Establishment of a risk model by integrating hypoxia genes in predicting prognosis of esophageal squamous cell carcinoma

Wanyi Xiao
Tianjin Cancer Institute: Tianjin Tumor Hospital
Peng Tang
Tianjin Cancer Institute: Tianjin Tumor Hospital
Zhilin Sui
Tianjin Cancer Institute: Tianjin Tumor Hospital
Xianxian Wu
Cancer Hospital and Shenzhen Hospital
Yueyang Yang
Tianjin Cancer Institute: Tianjin Tumor Hospital
Ningning Zhu
Tianjin Cancer Institute: Tianjin Tumor Hospital
Youming Han
Tianjin Cancer Institute: Tianjin Tumor Hospital
Lei Gong
Tianjin Cancer Institute: Tianjin Tumor Hospital
Zhentao Yu (yuzhentao@tjmuch.com)
Tianjin Cancer Institute: Tianjin Tumor Hospital https://orcid.org/0000-0002-5785-8492
Hongdian Zhang
Tianjin Cancer Institute: Tianjin Tumor Hospital

Primary research

Keywords: esophageal squamous cell carcinoma, hypoxia, hypoxia-related genes, nomogram, prognosis

Posted Date: March 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1008316/v3

Abstract

Esophageal squamous cell carcinoma (ESCC) has a dismal prognosis and hypoxia plays a key role in metastasis and proliferation of ESCC. Thus, we aimed to develop a hypoxia-based gene signature to assist in the diagnosis and prognosis. We performed consensus clustering analysis based on differentially expressed hypoxia related genes (HRGs) acquired from the GSE53625 dataset and The Cancer Genome Atlas (TCGA), and used weighted gene co-expression network analysis to filter out candidate modules which were then intersected with differentially expressed genes from clustered subgroups to obtain HRGs. The HRGs were used for stepAIC algorithm to construct risk score models and validated in TCGA database. Independent prognostic factors after univariate and multivariate COX analyses were used to construct the prognostic nomogram. Immunohistochemical were used to detect protein expression levels of relevant genes. And the relationship between risk scores and tumor microenvironment was explored. A hypoxia risk model containing 6 genes (PNPLA1, CARD18, IL-18, SLC37A2, ADAMTS18 and FAM83C) was constructed by screening key HRGs. Poorer prognosis in the high-risk group than in the low-risk group. And Cox regression analysis showed that risk score was independent prognostic factors. Nomogram based on risk scores could well predict 1-, 3-, and 5-year survival. P53, Wnt, and hypoxia signaling pathways may be some regulatory mechanisms of hypoxia associated with tumor microenvironment. In addition, we confirmed the high expression of BGN and low expression of IL-18 in ESCC tissues. Our study determined the prognostic value of a 6-hypoxia gene signature and a prognostic model, providing potential prognostic predictors and therapeutic targets for ESCC.

Introduction

Esophageal squamous cell carcinoma (ESCC) is the predominant type of esophageal cancer in China, accounting for more than 90% of pathological types(1). Due to the heterogeneous biological characteristics of ESCC(2), there are various mechanisms that promote tumor progression leading to poor prognosis.

Rapid growth and massive angiogenesis increase oxygen consumption(3), resulting in hypoxia. Hypoxia is closely related to abnormal biological behaviors, leading to aberrant gene expression changes(4), anti-apoptosis(5), suboptimal treatment outcomes(6) and ultimately to poor prognosis as tumor cells adapt to hypoxia(7). It has been found that ESCC tissues are invariably hypoxic. Hypoxia is involved in the migration and progression of ESCC and is associated with increased malignancy and poor prognosis(8). And hypoxia plays a key role in subsequent proteomic changes were reported(9). Therefore, focusing on hypoxia related markers may provide an efficient way to identify specific patient groups.

With advances in bioinformatics and high-throughput technologies(10), analyses that target the regulation networks are improving the current understanding of the molecular mechanisms. Since tumor hypoxia cannot be predicted based on clinical size, stage or differentiation, molecular biomarkers that can assess the hypoxic status of ESCC at an early stage and predict the prognosis are needed.

It has been reported that hypoxia and the tumor immune microenvironment also interact with each other(11). The adaptation of tumor cells to the hypoxic environment leads not only to aberrant gene expression but also to the formation of the tumor immune microenvironment. Exploring the effects of hypoxia on both and the potential remodeling effects will help reveal the complex interactions among them and provide new clues for the treatment of ESCC.

In this study, we aimed to identify and validate hypoxia-related genes (HRGs) in ESCC and develop a nomogram based on the HRG signature to assist in the prognostic risk prediction of patients and to facilitate personalized treatment.

Materials And Methods

2.1 Data acquisition and processing

We downloaded clinical information including follow-up data of 179 ESCC and 179 paired normal control samples from the Gene Expression Omnibus (GEO, GSE53625 dataset) based on GPL18109. Additionally, the RNA-FPKM data and clinical data of 82 ESCC samples were retrieved using the The Cancer Genome Atlas (TCGA) data portal. Corresponding normal samples included 11 TCGA paracancerous samples and 1445 samples of 54 noncancer sites in GTEx. The median value was calculated when more than one expression file had matched patients.

Tissue microarrays (TMAs) contained samples from 232 ESCC tissues and 58 adjacent normal esophageal tissues between 2009 and 2010 at Tianjin Medical University Cancer Institute and Hospital were obtained. Patients who had received neoadjuvant chemotherapy or radiotherapy and those with cancer types other than ESCC were excluded. Ethical approval was obtained from the Research Ethics Committee of Tianjin Medical University Cancer Institute and Hospital.

2.2 Extraction of HRGs

The list of 112 hypoxia genes was retrieved by enquiring “hypoxia” from The Cancer Single-Cell State Atlas (CancerSEA)(12) of all cancer types. The “HALLMARK-HYPOXIA” gene set was derived from The Molecular Signatures Database (MSigDB)(13). A total of 270 unique elements were identified as key genes involved in hypoxia activity.

2.3 Differential analysis of gene expression

Differentially expressed genes (DEGs) were determined using the “limma” package in R software. Multiple comparisons were adjusted using the Benjamini–Hochberg false discovery rate (FDR). An FDR < 0.05 and a log2 |fold change| > 2 were deemed as cutoff values.

2.4 Identification of hypoxia subgroups

Euclidean-based consensus clustering was performed on the GSE53625 dataset using the ConsensusClusterPlus R package. The clustering process was performed 500 times with each iteration containing 80% of the samples and the number of clusters set to 2–10. And the graphical output results included heatmaps of the consensus matrices, which displayed the clustering results, consensus cumulative distribution function (CDF) plots, and Delta area plots. The optimal number of clusters is usually chosen as the value of K corresponding to the last inflection point of the CDF and the smallest slope of Delta.

2.5 Weighted gene coexpression network analysis (WGCNA)(14)

The GSE53625 gene expression file for 179 ESCC samples was used to construct a scale-free network using the R package “WGCNA”. We set the minimal module size to 20 and the cut height to 0.25.

2.6 Construction and evaluation of the prognosis prediction model

The stepAIC algorithm running in the R “MASS” package was used to construct a optimal prognostic model, based on the combination of expression profiles that intersected from selected modules and DEGs and prognostic information from 179 ESCC samples. The risk score of each sample was divided into two groups according to the median value after sorting. Half of the samples were randomly taken as the training group to construct the model. While the other half (internal validation set) and all of the GSE53625 datasets were used as testing datasets to assess the robustness of the model. The same coefficient was used for the external validation dataset - TCGA. The area under the curve (AUC) of the receiver operating characteristic (ROC) curves was calculated using the “time ROC R” package.

2.7 Determining the classification features using Cox proportional risk regression models

In this study, significant prognostic variables obtained from the uni- and multivariate Cox regression model were then introduced into the final nomogram model. Calibration curves were plotted and higher overlap with the 45-degree standard curve indicated better predictive agreement. The “rms”, “foreign” as well as “survival” packages were used for nomogram construction and calibration curve plotting.

2.8 Protein-protein interaction (PPI) network and functional analysis

STRING website (http://string-db.org/) were used to explore the protein interaction relationship, and “clusterProfiler” R package were used for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis and for Gene Ontology (GO) analysis.

2.9 Implementation of gene set enrichment analysis (GSEA)

GSEA was conducted using GSEA software employing GSE53625 data. $P < 0.05$ and a Q value less than 0.25 were considered to denote significant enrichment.

2.10 The ESTIMATE and CIBERSORT algorithms

The stromal, immune and ESTIMATE score for each patient was calculated through the R “estimate” package. The fraction of 22 immune cell types was assessed through cell type identification (CIBERSORT; https://cibersort.stanford.edu/).

2.11 Immunohistochemical (IHC) analysis

TMAs were used for IHC to examine the protein expression level of selected HRGs. In brief, the tissue sections were deparaffinized (70℃, 2 hours), rehydrated and subjected to antigen repair with heated antigen retrieval solution (10 mmol/l sodium citrate buffer, pH 6.0) (100℃, 10 minutes). The activity of endogenous peroxidase was blocked with 0.3% H_2O_2 and 5% goat serum. Then, the sections were incubated with primary antibodies (1:50, BGN, Proteintech; 1:100, IL-18, ORIGENE) overnight at 4°C and then incubated with a biotinylated secondary antibody for 20 minutes at room temperature. Diaminobenzidine (Zhongshan Inc.) was used as a chromogen and produced a brown color, and then samples were counterstained with hematoxylin.

Two experienced pathologists who were blinded to the clinical data scored the staining results. Staining was assessed according to the intensity of staining (no staining, 0; weak, 1; moderate, 2; and strong, 3) and the percentage of positively stained tumor cells (0%, 0; 1 to 30% positive, 1; 31% to 70% positive, 2; 71-100% positive, 3). A total score of 0 to 9 was obtained by multiplying the results of the staining intensity and staining percentage scores, and tissues with a total score of 0 to 3 were considered to have low expression, while tissues with a score of 3 to 9 to have high expression.

2.12 Statistical analysis

Bioinformatics analysis and statistical analysis were conducted using R (version 4.0.2). Comparisons between two groups were presented via the Wilcoxon rank-sum test and chi–squared test, while multiple

comparisons were assessed via the Kruskal–Wallis test. The cutoff point of each subgroup was identified by the survminer package in R. Kaplan–Meier curves are presented between different subgroups, followed by the log-rank test. ROC curves for 1-, 3- and 5-year survival were delineated to evaluate the predictive efficacy of the risk score. The *P*-values were corrected by Bonferroni's test. A two-sided $P < 0.05$ was considered statistically significant.

Results

3.1 Screening differentially expressed hypoxia genes

After data preprocessing, 2822 DEGs in the GEO set and 4808 DEGs in TCGA were obtained. These are illustrated in the volcano plots in Figure 1A and 1B and heatmaps in Figure. S1. We then intersected the DEGs with the list of 270 HRGs as shown in the Venn diagrams in Figure 1C, 30 hypoxia genes (15 upregulated and 15 downregulated) were identified (Table S1).

A subsequent PPI network was constructed by uploading the aforementioned 30 target genes (Figure 1D). In the functional analysis of DEGs, the top 6 enriched GO annotations and KEGG pathways in each category were visualized intuitively as Circos plots (Figure 1E). The Kaplan-Meier analysis ($P < 0.05$, Figure 1F) showed that SLC2A1, PGM2, SULT2B1, and CA9 all had a lower risk of death except BGN.

3.2 Identifying distinct subgroups and intercluster prognosis analysis

A total of 179 tumor samples were consistently clustered based on expression of screened hypoxia genes. The consensus matrix (Figure 2A-B) helps us to determine the clearest division when divided into two clusters, named C1 and C2. Compared to other categorical numbers (Figure. S2), the consensus matrix graph corresponding to K = 2 showed that the distribution of 2 blue blocks on the diagonal along the white background was well defined (Figure 2C).

The results of Kaplan-Meier survival analysis revealed significant differences among C1 and C2 ($P = 0.014$, Figure 2D). Sorting the samples by cluster produced the gene expression heatmap (Figure. S3), which indicated the composition and quantity of clustering. Gene expression patterns differed among the subgroups, suggesting the credibility of the 2 structural clusters. A heatmap (Figure 2E) annotated by grade, T stage, N stage, stage, sex, and 5 real hub genes, demonstrated the heterogeneity between the two clusters.

3.3 Correlations between the obtained clusters and immune infiltration

We calculated the levels of 22 immune cell types in each sample and compared their differences in C1 and C2 (Figure 2F). The results showed that CD8 T cells, resting memory CD4 T cells, follicular helper T cells, regulatory T cells, activated NK cells, monocytes, M0 macrophages, activated dendritic cells and activated

mast cells had significantly different infiltration levels in different subgroups. Then, GSEA showed the activation of P53 signaling pathway molecules (Figure 2G).

3.4 Identification of modules associated with hypoxia

After sample clustering to detect outliers, restricted to 178 patients, different power values (1–20) were analyzed (Figure. S4A-E). The green module was significantly associated with consensus subgroups (C1, r = -0.46, *P* = 1e-10; C2, r=0.46, *P* = 1e-10, Figure 3A), suggesting that the module is suitable for identifying the hub genes associated with C1/C2.

3.5 Constructing and evaluating a hypoxia-related prognosis signature

The results of differential expression analysis among the obtained clusters are shown in Figure 3B. The hypoxia gene signature was constructed based on 259 differentially expressed HRGs (Figure 3C), identifying the risk score by using the 6 most relevant genes (Figure 4). The risk score was calculated as follows: risk score = (0.066265*PNPLA1 expression) + (-0.149270*CARD18 expression) + (-0.183367*IL-18 expression) + (-0.037724079*SLC37A2 expression) + (0.119388782*ADAMTS18 expression) + (-0.031834954*FAM83C expression). The samples were assigned a risk score and ordered to determine whether the expression level varied systematically with the risk score (Figure 4A-C). A higher percentage of patient deaths was associated with high-risk patients (*P* = 0.0063). Furthermore, the prognoses differed significantly between the two groups, as shown in Figure 4D. The results of ROC curve are shown in Figure 4E. The AUCs for 1-, 3-, and 5-year prognostic prediction were 0.71, 0.68 and 0.71, respectively, indicating the relatively excellent predictive efficacy of the model.

Then, the established prognostic signature was further validated in the test group, including the internal validation set and all GSE53625 and TCGA datasets. Expression level profiles for the 6 selected genes were obtained from 3 testing group samples, and the risk scores were calculated using the abovementioned method for each patient. Sorting the samples by risk score produced the heatmap shown in Figure 5A-C, Figure. S5A-C and Figure. S6A-C. The Kaplan-Meier survival curves showed that this risk model could effectively distinguish the survival (*P* = 0.0014 for the internal validation set, Figure. S5D, *P* < 0.001 for all datasets, Figure. S6D, and *P* = 0.043 for the TCGA dataset, Figure 5D). In the external validation group, the AUCs of the 1-, 3-, and 5-year OS were 0.64, 0.78, and 0.79, respectively (Figure 5E). Consistent results are presented in Figure. S5E and Figure. S6E.

3.6 Establishment of predictive nomogram

The univariate and multivariate Cox regression analyses were used to evaluate whether the predictive value of the model-incorporated prognostic signature was affected by other clinical factors (Figure 6A-B).

The results indicated that age (P = 0.010, HR = 1.011), stage, location and risk score (P < 0.001, HR = 1.869) were independent prognostic factors. The risk score integrated with age and stage was chosen to construct a nomogram model, as presented in Figure 6C. The calibration plot of the nomogram (Figure 6D) showed better consistency between the predicted OS outcomes and actual observations, indicating a good predictive performance of the hypoxia-related prognostic nomogram.

3.7 Correlation analysis of the risk score with clinicopathological features and immune infiltration

The high- and low-risk groups were closely correlated with the clinical phenotypes, shown by heat maps (Figure 7A). The results (Figure 7B, Table S2) showed that the stromal and immune scores were comparable between subgroups. However, the estimated score was higher in the high-risk group than in the low-risk group, and the difference was statistically significant (P = 0.03). Differences in the infiltration of 22 immune cell subtypes between subgroups are shown in Figure 7C. GSEA showed that the significantly enriched pathways were the P53 signaling pathway (Figure 7D), Wnt pathway (Figure 7E) and hypoxia pathway (Figure 7F). The results reconfirmed that esophageal cancer cells in a hypoxic state could affect the tumor immune microenvironment through an underlying regulatory mechanism.

3.8 Validation of the expression of selected HRGs

To verify the accuracy of the abovementioned HRGs, we further detected the protein expression levels of BGN and IL-18 according to previous publications and antibody availability. The clinical details of the 232 patients involved are presented in Table 1. BGN was highly expressed and mainly localized to the cytoplasm of cancer cells (Figure 8A-D). IL-18 is normally expressed in cancerous tissues, with a significantly higher percentage of expression in normal tissues than in tumor tissues (Figure 8E-H). Moreover, patients with higher BGN expression were predicted to have poorer survival (Figure 8I). In contrast, patients with high expression of IL-18 showed better prognosis (Figure 8J). High BGN expression was notably associated with tumor size, tumor invasion, and lymph node metastasis (P < 0.05 for all) of patients with ESCC (Table 1). For IL-18, no significant correlation between the IL-18 level and clinicopathological factor except lymph node metastasis (P < 0.001), was observed.

Discussion

Adaptation of tumor cells to a hypoxic environment leads to increased aggressiveness and a treatment-resistant tumor phenotype, contributing to a poor prognosis in various cancers(3, 15). Exploring mechanisms of ESCC progression can be beneficial for prognosis prediction, and some specific genomic alterations have also shown that hypoxia can induce upregulation of the expression of some genes, including CA9, VEGF, ADM and AK3(16, 17).

Since the predictive power of single indicators is limited and influenced by confounders, different hypoxia gene signatures have been reported(18, 19). Although 15 previously described hypoxia genes was confirmed in ESCC, the results obtained from 95 tumor paraffin samples may have limitations(20). We aimed to use high-throughput sequencing data and combine bioinformatic methods, constructing the first prognosis signature risk score model ever built containing key HRGs and first nomogram of ESCC that encompasses both clinical attributes and the risk score.

In this study we used publicly available databases from multiple sources for filtering rather than directly with published hypoxia genes, aiming to maximize the correlation of genes with hypoxia and to screen out those genes. Hypoxia genes with prognostic significance were used for consensus clustering and WGCNA, indicating hypoxia exposure and narrowing the scope of the study in an unbiased way.

Our 6-gene hypoxia signature can be used to calculate risk scores, and the classification of patients into high and low risk groups indicates different levels of hypoxia exposure. We developed the first prognostic prediction model with excellent prognostic power and generalizability. The obtained subtypes differed regarding clinical characteristic helps to screen patients before developing a treatment plan.

We also investigated critical features of the tumor hypoxic environment in hopes of providing clues for clinical diagnosis and immunotherapy. Hypoxia induces changes in the proportion of specific immune cells in ESCC. It was also confirmed that esophageal squamous carcinoma cells in hypoxia can affect the tumor immune microenvironment through P53, Wnt, and hypoxia signaling pathways. However, the current evidence is insufficient to elucidate the role of immune cells. The complex interaction between tumor cells and immune cells in hypoxic environments remains to be further explored.

Two of HRGs were selected for experimental validation. Upregulation of BGN was confirmed again. BGN, a member of the family of leucine-rich small proteoglycans (SLRPs), has been detected upregulation in esophageal(21, 22), pancreatic(23), gastric(24) and was correlated with tumor progression and poor prognosis, consistent with the findings of the present study. IL-18, the protein encoded by this gene, is a proinflammatory and immunomodulatory cytokine and a member of the IL-1 family. According to previous studies, IL-18 activation may be a double-edged sword that promotes tumor development and progression(25). In our study, it is suggested that this gene has a tumor suppressive effect. These findings are in line with data from studies on various other cancer types, especially ESCC(26).

The following are limitations of this study: First, from the perspective of data sources, potential selection bias could not be ruled out due to data obtained from the TCGA and GEO databases. The majority of these patients were white, which may result in inconsistent RNA-seq results and poor reproducibility of other races. Second, certain limitations and preferences could not be excluded as these were retrospective analyses. Third, Further in vitro and in vivo studies are needed to investigate the mechanisms of the key genes.

In conclusion, our study identified a 6-gene signature and a prognostic nomogram incorporating the gene signature and clinical prognostic factors associated with tumor prognosis. The constituents of the gene

signature may serve as potential prognostic predictors and provide therapeutic targets with future clinical applications for ESCC. The prognostic nomogram may reliably facilitate individualized treatment and medical decision making.

Declarations

Data Availability

The data described in the manuscript will be made available upon request.

Conflicts of interest

The authors declare no conflicts of interest.

Authors' Contributions

(I) Conception and design: Xiao WY, Tang P, Yu ZT and Zhang HD; (II) Administrative support: Gong L, Yu ZT and Zhang HD; (III) Provision of study materials or patients: Xiao WY, Sui ZL, Wu XX; (IV) Collection and assembly of data: Yang YY, Zhu Ningning, Han YM; (V) Data analysis and interpretation: Xiao WY, Tang P, Zhang HD, Gong L; (VI) Manuscript writing: Xiao WY; (VII) Final approval of manuscript: All authors.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (82002551 and 81772619), Science and Technology Project of Tianjin Municipal Health Commission (KJ20119).

Statement

A preprint has previously been published(27).

References

1. He Z, Ke Y. Precision screening for esophageal squamous cell carcinoma in China. Chin J Cancer Res. 2020;32(6):673-82.

2. Zhang J, Wang JL, Zhang CY, Ma YF, Zhao R, Wang YY. The prognostic role of FZD6 in esophageal squamous cell carcinoma patients. Clin Transl Oncol. 2020;22(7):1172-9.

3. Jing X, Yang F, Shao C, Wei K, Xie M, Shen H, et al. Role of hypoxia in cancer therapy by regulating the tumor microenvironment. Mol Cancer. 2019;18(1):157.

4. Chen F, Chu L, Li J, Shi Y, Xu B, Gu J, et al. Hypoxia induced changes in miRNAs and their target mRNAs in extracellular vesicles of esophageal squamous cancer cells. Thorac Cancer. 2020;11(3):570-80.

5. Wang Y, Xie Z, Lu H. Significance of halofuginone in esophageal squamous carcinoma cell apoptosis through HIF-1alpha-FOXO3a pathway. Life Sci. 2020;257:118104.

6. Macedo-Silva C, Miranda-Goncalves V, Lameirinhas A, Lencart J, Pereira A, Lobo J, et al. JmjC-KDMs KDM3A and KDM6B modulate radioresistance under hypoxic conditions in esophageal squamous cell carcinoma. Cell Death Dis. 2020;11(12):1068.

7. Vaupel P, Mayer A. Hypoxia in cancer: significance and impact on clinical outcome. Cancer Metastasis Rev. 2007;26(2):225-39.

8. Muz B, de la Puente P, Azab F, Azab AK. The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. Hypoxia (Auckl). 2015;3:83-92.

9. Mao Y, Wang Y, Dong L, Zhang Y, Zhang Y, Wang C, et al. Hypoxic exosomes facilitate angiogenesis and metastasis in esophageal squamous cell carcinoma through altering the phenotype and transcriptome of endothelial cells. J Exp Clin Cancer Res. 2019;38(1):389.

10. Liu M, An H, Zhang Y, Sun W, Cheng S, Wang R, et al. Molecular analysis of Chinese oesophageal squamous cell carcinoma identifies novel subtypes associated with distinct clinical outcomes. EBioMedicine. 2020;57:102831.

11. van Uden P, Kenneth NS, Rocha S. Regulation of hypoxia-inducible factor-1alpha by NF-kappaB. Biochem J. 2008;412(3):477-84.

12. Yuan H, Yan M, Zhang G, Liu W, Deng C, Liao G, et al. CancerSEA: a cancer single-cell state atlas. Nucleic Acids Res. 2019;47(D1):D900-D8.

13. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417-25.

14. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.

15. Akanji MA, Rotimi D, Adeyemi OS. Hypoxia-Inducible Factors as an Alternative Source of Treatment Strategy for Cancer. Oxid Med Cell Longev. 2019;2019:8547846.

16. Harris BH, Barberis A, West CM, Buffa FM. Gene Expression Signatures as Biomarkers of Tumour Hypoxia. Clin Oncol (R Coll Radiol). 2015;27(10):547-60.

17. Yang L, West CM. Hypoxia gene expression signatures as predictive biomarkers for personalising radiotherapy. The British journal of radiology. 2019;92(1093):20180036.

18. Shou Y, Yang L, Yang Y, Zhu X, Li F, Xu J. Identification of Signatures of Prognosis Prediction for Melanoma Using a Hypoxia Score. Front Genet. 2020;11:570530.

19. Yang Y, Li Y, Qi R, Zhang L. Constructe a novel 5 hypoxia genes signature for cervical cancer. Cancer Cell Int. 2021;21(1):345.

20. Winther M, Alsner J, Tramm T, Nordsmark M. Hypoxia-regulated gene expression and prognosis in loco-regional gastroesophageal cancer. Acta Oncol. 2013;52(7):1327-35.

21. Zhu YH, Yang F, Zhang SS, Zeng TT, Xie X, Guan XY. High expression of biglycan is associated with poor prognosis in patients with esophageal squamous cell carcinoma. International journal of clinical and experimental pathology. 2013;6(11):2497-505.

22. Tang L, Chen Y, Peng X, Zhou Y, Jiang H, Wang G, et al. Identification and Validation of Potential Pathogenic Genes and Prognostic Markers in ESCC by Integrated Bioinformatics Analysis. Front Genet. 2020;11:521004.

23. Yang L, Cui R, Li Y, Liang K, Ni M, Gu Y. Hypoxia-Induced TGFBI as a Serum Biomarker for Laboratory Diagnosis and Prognosis in Patients with Pancreatic Ductal Adenocarcinoma. Lab Med. 2020;51(4):352-61.

24. Liu Z, Liu S, Guo J, Sun L, Wang S, Wang Y, et al. Identification and Analysis of Key Genes Driving Gastric Cancer Through Bioinformatics. Genet Test Mol Biomarkers. 2021;25(1):1-11.

25. Saetang J, Chonpathompikunlert P, Sretrirutchai S, Roongsawang N, Kayasut K, Voravuthikunchai SP, et al. Anti-cancer effect of engineered recombinant interleukin 18. Adv Clin Exp Med. 2020;29(10):1135-43.

26. Tsuboi K, Miyazaki T, Nakajima M, Fukai Y, Masuda N, Manda R, et al. Serum interleukin-12 and interleukin-18 levels as a tumor marker in patients with esophageal carcinoma. Cancer Lett. 2004;205(2):207-14.

27. Xiao W, Tang P, Sui Z, Wu X, Yang Y, Zhu N, et al. Identification of a Hypoxia-Related Gene Signature and Establishment of a Nomogram for Predicting Prognosis in Esophageal Squamous Cell Carcinoma. https://www.researchsquare.com/article/rs-1008316/v1.

Tables

Table 1 Associations of BGN and IL-18 expression with clinicopathological variables of ESCC patients from the TMA dataset.

Clinicopathological variables	n	BGN expression		x^2	*P* value	IL18 expression		x^2	*P* value
		Low level	High level			Low level	High level		
Gender				1.924	0.165			3.513	0.061
Male	192	70 (36.5%)	122 (63.5%)			108 (56.3%)	84 (43.8%)		
Female	40	10 (25%)	30 (75%)			16 (40.0%)	24 (60.0%)		
Age（years）				0.837	0.360			0.595	0.440
≤66	114	36 (31.6%)	78 (68.4%)			58 (50.9%)	56 (49.1%)		
>66	118	44 (37.3%)	74 (62.7%)			66 (55.9%)	52 (44.1%)		
Tumor size				4.210	**0.040**			0.038	0.845
≤3.5	109	45 (41.3%)	64 (58.7%)			59 (54.1%)	50 (45.9%)		
>3.5	123	35 (28.5%)	88 (71.5%)			65 (52.8%)	58 (47.2%)		
Tumor Location				1.088	0.581			1.710	0.425
Upper	11	5 (45.5%)	6 (54.5%)			4 (36.4%)	7 (63.6%)		
Middle	171	56 (32.7%)	115 (67.3%)			91 (53.2%)	80 (46.8%)		
Lower	50	19 (38.0%)	31 (62.0%)			29 (58.0%)	21 (42.0%)		
Histological grade				3.312	0.191			13.439	0.179
I	6	4 (66.7%)	2 (33.3%)			1 (16.7%)	5 (83.3%)		
II	169	59 (34.9%)	110 (65.1%)			91 (53.8%)	78 (46.2%)		
III	57	17 (29.8%)	40 (70.2%)			32 (56.1%)	25 (43.9%)		
Tumor invasion depth				22.483	**0.000**			0.353	0.553
T1 - T2	56	34 (60.7%)	22 (39.3%)			28 (50.0%)	28 (50.0%)		
T3 - T4	176	46 (26.1%)	130 (73.9%)			96 (54.5%)	80 (45.5%)		
Lymph node metastasis				4.768	**0.029**			14.552	**0.000**
N0	128	52 (40.6%)	76 (59.4%)			54 (42.2%)	74 (57.8%)		
N1 - N3	104	28 (26.9%)	76 (73.1%)			70 (67.3%)	34 (32.7%)		

Abbreviations: BGN, biglycan; IL-18, interleukin 18; ESCC, esophageal squamous cell carcinoma; TMA, TMA tissue microarray.

Figures

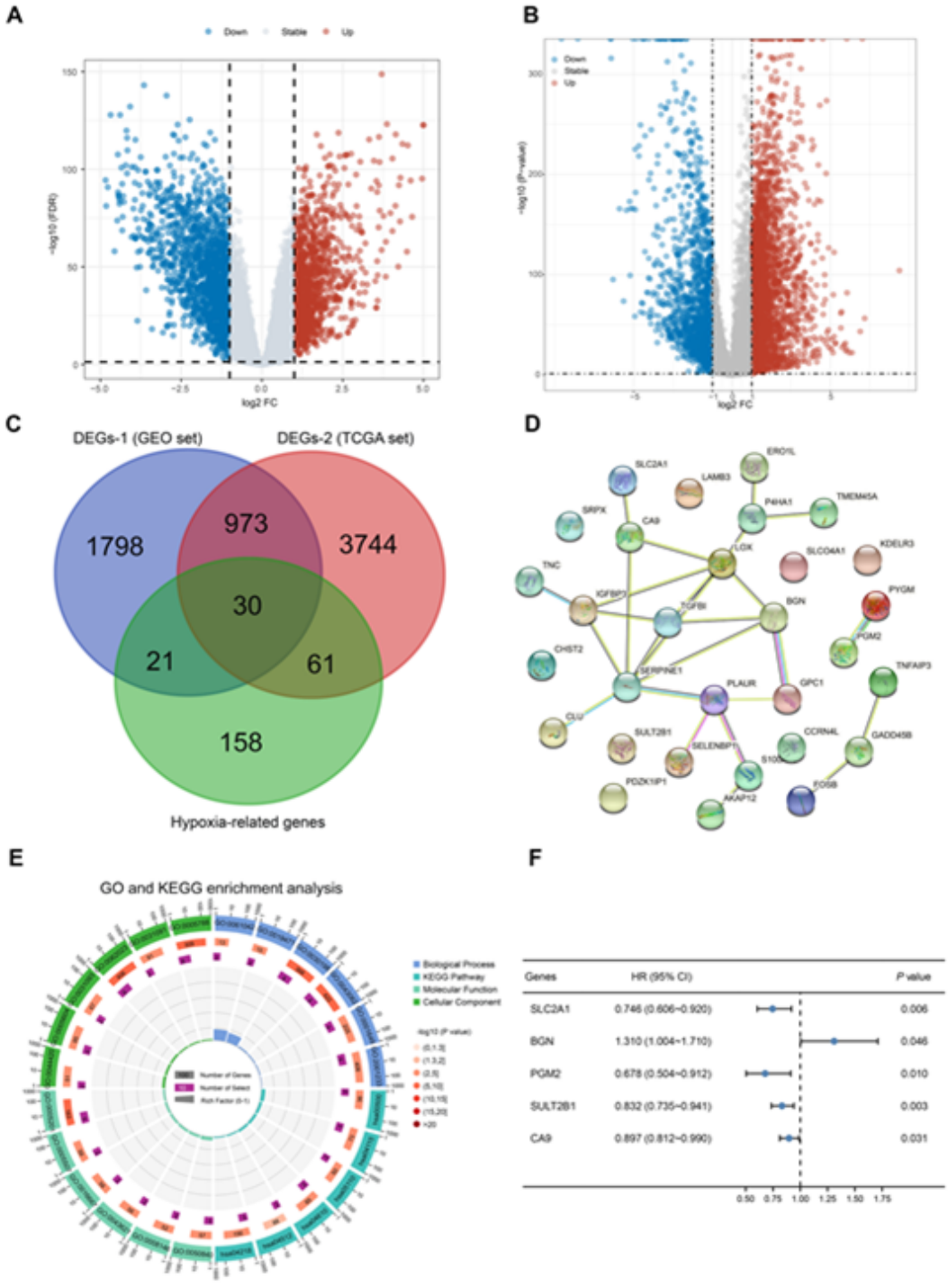

Figure 1

Identification of potential HRGs and related functional analysis. Differential expression analysis in ESCC (A-B). Volcano plots showing DEGs in (A) GES53625 and (B) TCGA-ESCC. (C) Venn diagrams of the 30 overlapping differentially expressed HRGs from three datasets. (D) PPI network of the differentially expressed HRGs constructed in STRING and visualized by Cytoscape. (E) The top 10 significant terms from the GO and KEGG analyses of differentially expressed HRGs.

HRGs, hypoxia related genes; ESCC, esophageal squamous cell carcinoma; DEGs, differentially expressed genes; PPI, protein–protein interaction; GO, Gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

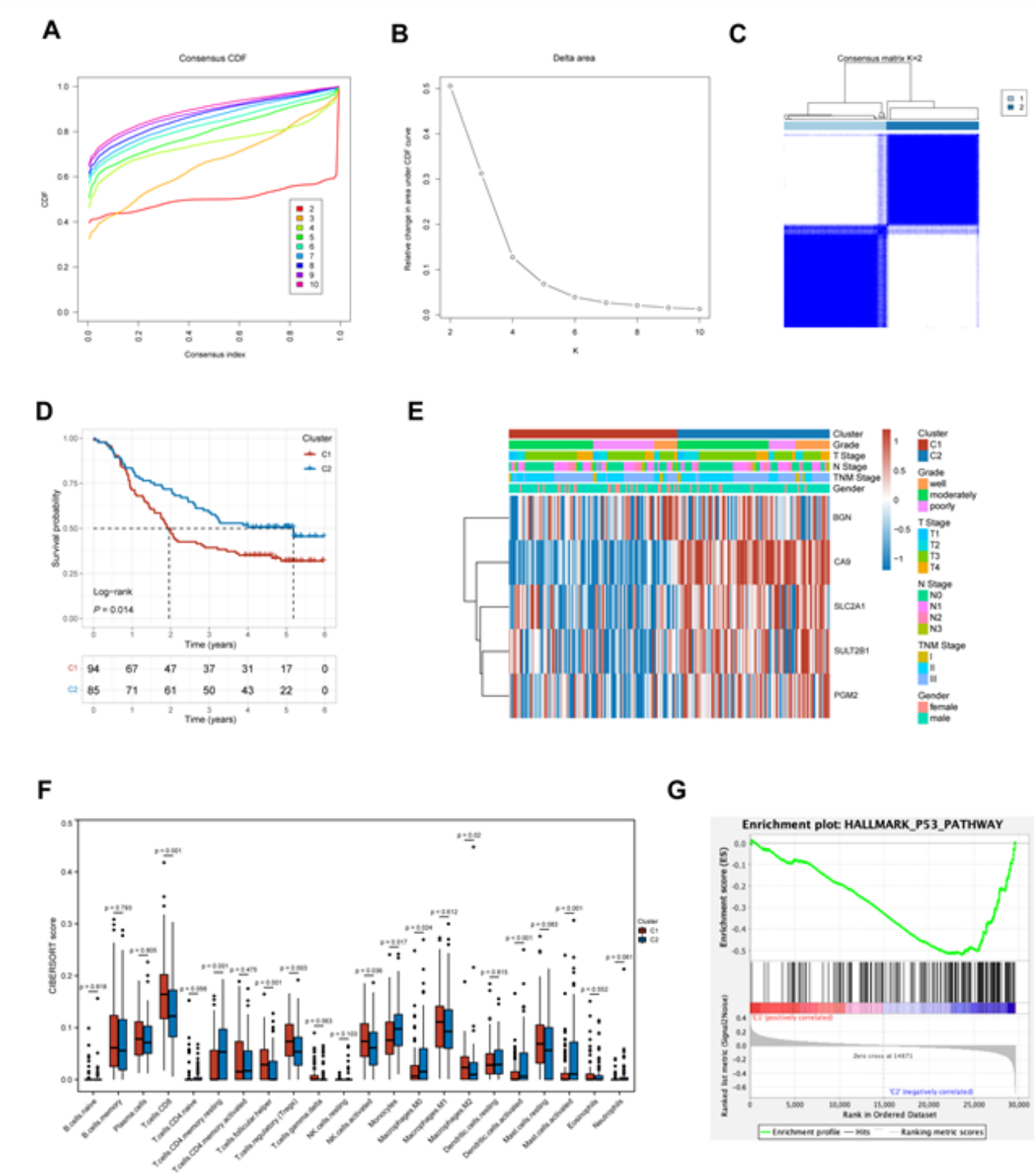

Figure 2

Analysis of subgroups with clinicopathological characteristics and immune infiltration. (A) Cumulative distribution map of clustering consistency. (B) Clustering Delta area map. (C) Display of the clustering results corresponding to K = 2. (D) The results of Kaplan-Meier survival analysis of different clusters. (E) A heatmap corresponding to the dendrogram annotated by grade, T stage, N stage, stage, sex, and 5 real hub

genes. (F) The infiltration of 22 immune cell subtypes in different clusters. (G) GSEA was used to demonstrate the correlation between genes from different clusters and the KEGG enriched pathways.

GSEA, gene set enrichment analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Figure 3

Screening of key HRGs. (A) The relationship of module features with consensus subgroups was assessed by eight gene modules obtained from WGCNA. (B) Volcano plots showing the DEGs in two clusters. (C) Venn diagrams of the 259 differentially expressed HRGs from intersection of green module genes and consensus clustering DEGs.

HRGs, hypoxia related genes; DEGs, differentially expressed genes; WGCNA, the weighted gene co-expression network analysis.

Figure 4

Constructing the prognostic hypoxia gene features in the training set. The samples were assigned a risk score and ordered to determine whether the expression level (A) and survival time (B) varied systematically with the risk score. (C) Expression levels of the 6 HRGs based on the risk score. (D) Survival curve distribution of the risk score. (E) ROC curves and AUCs of risk score classifications.

HRGs, hypoxia related genes; ROC, receiver operating characteristic; AUC, area under the curve.

Figure 5

Validating the prognostic hypoxia gene features in the external TCGA dataset. The samples were assigned a risk score and ordered to determine whether the expression level (A) and survival time (B) varied systematically with the risk score. (C) Expression levels of the 6 HRGs based on risk scores. (D) Survival curve distribution of the risk score. (E) ROC curves and AUCs of risk score classifications.

HRGs, hypoxia related genes.

Figure 6

Evaluating the independent role of the prognostic signature and building a predictive nomogram. Univariate (A) and multivariate (B) Cox regression analyses were used to evaluate the predictive value of

the model-incorporated prognostic signature. (C) Integrating the risk score with age and stage to construct a nomogram model. (D) The calibration curve of the nomogram.

Figure 7

Analysis of the correlation between the risk score and immune infiltration. (A) Expression and clinical features of the 5 prognostic genes in the high- and low-risk groups of the training set. (B) The results of the correlation analysis between the immune-related score and the risk score. (C) The infiltration of 22 immune cell subtypes in the high- and low-risk groups. (D-F) GSEA was used to demonstrate the correlation between HRG expression and the KEGG enriched pathways.

GSEA, gene set enrichment analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Figure 8

The expression of HRGs protein in ESCC and normal tissues was detected in TMA. Representative image of IHC staining for low expression of BGN in adjacent non-tumor tissue (A) and low expression (B), high expression case 1 (C), and high expression case 2 (D) of BGN in human primary tumor tissue, stained 5x, inset 40x magnification. Representative image of IHC staining for high expression of IL-18 in adjacent non-tumor tissue (E) and low expression case 1 (F), low expression case 2 (G) and high expression (H) of IL-18 in human primary tumor tissue, stained 5x, inset 40x magnification; (I) Kaplan-Meier curve of BGN expression in the TMA dataset. (J) Kaplan-Meier curve of IL-18 expression in the TMA dataset.

HRGs, hypoxia related genes; ESCC, esophageal squamous cell carcinoma; TMA, tissue microarray; BGN, biglycan; IL-18, interleukin 18.

Supplementary Files

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

- S.Fig.1.tif
- S.Fig.2.tif
- S.Fig.3.tif
- S.Fig.4.tif
- S.Fig.5.tif
- S.Fig.6.tif
- S.Table.1.pdf
- S.Table.2.pdf