Expression of pyroptosis-associated genes and construction of prognostic model for thyroid cancer

Lin Guo[1][#], Min Yuan[1][#], Senhe Jiang[2], Gang Jin[1]^, Ping Li[1]

[1]Department of Nuclear Medicine, The Second Affiliated Hospital of Harbin Medical University, Harbin, China; [2]Department of Otolaryngology, The First Affiliated Hospital of Harbin Medical University, Harbin, China

Contributions: (I) Conception and design: L Guo; (II) Administrative support: G Jin, P Li; (III) Provision of study materials or patients: S Jiang, L Guo; (IV) Collection and assembly of data: M Yuan, P Li; (V) Data analysis and interpretation: M Yuan, G Jin; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work as co-first authors.

Correspondence to: Gang Jin, Master's degree; Ping Li, PhD. Department of Nuclear Medicine, The Second Affiliated Hospital of Harbin Medical University, 246 Xuefu Road, Nangang District, Harbin 150000, China. Email: jin@hrbmu.edu.cn; h03713@hrbmu.edu.cn.

Background: Thyroid cancer (THCA) is a common endocrine malignancy. This study aimed to explore the expression of pyroptosis-related genes in THCA and establish a prognosis prediction model.

Methods: Differentially expressed pyroptosis-related genes (DEPRGs) were identified in The Cancer Genome Atlas (TCGA) and the Molecular Signatures Database (MSigDB). Subsequently, these genes were subjected to Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses. A least absolute shrinkage and selection operator (LASSO) regression model was employed to establish a DEPRG signature, and its reliability was validated through survival analysis and receiver operating characteristic (ROC) curves. Patients in TCGA-THCA cohort were stratified into two risk groups. The biological functions of the genes between the two risk groups were assessed through gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA). Finally, expression of DEPRGs was validated using datasets from the Gene Expression Omnibus (GEO) and the Human Protein Atlas (HPA) databases.

Results: Six DEPRGs were identified in TCGA dataset. Through LASSO Cox regression analysis, we determined *IL6*, *TP63*, *NOD1*, and *BAX* to be significant. Kaplan-Meier survival analysis demonstrated that patients with THCA expressing high levels of NOD1 and classified as low-risk individuals exhibited prolonged survival. The multifactorial ROC curves yielded area under the curve (AUC) scores exceeding 0.7 for risk score, age, and T-stage, affirming their significance as independent prognostic factors as determined by multivariate analysis. Additionally, we observed elevated expression levels of *BAX* and *NOD1* in THCA using data derived from the HPA database and the GEO dataset.

Conclusions: We established a novel DEPRG signature for predicting the prognosis of THCA, potentially offering a promising therapeutic marker for THCA treatment.

Keywords: Pyroptosis; thyroid cancer (THCA); gene signature; prognostic model

---

^ ORCID: 0000-0001-7028-4764.

Introduction

Thyroid cancer (THCA), the most common malignancy within the endocrine system (1), has witnessed a steady increase in its incidence in most parts of the world over the past few decades (2). As per the World Health Organization's Thyroid Tumor Classification Scheme (5th edition), malignancies arising from follicular cell-derived thyroid tumors can be categorized as follows: (I) follicular thyroid carcinoma; (II) infiltrative peritumoral follicular variant papillary carcinoma; (III) papillary thyroid carcinoma (PTC); (IV) eosinophilic cell carcinoma of the thyroid; (V) follicular carcinoma of follicular origin; and (VI) mesenchymal follicular cell-derived thyroid carcinoma. Notably, PTC stands out as the most prevalent malignancy derived from follicular cells, affecting both adult and pediatric populations (3). The substantial representation of PTC in our study stems from its status as the most common THCA type. Additionally, our dataset encompasses other THCA subtypes, such as follicular THCA, medullary THCA, and anaplastic THCA, allowing for a comprehensive analysis of this diverse group of malignancies. Many cancers are linked to lifestyle factors, including smoking and alcohol consumption (4).

While there have been notable advances in surgical resection, radioiodine therapy, drug therapy, targeted therapy, and immunotherapy (5,6), the optimal scope of surgery for differentiated thyroid cancer (DTC) remains a subject of debate. Radioactive iodine residual ablation serves as the standard adjuvant therapy for selected patients with DTC but necessitates a complete or near-complete thyroidectomy to prevent radionuclide absorption by residual thyroid tissue. Failure to achieve this can compromise its efficacy in eradicating micrometastatic disease. Recent study has demonstrated an increased risk of secondary malignancies following iodine therapy. Chemotherapy has been shown to exhibit limited efficacy due to the typically small size of these tumors (7). Progress has been made in the development of inhibitor-based therapies for THCA. However, two significant barriers hinder their clinical application in THCA syndromes: the challenge of managing adverse effects and the emergence of resistance (8). Therefore, there remains a pressing need to further refine treatment approaches to enhance disease response.

Pyroptosis is a form of programmed cell death accompanied by an inflammatory response (9,10). It has been associated with diseases such as arteriosclerosis and diabetic nephropathy. Study has indicated that pyroptosis plays pivotal roles in tumor proliferation, invasion, and metastasis. Pyroptosis induces cellular swelling, plasma membrane rupture, chromatin fragmentation, and the release of pro-inflammatory intracellular contents. Morphologically, pyroptosis differs from other cell death mechanisms, although they share some common features. In general, during pyroptosis, early-stage events involve DNA damage and chromatin condensation, followed by plasma membrane blistering and caspase activation, all without compromising cell membrane integrity (11). The relationship between pyroptosis and cancer is intricate, with a dual role in cancer progression and treatment. The impact of cell death varies across different cancer types. Studies have suggested that pyroptosis prevents colorectal tumor progression and inhibits tumor growth in hepatocellular carcinoma (12,13). Recent studies have uncovered novel pyroptosis-related characteristics in specific cancers. For example, pyroptosis-associated features have been used to predict prognosis and response to immunotherapy in patients with gastric cancer (14). Pyroptosis-related genes also predict the prognosis of ovarian cancer and play crucial roles in tumor immunity (15). Prognostic features of lung adenocarcinoma have been established based on regulators of cell death (16). Nonetheless, the prognostic value of apoptosis-related genes in THCA has not been extensively reported.

In this study, we identified differential expression patterns within both The Cancer Genome Atlas (TCGA) and the Molecular Signatures Database (MSigDB). Subsequently, we conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of pyroptosis-related genes. Furthermore, to establish a predictive model based on these differentially expressed pyroptosis-related genes (DEPRGs), we employed a cable regression model. The reliability of this model was assessed through survival analysis and receiver operating characteristic (ROC) curve analysis. Patients were stratified into two risk groups based on their risk scores. To further elucidate the biological functions of the genes between the risk groups, we employed gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA). Finally, we conducted experimental validation of the expression patterns of the DEPRGs.

In summary, our comprehensive analysis of the datasets from TCGA and the MSigDB databases led to the development of a predictive nomogram model. This model has the potential to forecast the prognosis of THCA and

　　　　　　　　　　　Transl Cancer Res 2023;12(12):3360-3383 | https://dx.doi.org/10.21037/tcr-23-810

3362

Guo et al. Pyroptosis-related model in THCA

may offer valuable insights into potential therapeutic targets for THCA treatment. We present this article in accordance with the TRIPOD reporting checklist (available at https://tcr.amegroups.com/article/view/10.21037/tcr-23-810/rc).

Methods

Data acquisition and data pre-processing

TCGA (17) is a pivotal resource for cancer researchers, encompassing a comprehensive array of data types, including clinical datasets, genomic variations, mRNA and miRNA expressions, methylation profiles, and more, across a spectrum of human cancer types and subtypes. For our study, we acquired samples from the TCGA database, specifically focusing on mRNA expression data pertaining to THCA. These mRNA expression values were provided in the fragments per kilobase of the exon model per million mapped fragments format. Concurrently, we obtained corresponding clinical information and survival data for the same Homo sapiens samples. The data platform utilized in TCGA was Illumina. Furthermore, we accessed THCA samples and normal controls from the Gene Expression Omnibus (GEO) database. Subsequently, we performed data normalization by applying the R package Limma (version 3.56.2) (18), transforming the data to log2 scale. This normalized expression distribution was then employed for generating boxplot visualizations. This study complies with the Declaration of Helsinki (as revised 2013).

TCGA dataset included 512 patients with THCA and 59 normal controls, while the GSE3467 dataset contained 9 patients with THCA and 9 normal controls. The GSE3678 dataset comprised seven patients with THCA and nine normal controls. The Hugo Gene Nomenclature Committee (HGNC) (19) is responsible for providing unique, standardized symbols for the human genome, including protein-coding genes. Each human gene in the HGNC database is assigned a unique numeric ID and symbol. mRNA expression profiles were extracted from the HGNC mRNA gene annotation file. We identified 27 pyroptosis-related genes from the MSigDB database and an additional 33 pyroptosis-related genes from literature reviews (20-23). After removing duplicate genes, we identified a total of 52 pyroptosis-related genes (available online: https://cdn.amegroups.cn/static/public/tcr-23-810-1.xlsx) for further analysis.

Analysis of differential expression of pyroptosis-related genes

The expression matrix for THCA was obtained from the TCGA-THCA dataset, and differentially expressed genes (DEGs) between the tumor and normal groups were analyzed using the Limma package. DEGs were defined based on the criteria of a P value <0.05 and an absolute fold change (|log2FC|) >1. Visualization of DEGs was achieved using the GGPLOT2 package (version 3.4.3) (24) to create a volcano plot highlighting the differential expression of genes. Additionally, the pheatmap package (25) was employed to visualize differences in DEG expression between patients with THCA and normal controls. Subsequently, DEGs associated with coke death were identified for further analysis, as indicated in the Venn diagram (version 1.7.3) (26).

GO and KEGG enrichment analyses of pyroptosis-related genes

GO enrichment analysis is a commonly employed method for conducting large-scale functional enrichment investigations encompassing biological processes (BPs), molecular functions (MFs), and cellular components (CCs) (27). KEGG (28) serves as a widely utilized database housing extensive information regarding genomes, biological pathways, diseases, and drugs, among other data. In our study, we performed GO annotation analysis on DEGs using the R package clusterProfiler (29), with significance determined at a P value <0.05. The Benjamini-Hochberg method was employed for P value correction.

Prognostic model construction

We developed a least absolute shrinkage and selection operator (LASSO) prognostic risk assessment model using R (version 4.1-8) (30) for the differentially expressed cell death genes in TCGA-THCA. Using this model, each TCGA-THCA sample was assigned a prognostic risk score. The cutoff for risk stratification was set at the median prognostic risk score, dividing patients into high- and low-risk groups based on this cutoff value. To visualize the association between risk scores and the prognostic genes of coke death, we generated heatmaps using the corrplot package (version 0.92) (31).

Prognostic and differential expression analysis of pyroptosis-related genes

The R survival package (version 3.5-7) (32) and the survival package (33) were employed to perform survival analysis and visualize pyroptosis-related genes as part of the model construction process. A box plot was generated using the R package GGPLOT2 to illustrate the differential gene expression between high- and low-risk groups.

GSEA and GSVA of high- and low-risk groups

We employed the clusterProfiler package (version 4.8.2) to conduct GSEA against the KEGG database (34), focusing on the TCGA-THCA expression profiles. For this analysis, we selected the KEGG gene set collection, specifically "C2.CP.KEGG.V7.4.symbols.gmt", as our reference gene set to discern significant differences between high- and low-risk populations. Significance was determined at a P value threshold of <0.05. To investigate disparities in BPs among distinct clustering groups, we computed pathway scores for each sample individually, based on their gene expression profiles, using the ssGSEA method implemented in the R-package GSVA (version 1.48.3) (35). Subsequently, we visualized the enrichment scores for each pathway in every sample using the PHEATMAP package (version 1.0.12). Pathway enrichment analysis was conducted with the R package Limma, and pathways with P values <0.05 were deemed significantly different. To illustrate the variations in pathways between high- and low-risk groups, we generated violin plots using the R package GGPLOT2.

Correlation analysis between prognosis gene and prognosis of pyroptosis

Based on the aforementioned models, each sample from the TCGA-THCA dataset underwent a prognostic risk assessment. The cutoff point was determined as the median value of the patients' prognostic risk scores, subsequently categorizing patients into high- and low-risk groups based on this established cutoff value. Survival analysis and visualization of these high- and low-risk groups were conducted using the R survival package and survminer package (version 0.4.9). To assess the model's predictive performance, time-dependent ROC analyses were carried out using the R package timeROC (version 0.4) (36).

Consistent clustering of pyroptosis-related genes

We employed the ConsensusClusterPlus package (version 1.64.0) in R (37) to consistently cluster the gene expression profiles of TCGA-THCA based on 52 coke-related genes, yielding three distinct subtypes denoted as C1, C2, and C3. Following the concordance clustering, we conducted a survival analysis of the TCGA-THCA cohort stratified by these subtypes, namely C1, C2, and C3. The relationship between survival outcomes and high- and low-risk groups was then visually represented using the ggalluvial package in R (version 0.12.5).

Nomogram model construction and correlation analysis

Clinical features of patients in TCGA-THCA were extracted using R, resulting in the identification of five features: age, gender, T, M, and N stages. Subsequently, univariate and multivariate Cox proportional hazards regression analyses were performed to assess patient prognostic risk scores in TCGA. These analyses aimed to explore factors associated with patient outcomes, and the results were visualized using the R package forestplot (version 3.1.1) (38). A nomogram is a Cartesian coordinate system that represents the functional relationship between multiple independent variables through a cluster of disjoint lines. To construct a nomogram model based on factors significantly associated with TCGA-THCA prognosis, we utilized the nomogram function within the R-package RMS (39). The constructed nomogram was then visualized using ggplot. To assess the predictive performance of the momogram model, time-dependent ROC analysis was conducted using the R package timeROC. This analysis involved evaluating the predictive scores of the nomogram model and comparing them to the overall survival (OS) status and survival time of the patients based on multiple factors.

Furthermore, calibration curves were employed to evaluate the accuracy and resolution of the nomograms. These curves were used to assess the prediction effect of the model on actual outcomes by plotting the fitting of the actual probability against the predicted probability under various conditions. The calibration curves primarily served to evaluate the model established using the Cox regression method in comparison to real-world observations. The construction of nomograms and calibration curves was

3364

Guo et al. Pyroptosis-related model in THCA

executed using the R package RMS. Additionally, decision curve analysis was employed as a straightforward method to evaluate clinical predictive models, diagnostic tests, and molecular markers. The R package GGDCA (version 1.1) (40) was employed to assess the nomogram model's performance concerning 1-, 3-, and 5-year patient survival.

Differential expression of pyroptosis-related genes in THCA and immunohistochemistry (IHC) analysis

The GEO datasets GSE3467 and GSE3678 comprise expression matrices pertaining to THCA. These matrices were subjected to normalization using the Limma package. A box plot was generated using the R package GGPLOT2 to illustrate the differential expression of prognosis-related genes in both normal control and tumor tissues. Immunohistochemical staining of cell death-related prognostic genes and mRNA expression of death-related genes across various carcinomas, were sourced from the Human Protein Atlas (HPA) (https://www.proteinatlas.org/) (41).

Statistical analysis

All dataset processing and statistical analyses were performed using R software (version 3.65). The ROC curve of gene expression and its impact on patient survival time, as well as survival status, was plotted utilizing the R timeROC package. Additionally, the area under the curve (AUC) was calculated to evaluate the diagnostic efficacy of gene expression in relation to patient survival. Correlation analysis between the risk score and prognostic genes was performed with Plzeň's correlation coefficient, implemented via the corrplot function within the CORRPLOT package in R.

Results

The study design is illustrated in *Figure 1*.

Data acquisition and data pre-processing

Gene expression matrices and clinical datasets for THCA were initially acquired from TCGA, while the gene expression datasets for GSE3467 and GSE3678 were sourced from GEO. The protein gene annotation file was retrieved from HGNC. Subsequently, 52 genes associated with THCA were extracted from both the MSigDB and relevant literature for subsequent analysis.

Analysis of differential expression of pyroptosis-related genes

After preprocessing the TCGA dataset, we conducted a differential expression analysis between disease and normal groups. We utilized the Limma package to analyze the pyroptosis-related gene expression matrix within the TCGA dataset. We applied the criteria of $|log2FC| >1$ and a P value <0.05 as thresholds for screening. The results are depicted in *Figure 2A,2B*. A total of 759 genes exhibited upregulation, while 1,108 genes showed downregulation. Among these, six genes were identified as differentially expressed and are related to cell pyroptosis (*Figure 2C*). These genes include *IL6*, *TP63*, *GSDMA*, *GSDMB*, *NOD1*, *BAX*, and *NLRP6*. Details of the alterations in these six genes can be found in *Table 1*.

Identification of prognostic genes

To comprehensively investigate the attributes of 52 pyroptosis-related genes within THCA, encompassing BPs, MFs, CCs, biological pathways, and gene expression patterns, we conducted enrichment analyses focusing on co-occurring genes associated with pyroptosis. Our findings from the GO functional enrichment analysis are visually depicted in a bubble plot (*Figure 3A*). The interconnectedness between cell pyroptosis-related genes and the outcomes of GO enrichment analysis is elucidated through a ring network plot (*Figure 3B*). Furthermore, the results of our KEGG enrichment analysis are presented with item annotation plots (*Figure 3C*). Additionally, the interrelation between cell pyroptosis-related genes and the findings of KEGG enrichment analysis is delineated via a network plot (*Figure 3D*). We have also established correlations between the expression profiles of cell pyroptosis-related genes and the results of both GO and KEGG enrichment analyses, visually depicted as a circular plot (*Figure 3E*) and a chord plot (*Figure 3F*), respectively. Within the realm of BPs, we observed enrichments in pathways such as midbody abscission, positive regulation of interleukin-1 production, positive regulation of cysteine-type endopeptidase activity, viral budding via the host Endosomal Sorting Complexes Required for Transport (ESCRT) complex, and mitotic cytokinetic processes. In CC, our analyses highlighted enrichments in the inflammasome complex and late endosome membrane. Moreover, MFs exhibited enrichments in cysteine-type endopeptidase activity involved in apoptotic processes, cysteine-type endopeptidase
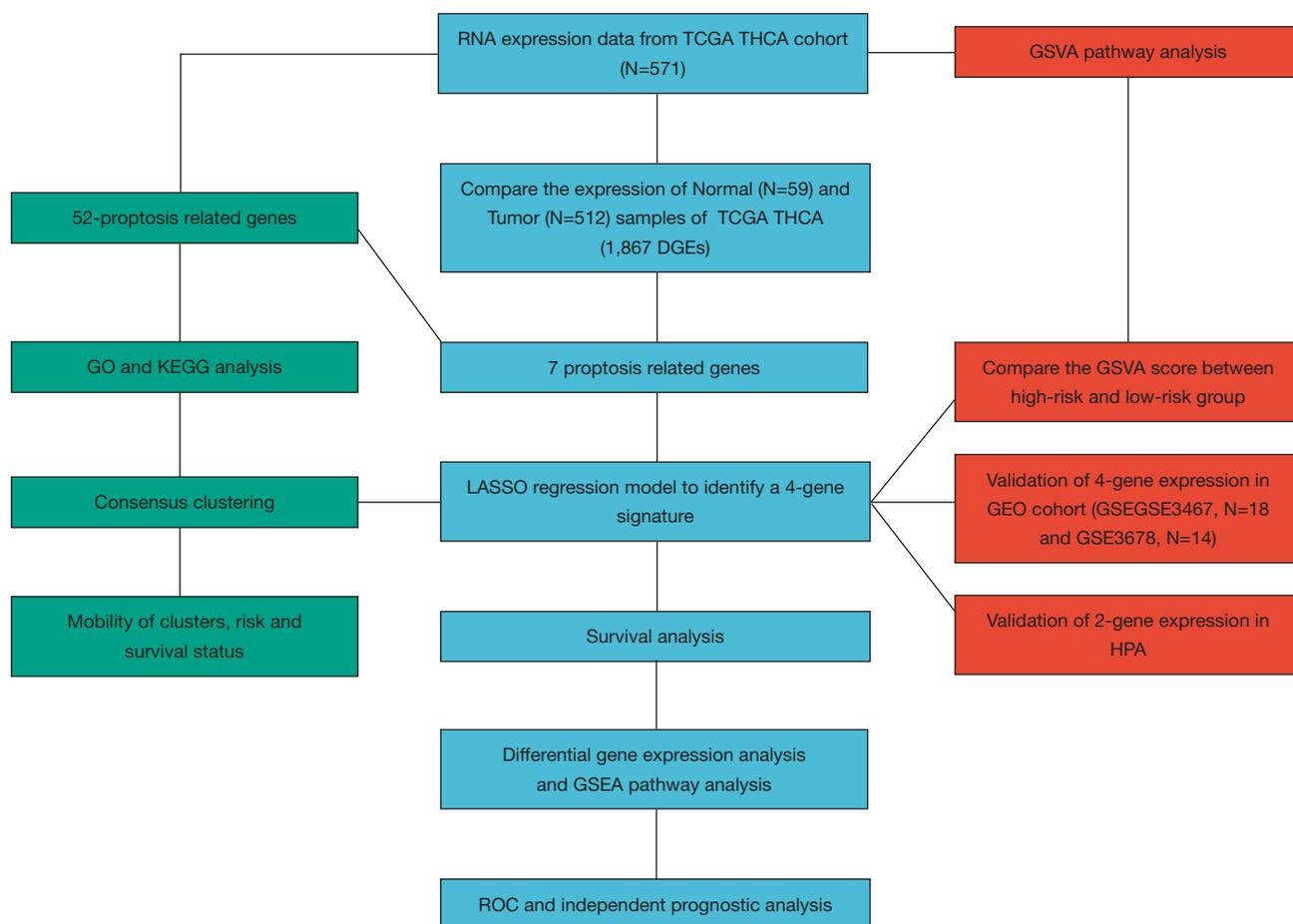
Figure 1 Study design. TCGA, The Cancer Genome Atlas; THCA, thyroid cancer; GSVA, gene set variation analysis; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; DGEs, differentially expressed genes; LASSO, least absolute shrinkage and selection operator; GEO, Gene Expression Omnibus; HPA, Human Protein Atlas; GSEA, gene set enrichment analysis; ROC, receiver operating characteristic.

activity involved in apoptotic signaling pathways, cysteine-type endopeptidase activity, cysteine-type peptidase activity, and phosphatidylinositol-4,5-bisphosphate binding, among others. Additionally, we identified that apoptosis-related genes demonstrated enrichments in pathways such as necroptosis, legionellosis, lipid and atherosclerosis, Salmonella infection, pathogenic Escherichia coli infection, and influenza A, among others. Detailed results of the GO and KEGG enrichment analyses can be found in https://cdn. amegroups.cn/static/public/tcr-23-810-2.xlsx and https://cdn. amegroups.cn/static/public/tcr-23-810-3.xlsx, respectively.

Identification of prognosis-related genes

Utilizing LASSO regression techniques to individually screen the six pyroptosis-related genes and ascertain the optimal lambda values, we identified four pyroptosis-related genes (*Figure 4A,4B*): *IL6*, *TP63*, *NOD1*, and *BAX*. Subsequently, logistic regression was employed to formulate the prediction model and compute the regression coefficients for each gene. The formula for calculating the risk score was as follows:

$$\text{Risk score} = (0.0662 \times IL6exp) + (-0.0024 \times TP63exp) \\ + (-0.0893 \times NOD1exp) + (-0.0189 \times Baxexp) \qquad [1]$$

Utilizing the prognostic risk formula for TCGA-THCA samples, we derived individual risk scores for each sample. Subsequently, we computed the relationship between these risk scores and the genes associated with cell death, which
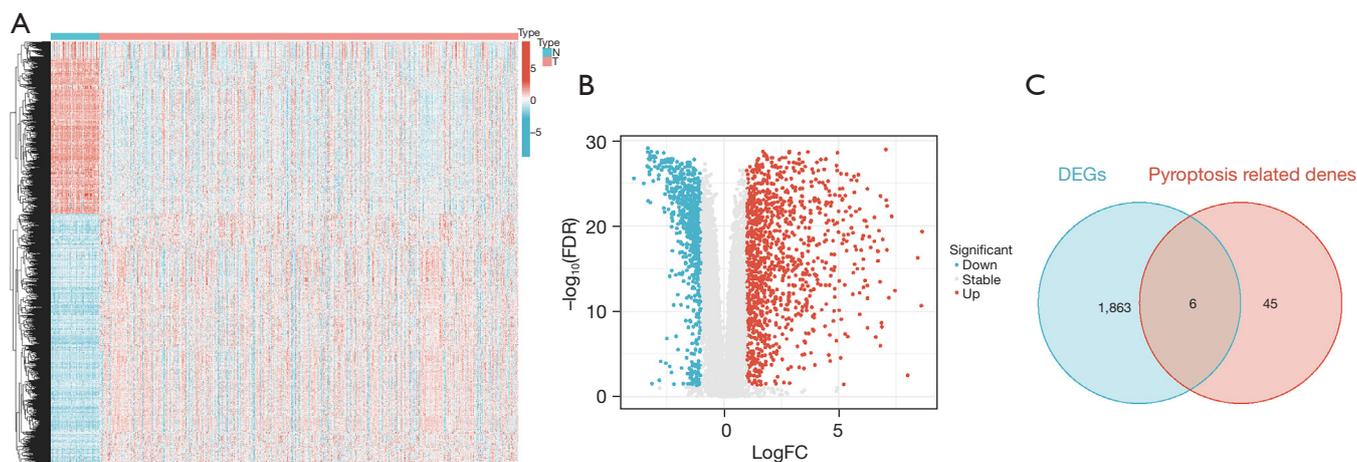
Transl Cancer Res 2023;12(12):3360-3383 | https://dx.doi.org/10.21037/tcr-23-810

3366

Guo et al. Pyroptosis-related model in THCA

Figure 2 Differential gene expression analysis of TCGA-THCA. (A) Heat map of differentially expressed gene correlation (darker the color, the stronger the correlation). Red indicates a negative correlation, and blue indicates a positive correlation. (B) Volcano plot of differentially expressed genes in the datasets. Upregulated genes are indicated in red, and downregulated genes are indicated in blue. Non-differentially expressed genes are indicated in gray. (C) Map of cell death-related genes and DEGs. N, normal group; T, tumor tissue; FC, fold change; DEGs, differentially expressed genes; FDR, false discovery rate; TCGA, The Cancer Genome Atlas; THCA, thyroid cancer.

Table 1 DEPRGs between normal tissues and tumor tissues in TCGA-THCA dataset

Gene	Normal_Mean	Tumor_Mean	logFC	P value	FDR
IL6	4.56772034	1.67823398	−1.4445304	0.00121779	0.00179814
TP63	0.13951356	0.74516309	2.41715086	7.64E−07	1.56E−06
GSDMA	0.14567797	0.79275547	2.44409324	2.72E−17	1.77E−16
GSDMB	1.47109322	3.14486738	1.09611051	5.84E−10	1.62E−09
NOD1	1.79304915	9.64505547	2.4273745	5.23E−19	4.33E−18
BAX	12.7574576	26.0353348	1.02913011	9.15E−27	4.38E−25

DEPRGs, differentially expressed pyroptosis-related genes; TCGA, The Cancer Genome Atlas; THCA, thyroid cancer; logFC, Log fold change; FDR, false discovery rate.

were integral to the model's construction (*Figure 4C*). Our findings revealed a positive correlation between risk scores and *IL6*, and conversely, a negative correlation with *TP63*, *NOD1*, and *BAX*. Heatmap of the differential expression of DEPRGs in normal and tumor tissues in the TCGA-THCA dataset (*Figure 4D*).

Differential expression analysis and prognostic analysis of cell death-related genes in high- and low-risk groups

The model was used to assess the prognostic risk for each sample within the TCGA-THCA dataset. The cutoff point was determined as the median value of patients' prognostic risk scores, thereby segregating patients into high-risk and

low-risk groups based on this threshold. Subsequently, we employed the R package Limma to compute the differential expression of genes between these high-risk and low-risk groups. The alterations in DEPRGs within these high- and low-risk groups are summarized in *Table 2*. To provide an overview of the four pyroptosis-related genes that contributed to the model's construction, we utilized a heatmap (*Figure 5A*). Following this, we utilized box plots to illustrate the differential expression of each gene within the high-risk and low-risk groups (*Figure 5B-5E*). Notably, our findings indicated higher expression levels of *Bax* and *NOD1* and lower expression levels of *IL6* and *TP63*. Specifically, *Bax*, *TP63*, and *NOD1* were significantly overexpressed within the low-risk population (P<0.05),

Figure 3 Functional enrichment analysis (GO) and pathway enrichment analysis (KEGG) of pyroptosis-related genes. (A,B) Results of GO functional enrichment analysis of pyroptosis-related genes: bubble plot (A) and ring network plot (B). (C,D) Results of KEGG pathway enrichment analysis of pyroptosis-related genes are presented in column plots (C) and network plots (D). (E) GO functional enrichment combined with logFC analysis of the pyroptosis-related genes is shown in a circle diagram. (F) KEGG pathway enrichment of pyroptosis-related genes combined with logFC analysis results are shown in a chord diagram. The vertical coordinates in the bubble plot (A) are GO terms. Color of the bubble indicates activation or inhibition of GO terms. Red indicates activation, and blue indicates inhibition. In the ring network diagram (B), the gradient-colored dots represent specific genes. Red represents upregulation, and blue represents downregulation. Yellow circles represent specific pathways. Red dots represent upregulated genes (logFC >0), and blue dots represent downregulated genes (logFC <0) in the loop graph (E). In the screening criteria for GO and KEGG enrichment analysis, a P value <0.05 and the FDR value (q.value) <0.05 were considered significantly enriched. ESCRT, Endosomal Sorting Complexes Required for Transport; CC, cellular component; BP, biological processes; MF, molecular functions; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; FC, fold change; FDR, false discovery rate.
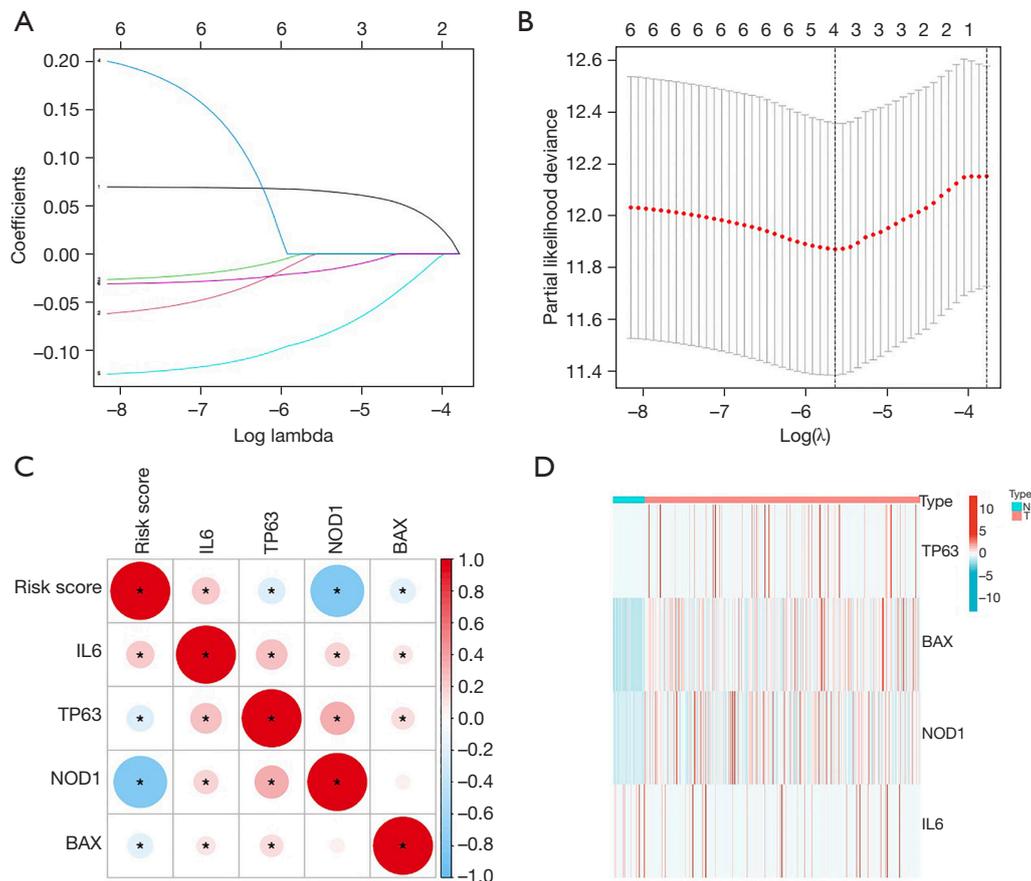
3368

Guo et al. Pyroptosis-related model in THCA

Figure 4 DEPRG prognostic model construction. (A) Variable trajectories of LASSO regression prognostic model for DEPRGs; (B) LASSO regression model construction based on DEPRGs; (C) heatmap of the association of risk scores with key prognostic genes. Blue indicates a negative association, and red indicates a positive association. The darker the color, the stronger the association. *, P<0.05. (D) Heatmap of differential expression of DEPRGs in normal and tumor tissues in the TCGA-THCA dataset, N, normal, normal tissue; T, tumor, tumor tissue; DEPRGs, differentially expressed pyroptosis-related genes; LASSO, least absolute shrinkage and selection operator; TCGA, The Cancer Genome Atlas; THCA, thyroid cancer.

Table 2 Variations in DEPRGs between the high- and low-risk groups

Gene	Low_Mean	High_Mean	logFC	P value	FDR
IL6	0.9834373	2.351649	1.257768	0.1269568	0.2530677
TP63	0.8444252	0.6268688	−0.429806	2.5659E−14	5.0318E−13
GSDMA	3.644904	2.620673	−0.4759432	4.7766E−10	5.3678E−09
GSDMB	1.063674	0.4848044	−1.133582	3.2441E−22	1.8535E−20
NOD1	14.92125	4.262019	−1.80776	3.768E−69	2.1184E−64
BAX	28.05693	23.97105	−0.2270641	2.7861E−06	1.782E−05

DEPRGs, differentially expressed pyroptosis-related genes; logFC, Log fold change; FDR, false discovery rate.
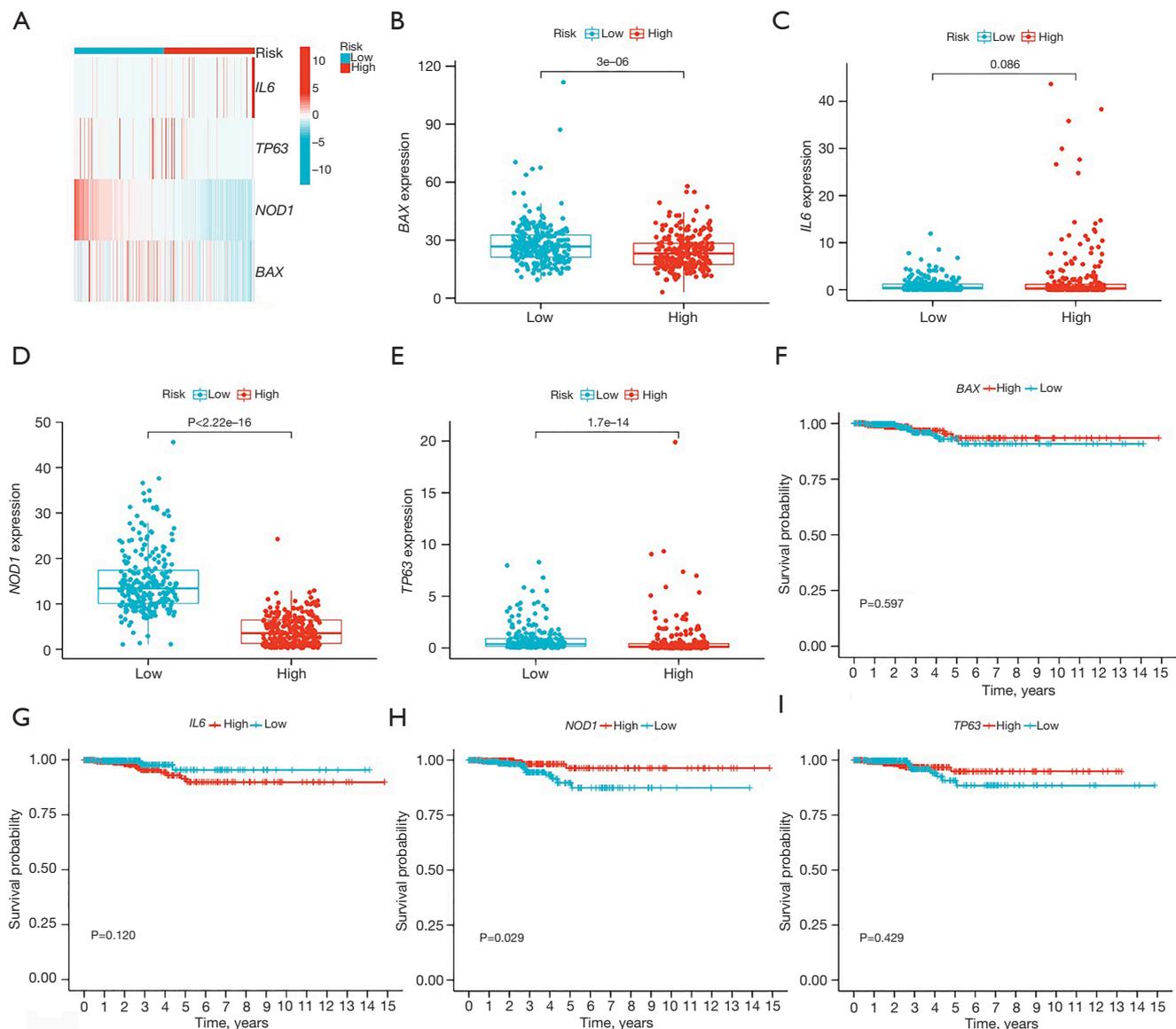
Figure 5 Differential expression analysis of key death-related prognostic genes in high-risk and low-risk groups and their prognostic relevance. (A) In the TCGA-THCA dataset, DEPRGs were differentially expressed in the high- and low-risk groups. (B-E) Differential expression of *BAX* (B), *IL6* (C), *NOD1* (D), and *TP63* (E) in the high- and low-risk groups. (F-I) KM curves were used to compare patient survival for each of the key prognostic genes: *BAX* (F), *IL6* (G), *NOD1* (H), and *TP63* (I) in the high- and low-risk groups. P≥0.05 indicates no statistical significance; P<0.05 indicates statistical significance; P<0.01 indicates high statistical significance; and P<0.001 indicates high statistical significance. TCGA, The Cancer Genome Atlas; THCA, thyroid cancer; DEPRGs, differentially expressed pyroptosis-related genes; KM curve, Kaplan-Meier curve.

while *IL6* exhibited a trend toward overexpression within the high-risk population (P=0.086).

Subsequently, we employed the R survival package in conjunction with the survminer package to analyze and visualize the survival outcomes associated with various cell death-related prognostic genes. Our results revealed that patients with high *NOD1* expression exhibited significantly improved survival times (P<0.05). Conversely, the expression levels of *IL6*, *TP63*, and *BAX* genes were not found to be associated with survival (*Figure 5F-5I*).

3370

Guo et al. Pyroptosis-related model in THCA

Table 3 GSEA of high- and low-risk groups in TCGA-THCA dataset

ID	Set size	Enrichment score	P value	Adj. P value	Q value
KEGG_OXIDATIVE_PHOSPHORYLATION	114	0.69343937	1.0702E-09	1.905E-07	1.4307E-07
KEGG_PARKINSONS_DISEASE	110	0.67640441	3.5824E-09	3.1883E-07	2.3946E-07
KEGG_CELL_ADHESION_MOLECULES_CAMS	105	−0.537413	1.0161E-06	6.0289E-05	4.528E-05
KEGG_HUNTINGTONS_DISEASE	157	0.55842937	4.6457E-06	0.00020673	0.00015526
KEGG_ALLOGRAFT_REJECTION	24	−0.7431423	3.7345E-05	0.00132948	0.00099849
KEGG_ASTHMA	16	−0.8040169	7.6238E-05	0.00226173	0.00169864
KEGG_ALZHEIMERS_DISEASE	135	0.55157927	9.4835E-05	0.00241151	0.00181113
KEGG_TYPE_I_DIABETES_MELLITUS	27	−0.6995838	0.00013701	0.00304849	0.00228952
KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_DEGRADATION	41	0.67863518	0.00039287	0.00777019	0.00583568
KEGG_GLYCINE_SERINE_AND_THREONINE_METABOLISM	22	0.79981019	0.00057476	0.01023077	0.00768366
KEGG_BUTANOATE_METABOLISM	26	0.75330099	0.00064699	0.01046952	0.00786298
KEGG_ARGININE_ANDPROLINE_METABOLIM	44	0.65383932	0.00139696	0.02072161	0.01556265
KEGG_LEISHMANIA_INFECTION	59	−0.5054466	0.00169835	0.02325434	0.01746482
KEGG_CARDIAC_MUSCLE_CONTRACTION	48	0.62449405	0.00187037	0.02378047	0.01785996
KEGG_VIRAL_MYOCARDITIS	55	−0.5105996	0.00214481	0.0254518	0.01911519
KEGG_PROPANOATE_METABOLISM	29	0.69840805	0.00360547	0.0401108	0.03012461
KEGG_GRAFT_VERSUS_HOST_DISEASE	24	−0.6233416	0.00407042	0.04261973	0.0320089

GSEA, gene set enrichment analysis; TCGA, The Cancer Genome Atlas; THCA, thyroid cancer.

GSEA of cell death-related genes in high- and low-risk groups

The TCGA-THCA expression profile was subjected to GSEA, distinguishing between high- and low-risk groups, utilizing the GSEA function in the clusterProfiler package. The reference gene set employed for this analysis was "C2. cp.kegg.v7.4.SYMBOLS.GMT". This analysis identified 17 differentially enriched pathways between the high- and low-risk groups (*Table 3*). For specific details regarding the genes enriched within each pathway, please see at https://cdn. amegroups.cn/static/public/tcr-23-810-4.xlsx. To elucidate the pathways associated with the low-risk group, we selected the top three enrichments (*Figure 6D-6F*). These pathways encompassed KEGG allograft rejection, KEGG asthma, and KEGG cell adhesion molecules (CAMs). Conversely, in the high-risk group, the top three enrichments (*Figure 6A-6C*) were KEGG Alzheimer's disease, KEGG Huntington's disease, and KEGG CAMs's disease, and KEGG oxidative phosphorylation.

GSVA of cell death-related genes in high- and low-risk groups

To investigate disparities in pathways between high- and low-risk groups, we performed a GSVA of the TCGA-THCA expression profile. Our analysis revealed 151 significantly distinct pathways in these high- and low-risk groups (adjusted P value <0.05) (available online: https:// cdn.amegroups.cn/static/public/tcr-23-810-5.xlsx and *Figure 7A*). Among these pathways, we selected the top four pathways that are enriched in the low- and high-risk groups, respectively (*Figure 7B* and *Table 4*). In the low-risk group, the following pathways exhibited enrichment: KEGG allograft rejection, KEGG CAMs, KEGG viral myocarditis, and KEGG asthma. Conversely, the high-risk group displayed enrichment in the following pathways: KEGG limonene and pinene degradation, KEGG propanoate and butanoate metabolism, KEGG valine, leucine, and isoleucine degradation, and KEGG butanoate metabolism.
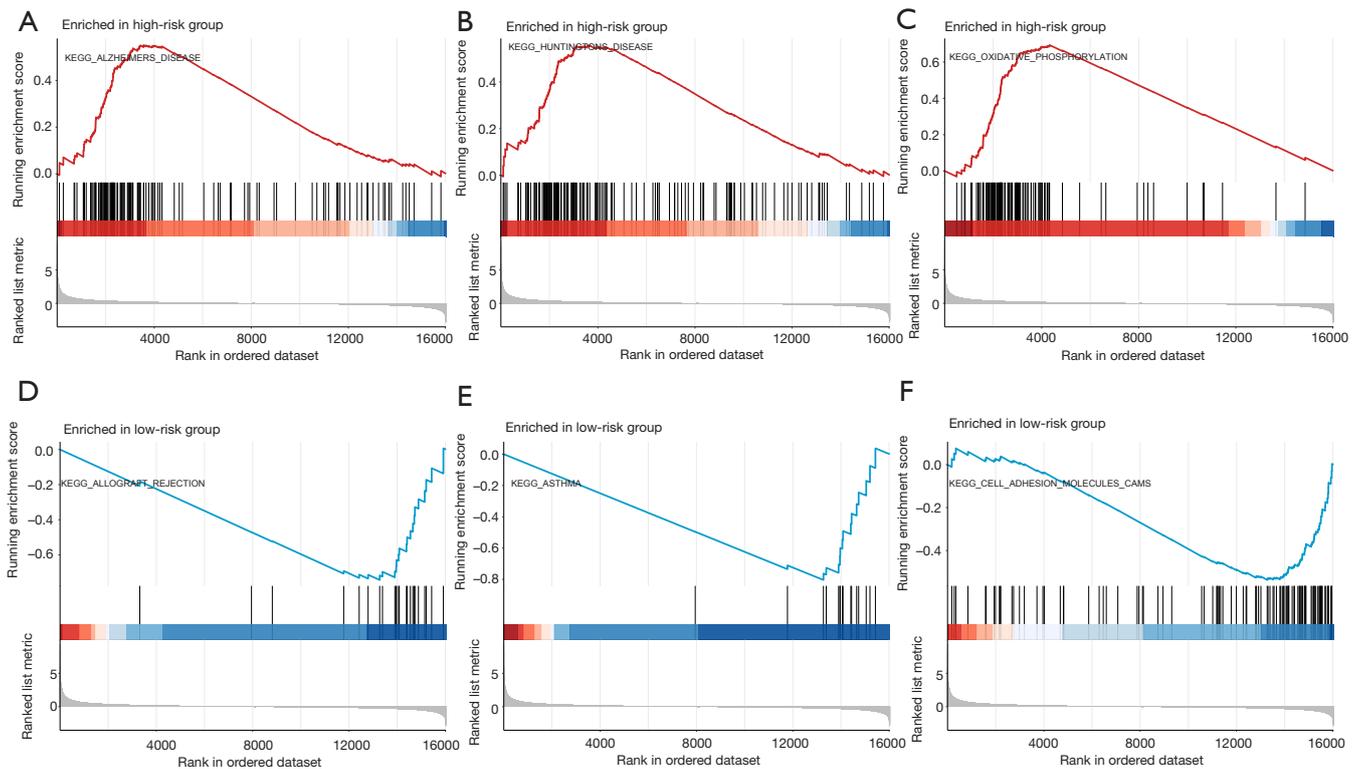
Figure 6 GSEA of cell death-related genes in high- and low-risk groups. (A-C) Different genes were associated with KEGG Alzheimer's disease (A), KEGG Huntington's disease (B), and KEGG oxidative phosphorylation (C) pathways in the high- and low-risk groups. (D-F) Differentially expressed genes were associated with KEGG allograft rejection (D), KEGG asthma (E), KEGG CAMs (F) pathways in the low-risk group. P<0.05 was considered to indicate statistical significance. GSEA, gene set enrichment analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; CAM, cell adhesion molecule.

Concordant cluster analysis and survival based on cell coke death-related genes

We employed the ConsensusClusterPlus package in R to consistently cluster the gene expression profiles of TCGA-THCA based on a set of 52 cell pyroptosis-related genes. The optimal clustering outcome was achieved at K=3, as illustrated in *Figure 8A,8B*. This yielded three distinct subtypes denoted as C1, C2, and C3, with the corresponding patient subgroups detailed in https://cdn.amegroups.cn/static/public/tcr-23-810-6.xlsx (see below). Following the concordant clustering, we conducted a survival analysis within the TCGA-THCA cohort, stratified by subtypes C1, C2, and C3. Notably, no significant difference was observed in OS among these three subtypes (P=0.206) (*Figure 8C*). Subsequently, a Sankey chart was generated to visualize the classification of patients into high- and low-risk groups, along with their respective survival status. Notably, a majority of patients within the C1 subtype were classified

as high-risk, while some fell into the low-risk category. For the C2 subtype, a portion of patients were identified as high-risk, while another subset was categorized as low-risk. The C3 subtype, although less common, exhibited a balanced distribution, with approximately half of its patients falling into the high-risk group and the other half into the low-risk group. Among the high-risk individuals, the majority were found to be alive, while only a few were deceased. Conversely, very few of the low-risk individuals were deceased. Importantly, patients with THCA displayed superior OS (*Figure 8D*).

Correlation analysis between prognosis and pyroptosis-related gene

The model was employed to assess the prognostic risk of individual samples within the TCGA-THCA dataset. The cutoff point was defined as the median value of patients' prognostic risk scores. Subsequently, patients were
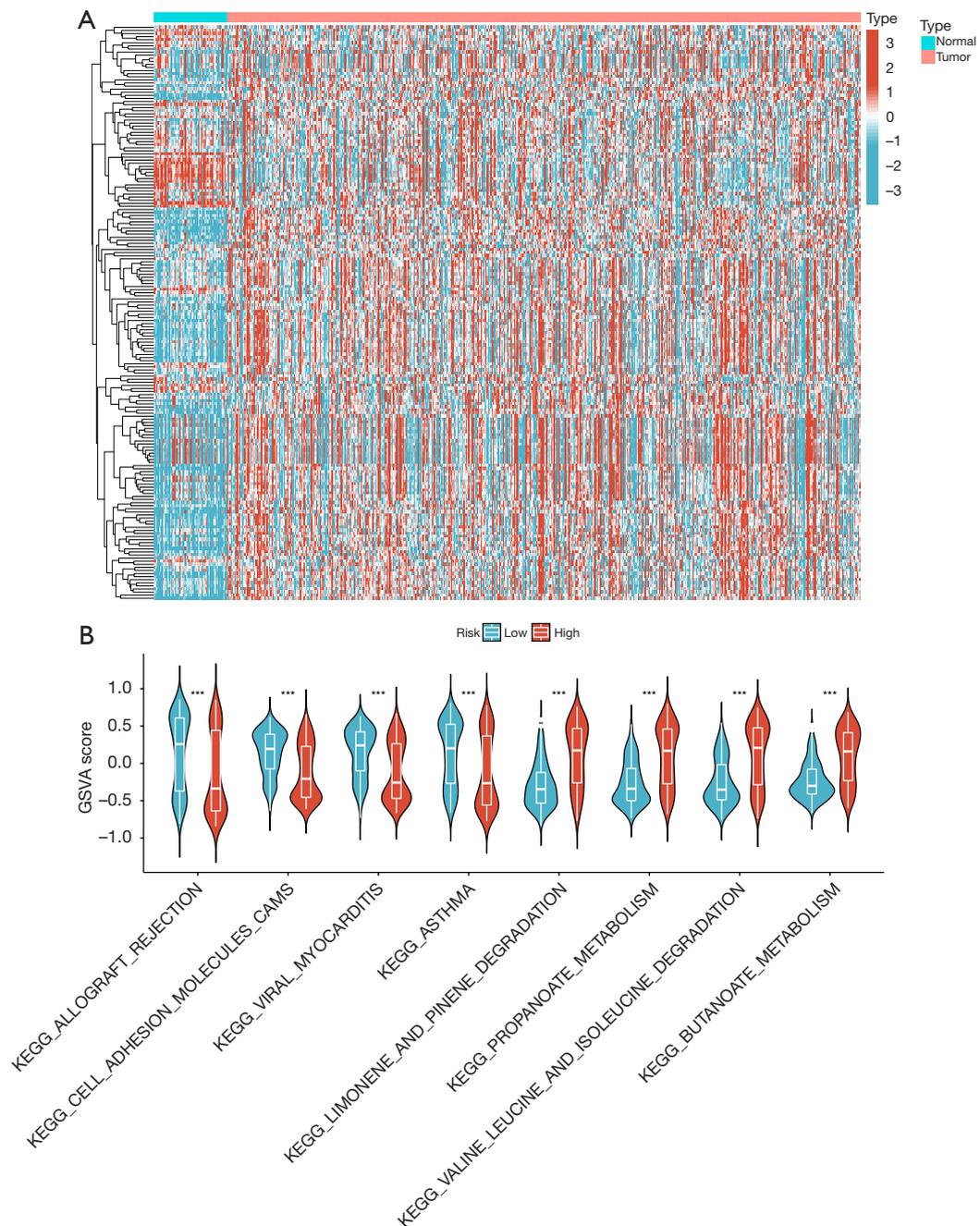
3372

Guo et al. Pyroptosis-related model in THCA

Figure 7 GSVA of high- and low-risk groups. (A) GSVA of TCGA-THCA dataset, and a heatmap of normal and tumor groups. (B) Top four pathways enriched in the high- and low-risk groups, respectively. Pathways enriched in the high-risk group were KEGG limonene and pinene degradation, KEGG propanoate butanoate metabolism, KEGG Valine leucine and isoleucine degradation, and KEGG butanoate metabolism. Pathways enriched in the low-risk group were: KEGG allograft rejection, KEGG cell adhesion molecules CAMs, KEGG viral myocarditis, and KEGG asthma. ***, P<0.001. GSVA, gene set variation analysis; TCGA, The Cancer Genome Atlas; THCA, thyroid cancer; KEGG, Kyoto Encyclopedia of Genes and Genomes; CAM, cell adhesion molecule.

Table 4 Top four pathways in GSVA enriched in high- and low-risk groups, respectively

ID	logFC	AveExpr	t	P value	Adj. P value	B
KEGG_LIMONENE_AND_PINENE_DEGRADATION	0.4027005	−0.0932206	12.5908106	7.56E−32	7.03E−30	61.5271241
KEGG_PROPANOATE_METABOLISM	0.36412521	−0.0844205	11.6198361	7.42E−28	1.97E−26	52.4684866
KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_ DEGRADATION	0.36151221	−0.0729168	11.0650592	1.17E−25	2.43E−24	47.4810468
KEGG_BUTANOATE_METABOLISM	0.3309971	−0.0689158	12.0576439	1.24E−29	5.68E−28	56.5034141
KEGG_ALLOGRAFT_REJECTION	−0.2685861	−0.0034524	−5.8158199	1.07E−08	2.79E−08	9.24563085
KEGG_CELL_ADHESION_MOLECULES_CAMS	−0.2595397	0.0120555	−8.8214921	1.79E−17	1.59E−16	28.9557849
KEGG_VIRAL_MYOCARDITIS	−0.2548962	0.01769772	−7.9998049	8.42E−15	4.75E−14	22.9282579
KEGG_ASTHMA	−0.2484736	0.01030598	−6.156239	1.51E−09	4.38E−09	11.1378064

GSVA, gene set variation analysis; logFC, Log fold change; AveExpr, average expression.

categorized into distinct risk groups using the R survival package and survminer package for survival analysis based on high and low-risk designations, and the results are illustrated in *Figure 9A*. Notably, the survival analysis yielded a log-rank P value of less than 0.05, indicating a more favorable prognosis in the low-risk group. To further evaluate the prognostic capabilities of the model within the TCGA-THCA dataset, time-dependent ROC analysis was conducted at 1-, 3-, and 5-year intervals based on the prognostic risk scores. All datasets exhibited AUC values exceeding 0.5, indicative of strong predictive performance. Specifically, both the 3- and 5-year datasets displayed AUC values exceeding 0.7, underscoring the model's heightened effectiveness in predicting survival at these time points (*Figure 9B*). Moreover, the AUC values associated with patients' age, T stage, and risk scores all exceeded 0.7, affirming their robust predictive utility. Baseline tables detailing demographic characteristics such as age, sex, T, M, and N stages, and risk scores are provided in *Table 5*. Furthermore, the clustering of risk factors within the LASSO regression prognostic model was visualized using a risk factor graph (*Figure 9C*). This graphical representation revealed a higher incidence of mortality within the population with elevated risk scores. AUC values for risk scores and clinical characteristics (*Figure 9D*). Subsequent validation of other clinical factors for survival outcomes underscored the superior predictive power of risk scores, age, and T stage for patients with THCA.

Validation of the prognostic independence of the model and the prognostic value of other risk factors

We investigated the potential associations between clinical characteristics, risk scores, and the prognosis of patients in TCGA-THCA. Initially, we performed univariate Cox regression analysis, which revealed significant relationships between age (HR =1.155, 95% CI: 1.080–1.235, P<0.001), risk score (HR =3.678, 95% CI: 1.975–6.851, P<0.001), and T stage (HR =2.443, 95% CI: 1.100–5.427, P=0.028) with OS (P<0.05) (*Figure 10A* and *Table 6*). Subsequently, we performed a multivariate Cox regression analysis, incorporating these factors as covariates. This analysis indicated that age (HR =1.147, 95% CI: 1.058–1.244, P<0.001) emerged as an independent prognostic factor for OS in patients with THCA; while the risk score (HR =2.592, 95% CI: 0.972–6.914, P=0.057) exhibited a trend towards becoming an independent prognostic factor for patient OS (*Figure 10B* and *Table 7*). Furthermore, when stratifying patients with THCA exhibiting an early T stage, we observed a notable trend indicating a potential association between the risk score and patient OS (P=0.202) (*Figure 10C*). In contrast, among patients exhibiting a late T stage, the risk scores demonstrated a significant association with patient OS (P=0.014) (*Figure 10D*). These findings suggested that the risk score may hold prognostic value, particularly in patients exhibiting an advanced T stage.
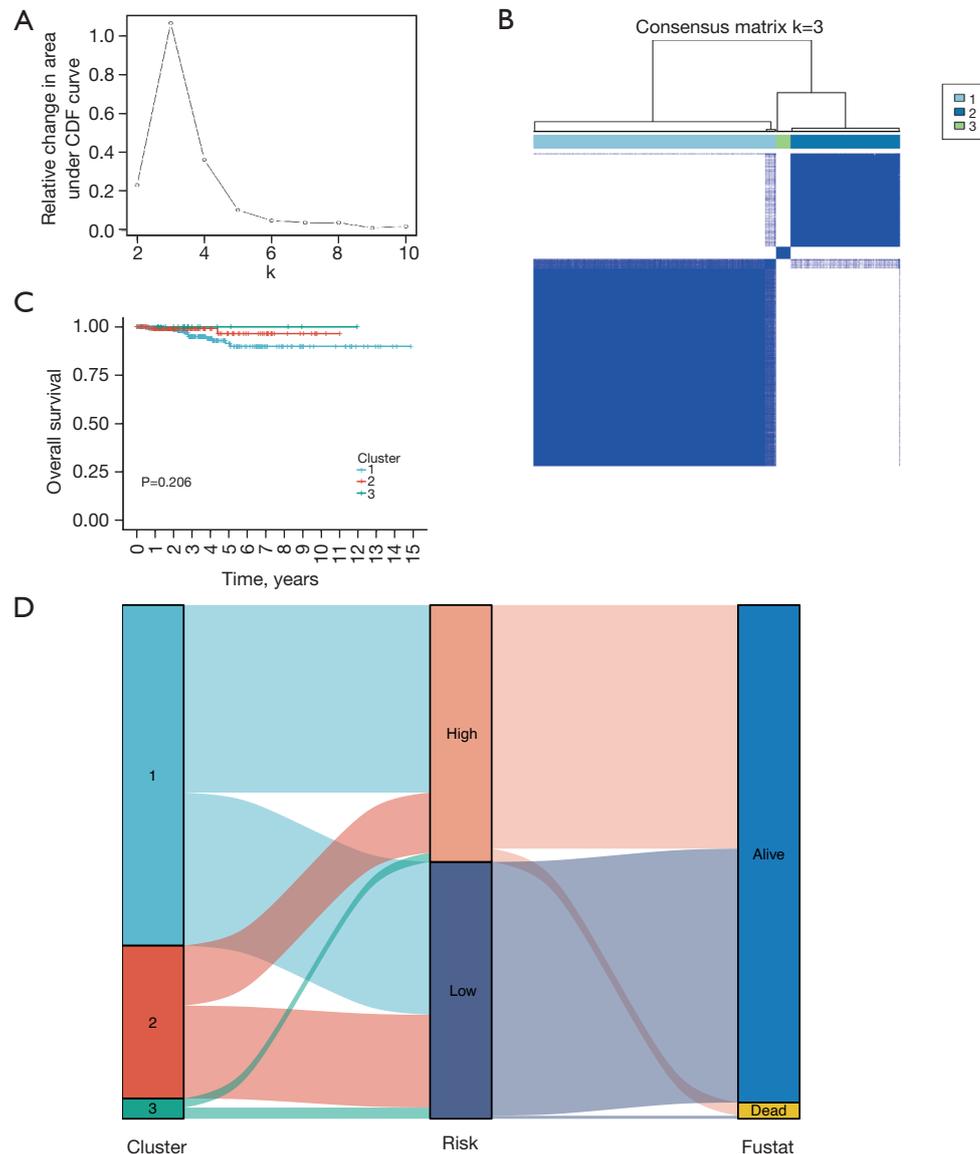
Figure 8 Concordance cluster analysis and survival. (A) Consistency cluster CDF. (B) Consistency cluster graph at K=3. (C) Survival differences among the three subtypes were obtained by concordance clustering. (D) Concordance clustering was used to obtain the Sankey map of the three types, high-risk group, low-risk group, and survival status. P<0.05 was considered to indicated statistical significance. CDF, cumulative distribution function.

Nomogram model construction and correlation analysis

In the analysis of clinical features within the TCGA-THCA dataset, we utilized R to eliminate features characterized as null or devoid of meaningful information. As a result of this data pre-processing, we narrowed down our dataset to five key features: age, gender, T, M, and N stages. Following this refinement, we employed both univariate and multivariate Cox proportional hazards regression analyses

to investigate prognostic risk scores and clinical factors. Notably, age emerged as a significant predictor of prognosis. Subsequently, we proceeded to construct a nomogram model designed to predict OS rates at 1, 3, and 5 years. This comprehensive model incorporated six essential variables: age, gender, T, M, and N stages, and prognostic risk scores, as illustrated in *Figure 11A*. To assess the predictive performance of this nomogram model, we employed the
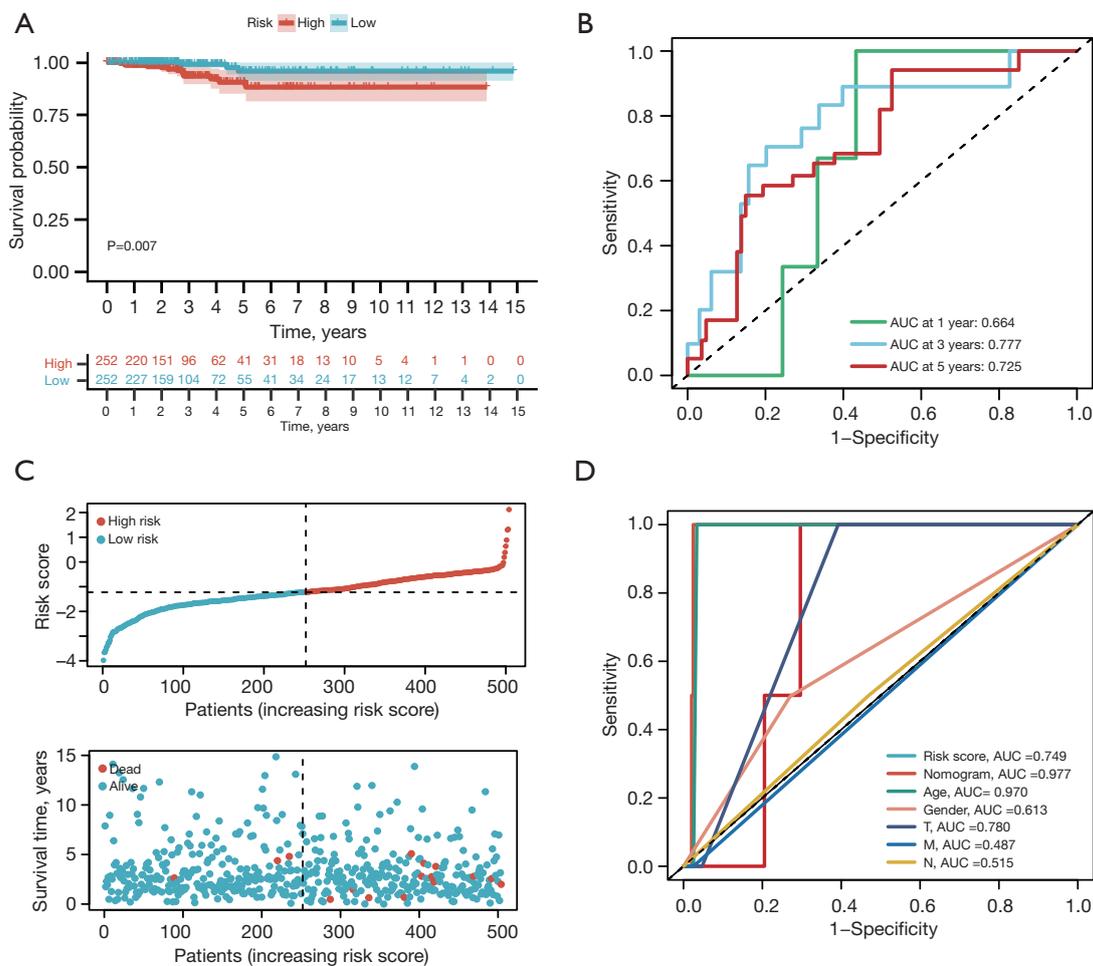
Figure 9 Prognostic model survival analysis. (A) TCGA-THCA datasets, survival curves of high- and low-risk groups differentiated by LASSO model risk scores. (B) ROC analysis of risk scores to assess sensitivity and specificity. (C) Relationship between risk score, mortality, and characteristic gene expression. (D) AUC values for risk score and clinical characteristics. P<0.05 was considered statistically significant. AUC, area under the curve; TCGA, The Cancer Genome Atlas; THCA, thyroid cancer; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic.

R package timeROC for time-dependent ROC analysis, focusing on OS status and survival time within the TCGA-THCA dataset. Remarkably, the 1-year AUC value for the nomogram exceeded 0.9, and the 3-year and 5-year AUC values outperformed those of the prognostic risk score. Furthermore, we conducted calibration analyses at 1, 2, and 3 years for variables identified through both univariate and multivariate Cox regression. Calibration curves were generated (*Figure 11B-11D*). In these plots, the vertical axis denotes the actual dataset's survival probability, while the horizontal axis represents the model-predicted survival probability at different time points. Various colored lines and data points correspond to predictions made by the

model at different time intervals. The closer these colored lines align with the gray ideal conditions line, the more accurate the model's predictive performance is at that specific time point. Collectively, the nomogram model reaffirms the reliability, prospective utility, and clinical applicability of our risk assessment model.

Identification of DEGs

The transcriptome expression microarray datasets GSE3467 and GSE3678 were subjected to normalization using the Limma package to conduct differential expression analysis. The results indicated that in the GSE3467 dataset, the

3376

Guo et al. Pyroptosis-related model in THCA

Table 5 Patient baseline dataset

Covariates	Risk	Total, n (%)	High, n (%)	Low, n (%)	Chi	P value
Age, year	≤65	240 (86.02)	108 (81.82)	132 (89.8)	3.0475	0.0809
	>65	39 (13.98)	24 (18.18)	15 (10.2)		
Gender	Female	203 (72.76)	97 (73.48)	106 (72.11)	0.0152	0.902
	Male	76 (27.24)	35 (26.52)	41 (27.89)		
Stage	Stage I–II	181 (64.87)	90 (68.18)	91 (61.9)	0.9429	0.3315
	Stage III–IV	98 (35.13)	42 (31.82)	56 (38.1)		
T	T1–2	166 (59.5)	84 (63.64)	82 (55.78)	1.4693	0.2255
	T3–4	113 (40.5)	48 (36.36)	65 (44.22)		
M	M0	272 (97.49)	129 (97.73)	143 (97.28)	0	>0.99
	M1	7 (2.51)	3 (2.27)	4 (2.72)		
N	N0	148 (53.05)	81 (61.36)	67 (45.58)	6.3385	0.0118
	N1	131 (46.95)	51 (38.64)	80 (54.42)		
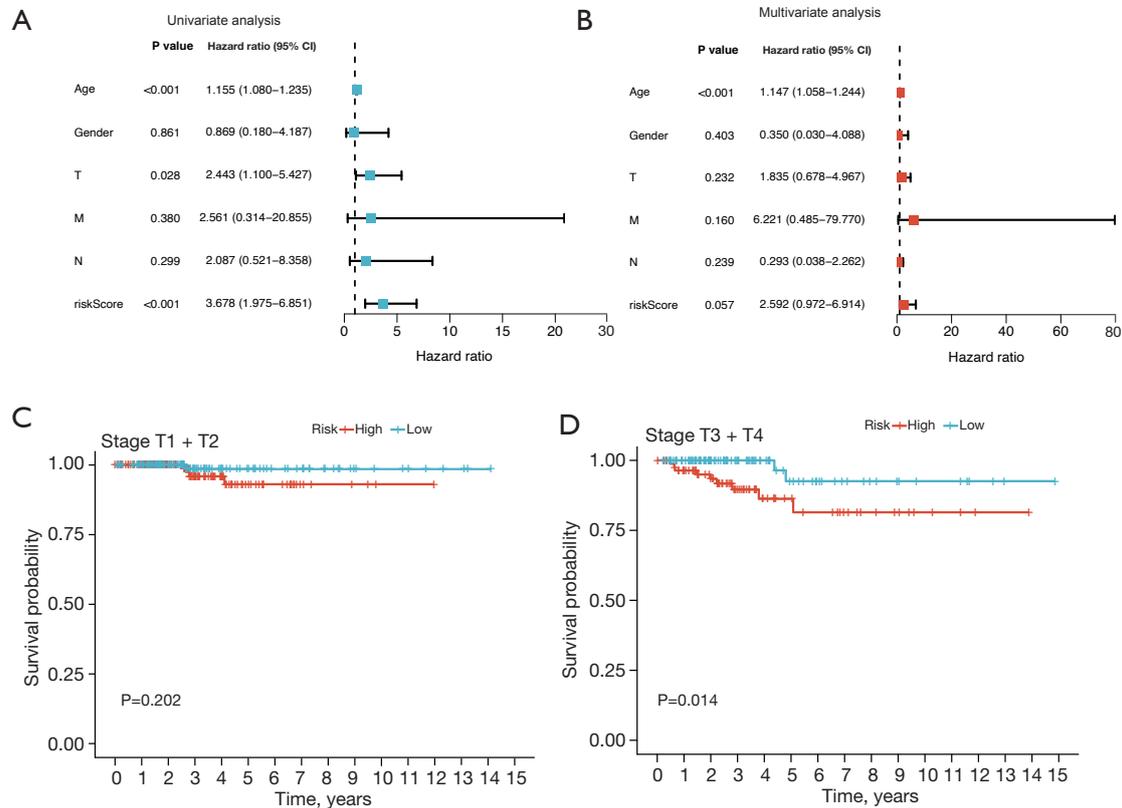
Figure 10 Prognostic risk score is a prognostic factor for OS of THCA. (A,B) Forrest plots for univariate (A) and multivariate (B) Cox regression analyses of each clinical index and risk score with survival. (C,D) Association of different risks with survival were assessed in patients with early T1 + T2 (C) and late T3 + T4 (D), classified according to T stage. P<0.05 was considered statistically significant. OS, overall survival; THCA, thyroid cancer.

Table 6 Univariate independent prognostic analysis

Covariates	HR	HR.95L	HR.95H	P value
Age	1.14731201	1.05826231	1.24385499	8.57E–04
Gender	0.35021186	0.03000393	4.08774258	0.40265139
T	1.83479304	0.67777362	4.96694683	0.23228511
M	6.22110043	0.48517109	79.7699856	0.16022265
N	0.29251221	0.03782106	2.26232171	2.39E–01
Risk score	2.59237623	0.97199705	6.91402767	5.70E–02

HR, hazard ratio; L, low; H, high.

Table 7 Multivariate independent prognostic analysis

Covariates	HR	HR.95L	HR.95H	P value
Age	1.15535705	1.08043547	1.23547398	2.43E–05
Gender	0.86870011	0.18024002	4.18686075	0.86075226
T	2.4428418	1.09958839	5.42700896	0.02830237
M	2.56064434	0.31439668	20.8554984	0.37958347
N	2.08680787	0.52102125	8.35813717	0.29877029
Risk score	3.67844748	1.97499631	6.85113982	4.05E–05

HR, hazard ratio; L, low; H, high.

expression of *TP63* and *IL6* was generally lower in tumor tissues, while *NOD1* and *BAX* exhibited significantly higher expression levels in tumor tissues compared to normal tissues (*Figure 12A,12B*). In contrast, within the GSE3678 dataset, only *NOD1* displayed significant differential expression, characterized by elevated expression in tumor tissues. Subsequently, we acquired IHC staining images from the THCA tissue microarray and *NOD1* mRNA expression data from the HPA database. Notably, *BAX* demonstrated a certain degree of expression in normal thyroid tissues (*Figure 12C*, left one and left two) but exhibited a moderate to high level of expression in tumor tissues, with staining localized to the cytoplasm and cytoplasmic membrane (*Figure 12C*, right one and right two). Furthermore, *NOD1* exhibited high expression levels in THCA, as confirmed through pan-cancer expression comparison (*Figure 12D*).

Discussion

THCA is a prevalent endocrine malignancy that has exhibited a significant increase in incidence over recent decades. Despite notable advancements in surgical

resection, radioactive iodine therapy, drug therapy, targeted therapy, immunotherapy, and the utilization of kinase inhibitors, these approaches have their respective limitations. Consequently, there remains a pressing need to further refine and optimize treatment modalities to enhance disease response.

Cancer manifests two fundamental characteristics: uncontrolled cell proliferation and the promotion of tumor-related inflammation (42). These attributes enable cancer cells to acquire genomic alterations, leading to genomic instability (43). Study has highlighted the pivotal role of pyroptosis in the proliferation, invasion, and metastasis of tumors. Pyroptosis induces cellular swelling, rupture of the plasma membrane, chromatin fragmentation, and the release of pro-inflammatory intracellular contents. This underscores the potential utility of pyroptosis-related genes as diagnostic markers and therapeutic targets (44).

In this study, we identified six DEPRGs from TCGA database. Furthermore, through LASSO Cox regression analysis, we identified *IL6*, *TP63*, *NOD1*, and *BAX* as candidate genes associated with the prognosis of TCGA-THCA patients. These four DEPRGs collectively constituted the prognostic risk score. Subsequently, we

3378

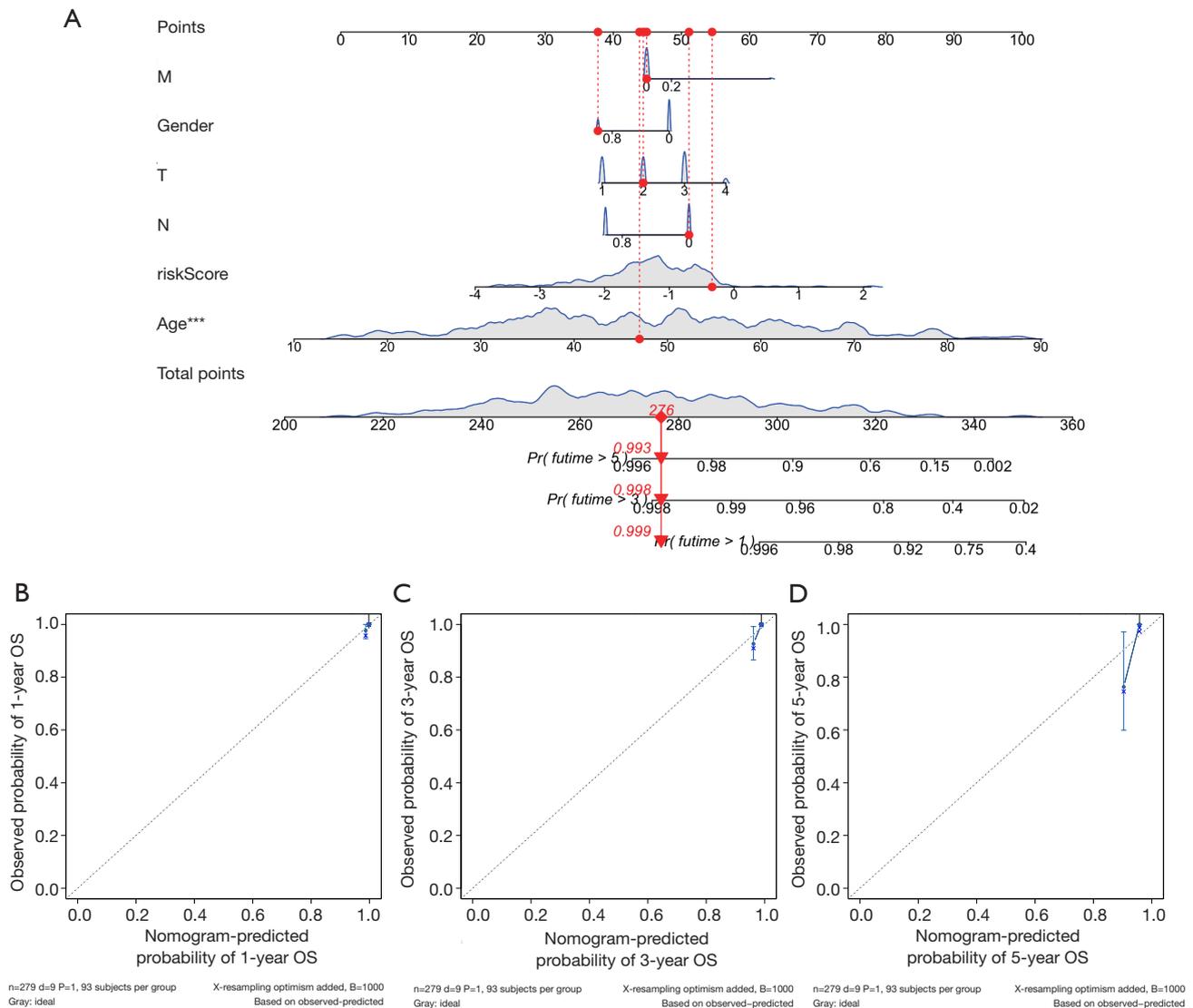Guo et al. Pyroptosis-related model in THCA

Figure 11 Nomogram model construction. (A). Nomograms established based on a combination of risk score, age, and TNM stage were used to predict the OS of patients with THCA. (B-D) Calibration charts for predicting 1-year (B), 3-year (C), and 5-year (D) survival. The horizontal and vertical axes represent the predicted survival probability and the actual survival probability, respectively. ***, represents highly significant differences with P<0.001. OS, overall survival; TNM, tumor-node-metastasis; THCA, thyroid cancer.

conducted an enrichment analysis of pyroptosis-related genes using GO and KEGG. We categorized all TCGA-THCA patients into two risk groups based on their prognostic risk scores and employed box plots to illustrate the differential expression of each gene within the high- and low-risk groups. Our findings revealed that *BAX* and *NOD1* exhibited higher expression levels, whereas *IL6* and *TP63* showed lower expression levels. Specifically, *BAX*, *TP63*, and *NOD1* were significantly overexpressed in the low-risk population, while *IL6* tended to be overexpressed in the

high-risk population. We performed survival analysis and visualization using the R survival package and survminer package. The results demonstrated that patients with high *NOD1* expression exhibited prolonged survival. However, the expression levels of *IL6*, *TP63*, and *BAX* genes were not associated with survival outcomes. Further analysis involved the exploration of the biological functions of genes within the two risk groups through GSEA and GSVA. GSEA revealed 18 distinct enrichment pathways between the high-risk and low-risk groups, and the pathways of the
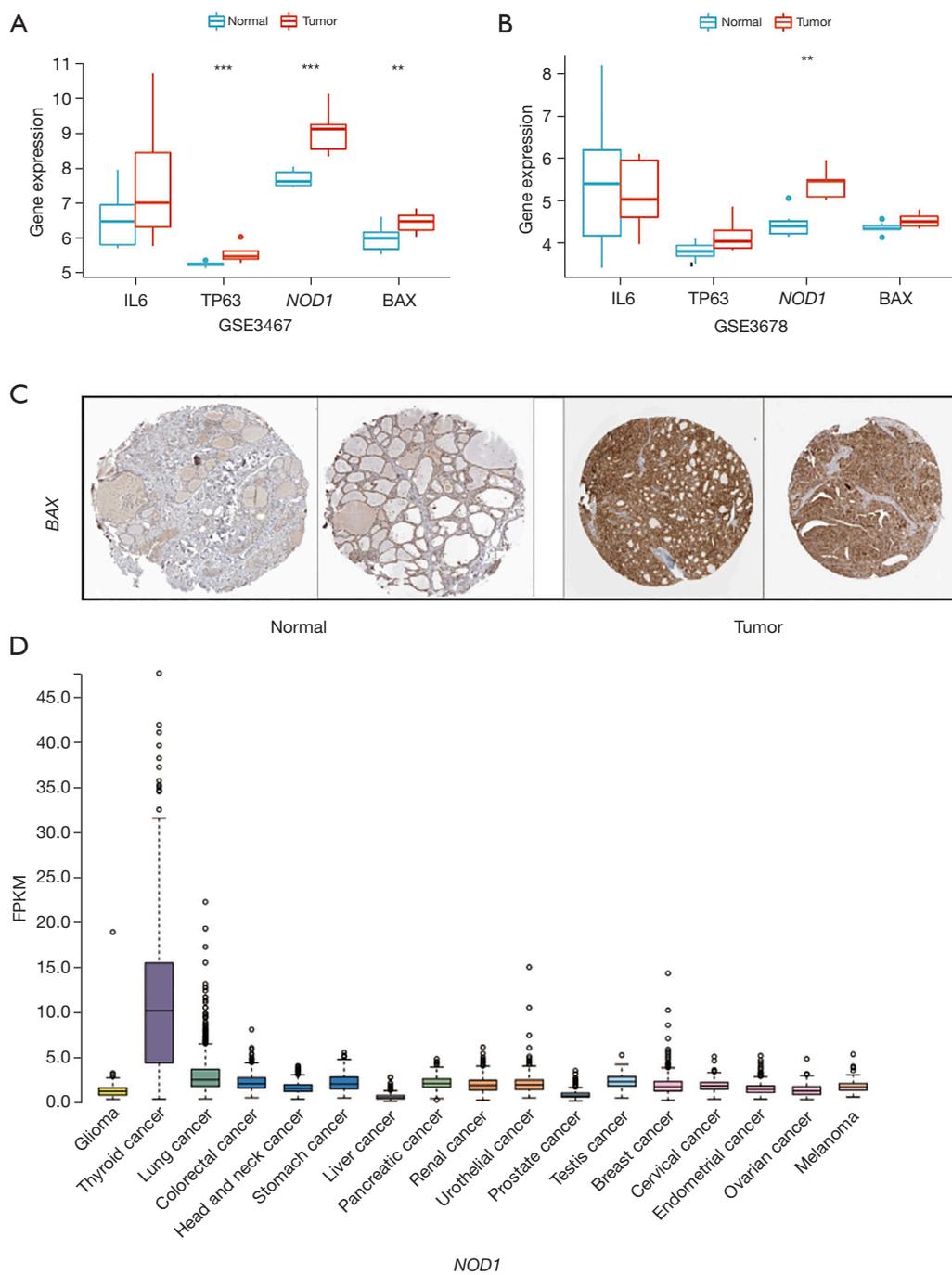
Figure 12 Validation of pyroptosis-related gene expression. (A) Differential expression of key pyroptosis-related genes between THCA tissues (tumor group) and normal tissues (normal group) in the GSE3467 dataset. (B) Differential expression of key pyroptosis-related genes between THCA tumor and normal tissues in the GSE3678 dataset. (C) Expression of BAX in normal thyroid tissues (left one and left two) and tumor tissues (right one and right two) in TCGA-THCA. From the Human Protein Atlas database (Normal: https://www.proteinatlas.org/ENSG00000087088-BAX/pathology/thyroid+cancer#img; Tumor: https://www.proteinatlas.org/ensg00000087088-bax/pathology/thyroid+gland), immunohistochemical method, hematoxylin staining. (D) mRNA expression of *NOD1* in pan-cancerous tissues in TCGA. **, P<0.01; ***, P<0.001. P<0.05 was considered statistically significant. THCA, thyroid cancer; TCGA, The Cancer Genome Atlas; THCA, thyroid cancer; FPKM, fragments per kilobase million.

low- and high-risk groups were selected for display. GSVA revealed 151 different pathways between the high- and low-risk groups, with the top four pathways displayed for each group. We conducted a consensus clustering of gene expression across 52 pyroptosis-related genes, revealing that K=3 was the optimal cluster. Subsequent survival analysis demonstrated no significant differences in OS among the three clusters. To enhance prognostic modeling, we developed a nomogram that incorporated the prognostic risk score along with clinical factors, including age, gender, T, M, and N stages. We assessed the performance of the nomogram through time-dependent ROC analysis at 1-, 3-, and 5-year intervals in the TCGA-THCA dataset, revealing its superiority over the prognostic risk score alone. Our consensus clustering-based survival analysis and multivariable Cox analysis were visually presented using nomograms. Finally, we validated the expression of DEPRGs using GEO and HPA datasets.

Multivariate analysis demonstrated that both risk score and age were independent prognostic factors. Examination of the HPA and GEO databases revealed high expression levels of both *BAX* and *NOD1* in THCA.

Some *BAX* activators have recently emerged as promising candidates for cancer therapy (14). Study has shown that co-targeting of *BAX* and BCL-XL is substantially effective in overcoming cancer resistance to apoptosis (45). Nevertheless, certain study has reported no significant association between *BAX* polymorphism and cancer susceptibility; however, *BAX* polymorphism has shown a significant association with poor prognosis (46). Our findings indicated a significant correlation between THCA and *BAX* expression levels.

The recognition of Microbe-Associated Molecular Pattern (MAMP) by NOD1 has been demonstrated to impact hepatocarcinogenesis and the development of liver damage (47). Furthermore, *NOD1* has been found to upregulate the tumorigenicity of human cervical squamous cell carcinoma, thereby promoting cancer progression and metastasis (48). A study has revealed that *NOD1/* CARD4 polymorphisms can alter the balance between proinflammatory and anti-inflammatory cytokines, thereby regulating the risk of infection, chronic inflammation, and cancer development (49). *NOD1*/CARD4 polymorphism has been associated with gastric cancer, colorectal cancer, mammary cancer, ovarian cancer, prostate cancer, testis cancer, lung cancer, larynx cancer, liver cancer, gallbladder cancer, bile cancer, pancreas cancer, small intestine cancer, kidney cancer, bladder cancer, skin

cancer, non-thyroid endocrine cancer, lymphoma, and leukemia (49). While studies have not explored the relationship between THCA and *NOD1*, our research demonstrated an association between *NOD1* expression and the development of THCA, potentially mediated through the enrichment pathway we investigated. Further research is warranted.

IL6 is produced and secreted by various cell types, including tumor cells, and is associated with the proliferation and differentiation of malignant cells. Elevated levels of IL6 are commonly observed in both serum and tumor tissues across various cancer types, such as colorectal, breast, prostate, ovarian, pancreatic, lung, renal cell, cervical, and multiple myeloma cancers (17). In this study, we also observed significant disparities in IL6 expression between THCA tissues and normal thyroid tissues, thereby corroborating previous findings.

We identified 18 distinct enrichment pathways through GSEA and 151 significantly different enrichment pathways through GSVA. From this comprehensive pool of pathways, we focused on the top-ranked enrichment pathways for further investigation. The results revealed differences in gene expression between high- and low-risk groups, suggesting potential implications across multiple facets.

Functional enrichment analysis was conducted on all potential DEPRGs. GO enrichment analysis revealed that the genes associated with cell death were predominantly enriched in BPs such as host mitotic cell dynamics processes. In terms of MFs, these genes were associated with cysteine-type peptidases and binding to phosphatidylinositol-4,5-bisphosphate. These MFs are integral to apoptosis signaling pathways, indicating the involvement of the four candidate pyroptosis-related genes in the initiation and progression of THCA. Additionally, KEGG enrichment analysis demonstrated that the four candidate pyroptosis-related genes were primarily associated with "cell adhesion molecule" and "oxidative phosphorylation", which are implicated in cancer-related signaling pathways. These results are consistent with previous research (50) and corroborate the accuracy of our findings. In contrast to previous studies on THCA and pyroptosis-related genes, we also performed a GSVA, which revealed that these four crucial pyroptosis-related genes were also associated with "cell adhesion molecule" and cancer-related signaling pathways. It is plausible that the four-cell death-related genes may be involved in the development of THCA by regulating the expression of related genes. Additionally, through KEGG analysis and GSVA, we found that these

genes were linked to signaling pathways associated with "asthma". This novel finding has the potential to contribute to future research endeavors.

We validated the expression of DEPRGs using GEO and HPA datasets. Subsequently, we performed survival predictions using the nomogram model and compared the outcomes to actual analyses. The results exhibited statistical significance, underscoring the reliability of our prediction model. Furthermore, the nomogram model reaffirmed the reliability, prospective value, and clinical applicability of our risk model. Our findings are in alignment with those of a prior study that established an association between NOD1 and THCA (50). Additionally, we identified distinct variants of the scorch gene in other types of cancer (50). Notably, the previous study exclusively relied on TCGA dataset, while we leveraged a joint analysis of both the GEO and TCGA datasets, thereby enhancing external validation. Furthermore, our study introduced a well-defined technology roadmap to streamline the workflow, ensuring a more organized approach. Moreover, we expanded our study by incorporating additional results from IHC analyses, nomogram utilization, and model validation assessments. This comprehensive approach brought us closer to the clinical analysis results. Both studies culminated in clinical analyses, which corroborate our findings. Notably, our study featured a larger sample size and greater reliability compared to the previous study (50). Furthermore, while our study primarily focused on PTC, it also encompassed samples from other subtypes of THCA. This inclusivity permitted us to undertake a more comprehensive analysis. In contrast, the 2022 article (50) exclusively examined PTC, indicating certain limitations in their study. We anticipate the expansion of our sample size and further validation of our findings in future research endeavors.

At present, there is limited research on THCA at the cellular level in relation to cell death. Investigating its potential mechanisms holds substantial future significance. Examining the prognostic relevance of pyroptosis-related genes will establish a fundamental basis for forthcoming mechanistic investigations. Nonetheless, our study has a few limitations. For instance, additional basic experiments are required to explore the correlation between the model and the tumor microenvironment.

Conclusions

In summary, we conducted a comprehensive analysis of datasets sourced from TCGA and MSigDB datasets. Using

these datasets, we developed a predictive nomogram model encompassing four genes: *IL6*, *TP63*, *NOD1*, and *BAX*, in conjunction with five clinical factors. This model may serve as a valuable tool for forecasting the prognosis of THCA and offering pertinent insights for optimizing therapeutic strategies for THCA.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://tcr.amegroups.com/article/view/10.21037/tcr-23-810/rc

Peer Review File: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-23-810/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tcr.amegroups.com/article/view/10.21037/tcr-23-810/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study complies with the Declaration of Helsinki (as revised 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Trigo JM, Capdevila J, Grande E, et al. Thyroid cancer: SEOM clinical guidelines. Clin Transl Oncol

3382

Guo et al. Pyroptosis-related model in THCA

2014;16:1035-42.

2. Kim J, Gosnell JE, Roman SA. Geographic influences in the global rise of thyroid cancer. Nat Rev Endocrinol 2020;16:17-29.

3. Baloch ZW, Asa SL, Barletta JA, et al. Overview of the 2022 WHO Classification of Thyroid Neoplasms. Endocr Pathol 2022;33:27-63.

4. Seib CD, Sosa JA. Evolving Understanding of the Epidemiology of Thyroid Cancer. Endocrinol Metab Clin North Am 2019;48:23-35.

5. Grewal RK, Ho A, Schöder H. Novel Approaches to Thyroid Cancer Treatment and Response Assessment. Semin Nucl Med 2016;46:109-18.

6. Naoum GE, Morkos M, Kim B, et al. Novel targeted therapies and immunotherapy for advanced thyroid cancers. Mol Cancer 2018;17:51.

7. Schneider DF, Chen H. New developments in the diagnosis and treatment of thyroid cancer. CA Cancer J Clin 2013;63:374-94.

8. Ratajczak M, Gaweł D, Godlewska M. Novel Inhibitor-Based Therapies for Thyroid Cancer-An Update. Int J Mol Sci 2021;22:11829.

9. Fang Y, Tian S, Pan Y, et al. Pyroptosis: A new frontier in cancer. Biomed Pharmacother 2020;121:109595.

10. Kovacs SB, Miao EA. Gasdermins: Effectors of Pyroptosis. Trends Cell Biol 2017;27:673-84.

11. Ruan J, Wang S, Wang J. Mechanism and regulation of pyroptosis-mediated in cancer cell death. Chem Biol Interact 2020;323:109052.

12. Ma X, Guo P, Qiu Y, et al. Loss of AIM2 expression promotes hepatocarcinoma progression through activation of mTOR-S6K1 pathway. Oncotarget 2016;7:36185-97.

13. Shao W, Yang Z, Fu Y, et al. The Pyroptosis-Related Signature Predicts Prognosis and Indicates Immune Microenvironment Infiltration in Gastric Cancer. Front Cell Dev Biol 2021;9:676485.

14. Ye Y, Dai Q, Qi H. A novel defined pyroptosis-related gene signature for predicting the prognosis of ovarian cancer. Cell Death Discov 2021;7:71.

15. Lin W, Chen Y, Wu B, et al. Identification of the pyroptosis-related prognostic gene signature and the associated regulation axis in lung adenocarcinoma. Cell Death Discov 2021;7:161.

16. Liu Z, Ding Y, Ye N, et al. Direct Activation of Bax Protein for Cancer Therapy. Med Res Rev 2016;36:313-41.

17. Wang Z, Jensen MA, Zenklusen JC. A Practical Guide to The Cancer Genome Atlas (TCGA). Methods Mol Biol 2016;1418:111-41.

18. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.

19. Bruford EA, Antonescu CR, Carroll AJ, et al. HUGO Gene Nomenclature Committee (HGNC) recommendations for the designation of gene fusions. Leukemia 2021;35:3040-3.

20. Karki R, Kanneganti TD. Diverging inflammasome signals in tumorigenesis and potential targeting. Nat Rev Cancer 2019;19:197-214.

21. Xia X, Wang X, Cheng Z, et al. The role of pyroptosis in cancer: pro-cancer or pro-"host"? Cell Death Dis 2019;10:650.

22. Wang B, Yin Q. AIM2 inflammasome activation and regulation: A structural perspective. J Struct Biol 2017;200:279-82.

23. Man SM, Kanneganti TD. Regulation of inflammasome activation. Immunol Rev 2015;265:6-21.

24. Ito K, Murphy D. Application of ggplot2 to Pharmacometric Graphics. CPT Pharmacometrics Syst Pharmacol 2013;2:e79.

25. Kolde R, Kolde MR. Package 'pheatmap'. R Package 2015;1:790.

26. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 2011;12:35.

27. Yu G. Gene Ontology Semantic Similarity Analysis Using GOSemSim. Methods Mol Biol 2020;2117:207-15.

28. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27-30.

29. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284-7.

30. Friedman J, Hastie T, Tibshirani R, et al. Package 'glmnet'. Package 'Glmnet'. CRAN R Repositary 2021. Available online: https://cran.r-project.org/web/packages/glmnet/index.html

31. Wei T, Simko V, Levy M, et al. Package 'corrplot'. Statistician 2017;56:e24. Available online: https://github.com/taiyun/corrplot

32. Therneau TM, Grambsch PM. Modeling Survival dataset: Extending the Cox Model; 2013. Available online: https://link.springer.com/chapter/10.1007/978-1-4684-6316-3_5

33. Kassambara A. Drawing Survival Curves using 'ggplot2' [R package survminer version 0.2.0]. 2017. Available online: https://cran.rproject.org/web/packages/survminer/survminer.pdf

34. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.

35. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 2013;14:7.

36. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. Stat Med 2013;32:5381-97.

37. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics 2010;26:1572-3.

38. Jiang H, Li Y, Wang T. Comparison of Billroth I, Billroth II, and Roux-en-Y reconstructions following distal gastrectomy: A systematic review and network meta-analysis. Cir Esp (Engl Ed) 2021;99:412-20.

39. Jin C, Cao J, Cai Y, et al. A nomogram for predicting the risk of invasive pulmonary adenocarcinoma for patients with solitary peripheral subsolid nodules. J Thorac Cardiovasc Surg 2017;153:462-469.e1.

40. Tataranni T, Piccoli C. Dichloroacetate (DCA) and Cancer: An Overview towards Clinical Applications. Oxid Med Cell Longev 2019;2019:8201079.

41. Thul PJ, Lindskog C. The human protein atlas: A spatial map of the human proteome. Protein Sci 2018;27:233-44.

42. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646-74.

43. Shen Z. Genomic instability and cancer: an introduction. J Mol Cell Biol 2011;3:1-3.

44. Sartorelli V, Lauberth SM. Enhancer RNAs are an important regulatory layer of the epigenome. Nat Struct Mol Biol 2020;27:521-8.

45. Lopez A, Reyna DE, Gitego N, et al. Co-targeting of BAX and BCL-XL proteins broadly overcomes resistance to apoptosis in cancer. Nat Commun 2022;13:1199.

46. Feng Y, Chen X, Zheng Y, et al. Prognostic value and susceptibility of BAX rs4645878 polymorphism in cancer: A systematic review and meta-analysis. Medicine (Baltimore) 2018;97:e11591.

47. Omaru N, Watanabe T, Kamata K, et al. Activation of NOD1 and NOD2 in the development of liver injury and cancer. Front Immunol 2022;13:1004439.

48. Zhang Y, Li N, Yuan G, et al. Upregulation of NOD1 and NOD2 contribute to cancer progression through the positive regulation of tumorigenicity and metastasis in human squamous cervical cancer. BMC Med 2022;20:55.

49. Fisher ML, Balinth S, Mills AA. p63-related signaling at a glance. J Cell Sci 2020;133:jcs228015.

50. Liang Y, Zhang Q, Xin T, et al. A four-enhancer RNA-based prognostic signature for thyroid cancer. Exp Cell Res 2022;412:113023.