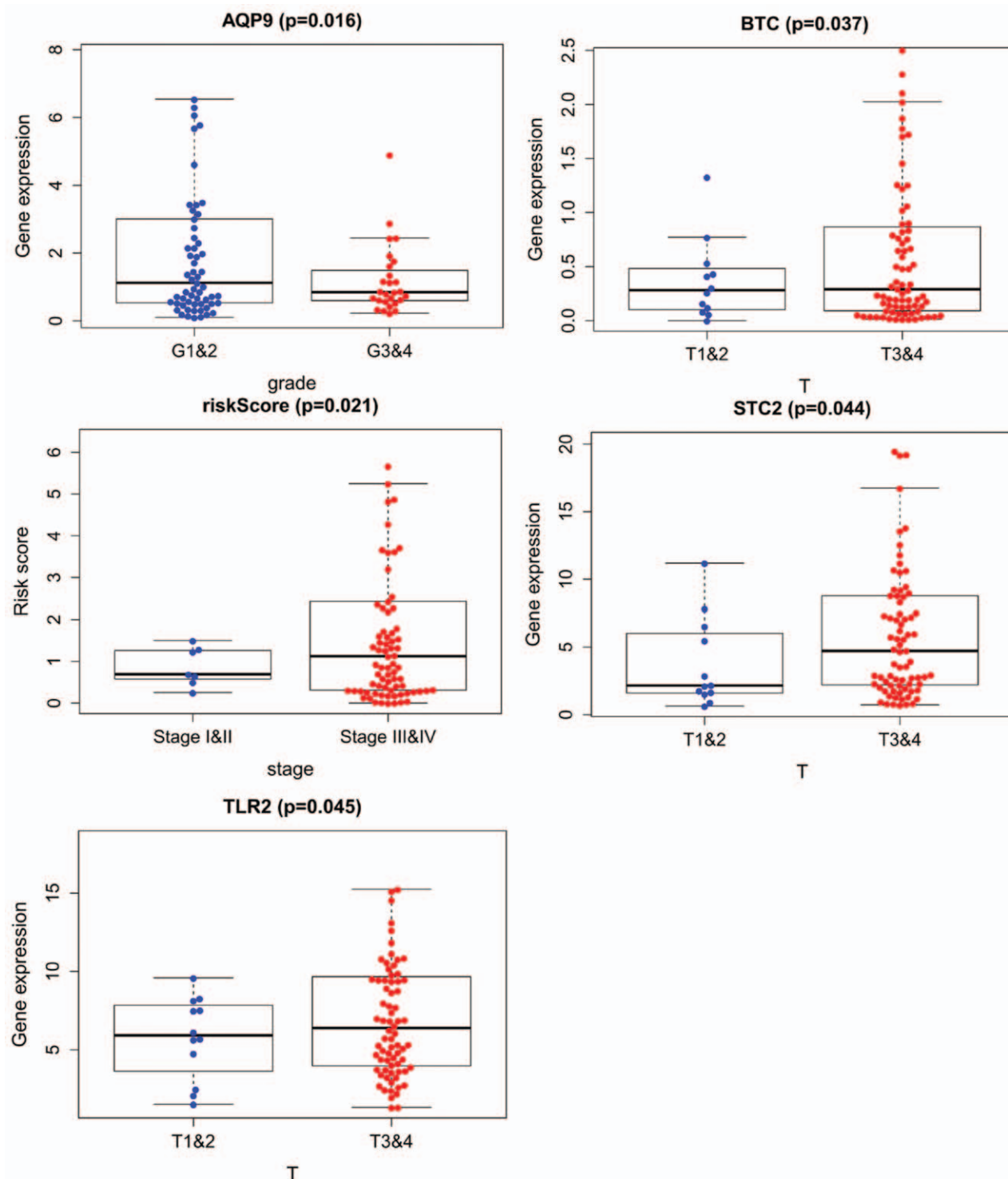
Figure 7. BTC (*P* = .037), STC2 (*P* = .044), and TLR2 (*P* = .045) showed the gene expression was positively correlated with T stage. AQP9 (*P* = .016) was negatively correlated with grade. The risk score (*P* = .021) was positively correlated with Stage.

3.4. Risk curve

The patient's risk value is sorted from low to high. The survival status and survival time of the 2 groups of patients are represented by scatter plots (Fig. 5B). The difference in prognosis-related immune gene expression between the 2 groups is represented by a heatmap (Fig. 5C).

3.5. Independent prognosis analysis

We performed univariate independent prognostic analysis of the prognosis of LSCC by sex, age, stage, grade, T, N stage, and risk scores (Fig. 6A and B): women have a better prognosis than men (*P* = .011). High risk scores (*P* < .001) and lymph node metastasis (*P* = .042) are independent risk factors for prognosis. After

multivariate regression analysis of all indicators, the risk score (P=.004) and sex (P=.011) can still independently predict the prognosis.

3.6. Clinical correlation analysis

The clinical data were transformed into binary variables to find out the correlation between the individual immune genes in the model and the clinical traits (Fig. 7). BTC (P=.037), STC2 (P=.044), and TLR2 (P=.045) showed the gene expression was positively correlated with T stage. The gene expression in T III–IV grade was higher than that in T I–II grade. AQP9 (P=.016) was negatively correlated with grade and the expression of this gene in Grade 3 and 4 was lower than that in Grade 1 and 2. The risk score (P=.021) was positively correlated with Stage, the gene expression level of Stage III and IV was higher than that of Stage I and II.

4. Discussion

In the past 10 years, great progress has been made in the research and treatment of LSCC.[1] However, treatment resistance will increase the recurrence rate of LSCC, and patients undergoing total laryngeal surgery have a worse quality of life.[4] This paper screened out differentially expressed genes to provide insights for subsequent research, including targets for immunotherapy. The progression and metastasis of LSCC are related to the interaction between tumor cells and surrounding cells, which is caused by the change of adhesion molecules between cells and cells.[10–12] There is an inseparable connection between tumor cells and immune cells for surveillance.[13] Some immune genes are differentially expressed in normal tissues and tumor tissues, and some are correlated with prognosis. Therefore, we screened these differentially expressed immune genes, and then we can carry out risk assessment for patients with LSCC.

Our analysis identified 8 prognostic-related immune genes to construct a risk prediction model. Among them, TLR2 was once considered to be a molecule that initiates the activation of NLRP3 inflammatory bodies and is considered to be related to asthma. TLR2 deficiency can effectively inhibit airway inflammation and reduce the production of related immunoglobulin (Ig E) and inflammatory cells.[14,15] AQP9 is upregulated in a variety of cancer tissues, and some scholars conducted in vitro cell experiments on prostate cancer, knocking out AQP9 gene and found to accelerate the apoptosis of androgen-independent prostate cancer cells.[16,17] Some scholars have found that silent HOTAIR may down-regulate STC2 by competitively binding to miR-206, thereby inhibiting the biological function of head and neck squamous cell carcinoma.[18] Inhibition of STC2 is also thought to hinder the proliferation, invasion, and metastasis of glioblastoma cells.[19] EPO is considered to be related to tumor proliferation, metastasis, and drug resistance.[20] There are EPO receptors in most tumor cells, but there are studies saying that there is no specific reason to believe that EPO has a role in tumor development, and its control of tumor cell proliferation is very gentle.[21] In our study, EPO as a protective factor of LSCC its role still needs to be proved by follow-up experiments. RBP1 is considered to be related to bladder cancer, tongue cancer, breast cancer, and other cancers. It is reported that RBP1 affects retinoic acid metabolism and leads to cell differentiation and tumor progression by reducing retinol transport and preventing retinol ester formation and nuclear retinoic acid receptor (RARs) activity.[22] BTC is one of the least-known ligands among endogenous EGFR ligands. BTC promotes wound healing better by promoting cell migration, but it reduces receptor affinity and attenuates the induction of EGFR phosphorylation, which is thought to be related to cancer progression.[23] Some genes in our model have been discovered and studied, but some genes still need follow-up exploration and cell experiment support.

Research limitations: there is only one patient with distant metastasis in this sample, and the extreme value will cause inaccurate research results, so the risk score is not precise for those patients with distant metastasis. The subsequent expansion of the sample size will enable the risk model to more widely predict the prognosis of various laryngeal cancer patients.

In summary, by mining the database, we constructed a differential immune gene model related to prognosis and estimated the survival rate of patients with LSCC. In the future, we hope to verify and improve our results through clinical in vitro or in vivo experiments.

5. Conclusions

In our study, the immune gene model related to prognosis effectively predicts the prognosis of patients with LSCC. The interaction network with transcription factor (TF) clearly expresses its associated pathway, providing a direction for follow-up molecular research and laying a molecular foundation for immunotherapy.

Author contributions

Conceptualization: Sihan Chen, Min Fu.
Data curation: Ming Sun, Sihan Chen.
Formal analysis: Min Fu.
Funding acquisition: Min Fu.
Project administration: Min Fu.
Visualization: Ming Sun, Sihan Chen.
Writing – original draft: Ming Sun.
Writing – review & editing: Ming Sun, Min Fu.

Correction

The sentence "There is no fund support for this project" originally appeared in the footnote of the article incorrectly and has since been removed.

References

[1] Obid R, Redlich M, Tomeh C. The treatment of laryngeal cancer. Oral Maxillofac Surg Clin North Am 2019;31:1–1.

[2] Gatta G, Botta L, Sanchez MJ, et al. Prognoses and improvement for head and neck cancers diagnosed in europe in early 2000s: The eurocare-5 population-based study. Eur J Cancer 2015;51:2130–43.

[3] Cortesina G, Martone T. Molecular metastases markers in head and neck squamous cell carcinoma: review of the literature. Acta Otorhinolaryngol Ital 2006;26:317–25.

[4] Erkul E, Yilmaz I, Gungor A, et al. Microrna-21 in laryngeal squamous cell carcinoma: diagnostic and prognostic features. Laryngoscope 2017;127:E62–6.

[5] Sharma S, Ciufo S, Starchenko E, et al. The NCBI BioCollections Database. Database (Oxford) 2018;2018:bay006.

[6] Li B, Severson E, Pignon JC, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. Genome Biol 2016;17:174.

[7] Bhattacharya S, Dunn P. Immport, toward repurposing of open access immunological assay data for translational and clinical research. Sci Data 2018;5:180015.

[8] Zheng R, Wan C, Mei S, et al. Cistrome data browser: expanded datasets and new tools for gene regulatory analysis. Nucleic Acids Res 2019;47:D729–35.

[9] Otasek D, Morris JH, Boucas J, et al. Cytoscape automation: empowering workflow-based network analysis. Genome Biol 2019;20: 185.

[10] Hagedorn H, Schreiner M, Wiest I, et al. Defective basement membrane in laryngeal carcinomas with heterogeneous loss of distinct components. Hum Pathol 1998;29:447–54.

[11] Antonelli AR, Nicolai P, Cappiello J, et al. Basement membrane components in normal, dysplastic, neoplastic laryngeal tissue and metastatic lymph nodes. Acta Otolaryngol 1991;111:437–43.

[12] Visser R, van der Beek JM, Havenith MG, et al. Immunocytochemical detection of basement membrane antigens in the histopathological evaluation of laryngeal dysplasia and neoplasia. Histopathology 1986;10:171–80.

[13] Fridman WH, Pages F, Sautes-Fridman C, et al. The immune contexture in human tumours: impact on clinical outcome. Nat Rev Cancer 2012;12:298–306.

[14] Wu HM, Zhao CC, Xie QM, et al. Tlr2-melatonin feedback loop regulates the activation of nlrp3 inflammasome in murine allergic airway inflammation. Front Immunol 2020;11:172.

[15] Dai MY, Chen FF, Wang Y, et al. Particulate matters induce acute exacerbation of allergic airway inflammation via the tlr2/nf-kappab/ nlrp3 signaling pathway. Toxicol Lett 2020;321:146–54.

[16] Lv Y, Huang Q, Dai W, et al. Aqp9 promotes astrocytoma cell invasion and motility via the akt pathway. Oncol Lett 2018;16:6059–64.

[17] Chen Q, Zhu L, Zheng B, et al. Effect of aqp9 expression in androgen-independent prostate cancer cell pc3. Int J Mol Sci 2016;17:738.

[18] Li T, Qin Y, Zhen Z, et al. Long non-coding rna hotair/microrna-206 sponge regulates stc2 and further influences cell biological functions in head and neck squamous cell carcinoma. Cell Prolif 2019;52:e12651.

[19] Feng L, Ma J, Ji H, et al. Mir-184 retarded the proliferation, invasiveness and migration of glioblastoma cells by repressing stanniocalcin-2. Pathol Oncol Res 2018;24:853–60.

[20] Les D, Sadus-Wojciechowska M, Rutkowski T, et al. The endogenous erythropoietin in correlation with other erythrocytic parameters in patients with head and neck squamous cell carcinoma treated with platinum-based induction chemotherapy. Contemp Oncol (Pozn) 2019;23:178–82.

[21] Fandrey J. Erythropoietin receptors on tumor cells: what do they mean? Oncologist 2008;13(suppl):16–20.

[22] Gao L, Wang Q, Ren W, et al. The RBP1-CKAP4 axis activates oncogenic autophagy and promotes cancer progression in oral squamous cell carcinoma. Cell Death Dis 2020;11:488.

[23] Rush JS, Peterson JL, Ceresa BP. Betacellulin (BTC) biases the EGFR to dimerize with ErbB3. Mol Pharmacol 2018;94:1382–90.