Characterization of pyroptosis-related genes in esophageal cancer and construction of a prognostic model

S.-C. JIANG¹, S.-H. TAO², S.-Y. CHEN³, H. XIE⁴, Y.-J. FENG⁵

¹Division of Thoracic and Endocrine Surgery, University Hospitals and University of Geneva, Geneva, Switzerland
²General Surgery Department, The Fourth Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China
³Jiangxi Institute of Applied Science and Technology, Nanchang, Jiangxi, China
⁴Thoracic Surgery Department, Jiangxi Province Cancer Hospital, Nanchang, Jiangxi, China
⁵Department of Oncology, Shanghai Artemed Hospital, Shanghai, Shanghai, China

Abstract. – OBJECTIVE: Esophageal cancer (EC) is a highly malignant digestive system tumor that often lacks evident early symptoms and has a poor prognosis. Pyroptosis, a form of programmed cell death, has been shown to be associated with the occurrence and progression of many malignancies. However, its role in esophageal cancer remains unclear. This work aimed to evaluate the prognostic value of pyroptosis-related genes (PRGs) in EC using data from The Cancer Genome Atlas (TCGA) cohort.

MATERIALS AND METHODS: The RNA-seq data from 171 esophageal samples in the TCGA database were employed. Differential expression genes (DEGs) between tumor and non-tumor samples were compared. Protein-protein interaction (PPI) networks were constructed using the STRING database, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment and Gene Ontology (GO) analyses were performed using the “clusterProfiler” package in R. Furthermore, based on the DEGs, all esophageal cancer cases were classified into three subtypes. A risk model for gene features was established using the LASSO regression method, and EC patients in the TCGA cohort were divided into high-risk and low-risk groups.

RESULTS: A total of 614 PRGs were identified. Among them, 32 DEGs (31 upregulated and 1 downregulated) were found between normal esophageal tissue and EC tissue. PPI analysis identified key genes including IL-1β, CASP1, AIM2, HMGB1, GSDMD, PYCARD, IL-18, BAK1, and TP53. On the other hand, the low-risk group exhibited a significantly higher survival rate than the high-risk group (p < 0.001). Combined with the clinical characteristics of the TCGA cohort, it was found that the risk score was an independent prognostic factor for overall survival (OS) prediction in EC patients. KEGG and GO analyses revealed the enrichment of genes associated with cell proliferation in the high-risk group.

CONCLUSIONS: PRGs play a crucial role in the occurrence and development of EC and can be used to predict the prognosis of EC patients.

Key Words: Esophagus cancer, Pyroptosis-related genes, Prognostic model, Prognostic value.

Introduction

Esophageal cancer (EC) is a worldwide malignant disease and is among the most frequent malignancies of the digestive tract. Based on recent data, in 2018, there were over 570,000 new cases of EC and over 500,000 deaths caused by EC worldwide, ranking 7th and 6th, respectively, which have been affecting human health. Among many countries, China has a high incidence of EC, in which esophageal squamous cell carcinoma (ESCC) is predominant, accounting for about 70% of the global incidence. With the improvement of medical diagnosis and treatment, more and more patients with esophageal cancer can benefit from early diagnosis and treatment. When it comes to treating EC, surgical excision is still the primary method, often combined with radiotherapy and chemotherapy, for optimal results. In recent years, radiotherapy and chemotherapy...
have been a focal point of research in this field\textsuperscript{4,5}. Although much progress has been achieved in treating EC, its prognosis is still poor, and the 5-year survival rate is below 25\%\textsuperscript{6}. At present, there is no good way of predicting the EC prognosis. The mainstream prediction of EC prognosis is still based on the TNM stage. However, studies\textsuperscript{7} have shown that patients receiving similar treatment regimens during the same TNM phase frequently have different clinical outcomes. This is because of the heterogeneity and individual differences of tumors, and TNM staging do not entirely reflect the progression of internal biological processes together with the pathologic state of the tumor\textsuperscript{8}. In conclusion, there are still many uncertainties in EC, and the prognosis is still not optimistic. It is extremely important to find reliable biological markers and prognostic models for early diagnosis and treatment. The individualized formulation of clinically targeted therapy programs is of great significance.

Pyroptosis is a new form of cell death. Cell death is the terminal stage of cell metabolism and plays an essential role in biological metabolism and development together with the progression of diseases. Cell death is divided into necrosis and programmed cell death (PCD). Pyroptosis is a form of PCD. The word “pyroptosis” comes from the Greek root Pyro, relating to heat or fire, while “ptosis” means to fall. Therefore, the combination of the two words reflects the inflammatory nature of this mode of cell death\textsuperscript{9}. The concept of pyroptosis was discovered and proposed in 2001 by Cookson and Brennan\textsuperscript{10} to characterize the manner of PCD that accompanied the inflammatory response. During pyroptosis, cells undergo nuclear pyknosis, DNA fragmentation, swelling, and finally rupture, accompanied by obvious blebbing and the release of numerous inflammatory factors, for instance, interleukin-18 (IL-18) and interleukin-1β (IL-1β)\textsuperscript{11,12}. In the microscopic view, pyroptosis cells first develop numerous vesicles, which act on the cell membrane to form holes and eventually lead to the rupture of cell membranes and inflammatory factor release. In 2006, pyroptosis was described\textsuperscript{13} as the programmed death of cells mediated by caspase-1. However, how the activation of inflammatory caspases causes pyroptosis has remained unanswered. This question was not answered until the publication of two independent studies\textsuperscript{14,15} in 2015. The authors all discovered that Gasdermin D (GSDMD) is a substrate for inflammatory caspases. On the one hand, caspase-1, which is activated, cleaves the precursors of IL-18 along with IL-1β for the inflammatory factor release, and it cleaves Gasdermin-D (GSDMD) protein that eventually causes cells to create pores in the cell membrane and result in pyroptosis\textsuperscript{16}. Gasdermin proteins are crucial in the development of pyroptosis. They undergo cleavage and multimerization, resulting in the cleavage of C- and N-terminal linkage domains. This activation leads to the release of the N-terminal region. Such regions are associated with cell membrane lipids and phosphatidyl Inositol, etc. and are combined and localized in the pores of the cell membrane, where they aggregate on the inside of the cell membrane to form pores, triggering cell pyroptosis\textsuperscript{17,18}. Pyroptosis was considered one of the key mechanisms against infection in the early days of its discovery, but with the increasing and in-depth study of pyroptosis, follow-up research\textsuperscript{19,20} have indicated that it also has a critical role in tumor progression. Proteins of the Gasdermin family together with pro-inflammatory cytokines have been described\textsuperscript{21} to be implicated in tumorigenesis, invasion as well as metastasis. Recent research\textsuperscript{22} also demonstrated the critical effect of pyroptosis on the anti-tumor function. Based on available studies\textsuperscript{22,24}, we know that pyroptosis acts in tumorigenesis together with anti-tumor processes, but its function in the clinical value and prognosis is poorly examined. Hence, we implemented a systematic study to assay the expression levels of PRGs in both cancer and normal esophageal tissues, to analyze the prognostic value of such genes, and to examine the association of pyroptosis with the immunological aspects and tumor microenvironment. This lays the basis for the therapeutic and prognostic assessment of EC.

Materials and Methods

Pyroptosis-Related Gene Datasets and Patient Samples

The data in the current study were taken from the TCGA, a public database of 171 esophageal RNA-seq data (available at: https://portal.gdc.cancer.gov), and their clinical data were downloaded. The gene-expression data were normalized by ap-
plying the scaling approach of the ‘limma’ package (available at: https://www.bioconductor.org/packages/limma/). Through bioinformatics data mining and previous literature reports, 614 pyroptosis-related genes (PRGs) associated with were screened25 (Supplementary Table I).

Identification of Differentially Expressed PRGs

PRGs expression in TCGA was assessed to determine DEGs between tumor and non-tumor samples. When \( p < 0.05 \), the DEGs difference between the two samples was considered to be statistically significant. DEGs were determined with the “limma” package of R software. With an interacting gene search tool (STRING, version 1.0, available at: https://string-db.org/), the protein-protein interaction (PPI) network was examined to detect the interactions of PRGs in the current work. DEGs are presented in Supplementary Table II.

Development and Verification of Prognostic Models for PRGs

The PRGs were screened by utilizing the univariate regression analysis. \( p < 0.2 \) was defined as the threshold value by which survival-associated genes were determined for follow-up analysis. Cox regression analysis was applied to assess the PRGs prognostic value in the TCGA database. The candidate genes most correlated with pyroptosis were screened utilizing the least absolute shrinkage and selection operator (LASSO)-Cox regression model, which was employed for building the prediction model. LASSO-Cox regression analysis was carried out with the ‘glmnet’ package of R software. Risk scores were acquired from standardized EC mRNA expression data in the TCGA dataset. Risk score = \( (X: \text{coefficients}, Y: \text{gene expression level}) \). Patients were categorized as low- and high-risk groups based on median risk score, and the analysis of OS was conducted for both groups. Principal component analysis (PCA) based on PRG characteristics was implemented by utilizing the “prcomp” function in “stats” R package (available at: https://www.r-project.org/). ROC curves were plotted with “survminer” as well as “time-ROC” software packages of R. Lastly, multivariate and univariate Cox regression analyses were implemented to identify the model’s independent prognostic value.

Gene Set Enrichment Analyses

The analysis of KEGG and GO were conducted with “clusterProfiler” package of R (available at: https://cran.r-project.org/mirrors.html).

Statistical Analysis

Gene expression levels of tumor and non-tumor tissues were investigated through the Mann-Whitney U test. The comparison of count data was conducted with \( \chi^2 \) test. LASSO regression was employed for calculating the coefficients of prognostic characteristics. The survival rates between the subgroups were compared through a log-rank test together with the Kaplan-Meier method. Correlation analysis was conducted with the Pearson’s correlation test. Multivariate and univariate Cox regression models examined the model’s independent risk factors. Statistical analyses were conducted via IBM SPSS (version 26.0, IBM Corp., Armonk, NY, USA) and R software (version 4.1.1) (available at: https://www.r-project.org/). In all tests, a two-sided \( p < 0.05 \) was deemed statistically significant.

Result

Identification of DEGs between Tumor and Normal Tissues

In TCGA data, the expression levels of 614 PRGs in 160 tumors and 11 normal tissues were compared, and 32 DEGs (all \( p < 0.01 \)) were detected. Of these, one gene (ELANE) was down-regulated, and the other 31 genes (IL-18, PRKACA, NOD1, SCAF11, CHMP6, CASP4, CASP3, CASP8, CHMP4A, GPX4, CYCS, LCG1, CHMP4B, NOD2, IRF1, GSDMD, BAK1, HMGB1, CASP5, TP53, GSDMC, CASPI, BAX, PYCARD, GZMA, GZMB, IL-1β, TNF, AIM2, TP63, and IL-1A) were enriched in tumor group. The RNA levels of such genes are displayed as a heat map in Figure 1A (red and green denote the high and low expression levels, separately). To further investigate the interactions of such PRGs, we performed PPI analysis, and the findings are presented in Figure 1B. The minimal requirement for PPI analysis was a score of 0.4, and we identified IL-1β, CASPI, AIM2, HMGB1, GSDMD, PYCARD, IL-18, BAK1, TP53 as pivotal genes. All of these genes are DEGs between tumor and normal tissues. Figure 1C displays the correlation network comprising all PRGs (blue: negative correlation; red: positive correlation).

Tumor Classification based on DEGs

To investigate the association between the expression of 32 pyroptosis-associated DEGs and EC subtypes, we conducted a consensus clustering analysis of all 161 individuals with EC in the
Figure 1. Differential expression of 32 PRGs and their interactions. A. Heat map between tumor tissue (T, red) and normal EC tissue (N, bright blue) (red and green represent the high and low expression levels, separately). *p < 0.05; **p < 0.01; ***p < 0.001. B, PPI network revealing interaction network of PRGs (with the interaction score of 0.4). C, Net of interrelationships of PRGs (blue line and red line denote the negative and positive correlation). The depth of the color indicates the correlation strength.)
By raising the clustering variable (k) in a range of 2-10, we suggested the highest intra-group association and lower inter-group association when k = 3, reflecting that the 161 individuals with EC could be well grouped into three clusters based on the 32 DEGs (Figure 2A). Through these three categories, we examined the expression of the above genes and the clinical tumor characteristics of individuals with EC and found that there was no remarkable difference in the TNM stage of clinical tumors of esophageal cancer between these three clusters (Figure 2B). However, further comparison of the OS of the three groups presented a remarkable difference in the OS of patients with EC when k = 3 (p = 0.014, Figure 2C).

Progression of Prognostic Gene Models in the TCGA Cohort

Through the survival data of the TCGA dataset, 159 EC samples with full survival information were matched. The survival-associated genes were screened initially via univariate Cox regression analysis. Forty-five genes that met the p < 0.2 criterion were retained to perform the subsequent analysis, of these, 13 genes (ARHGAP10, BACH2, BNC2, FBLN2, KLF12, PDLIM4, PSTPIP2, SLIT3, SUS6D, TECPR2, TSHZ3, YAP1 and ZFPM2) were protective genes with the HRs of < 1. The residual 32 genes were linked to elevated risk with HRs > 1 (Figure 3A). A 13-gene signature was built based on the optimal λ value carrying out a LASSO Cox regression analysis (Figure 3B-C). Risk scores were presented as below: risk score = (0.0004 \times \text{ATP5MK exp.}) + (0.0219 \times \text{CA8 exp.}) + (0.0473 \times \text{CDH19 exp.}) + 0.0055 \times \text{CLDN12 exp.}) + (0.0009 \times \text{CYCS exp.}) + (0.0028 \times \text{ELF1N-ASI exp.}) + (0.0022 \times \text{GDF15 exp.}) + (0.0010 \times \text{NT5C3A exp.}) + (0.0082 \times \text{PRDX4 exp.}) + (-0.0228 \times \text{PSTPIP2 exp.}) + (0.0276 \times \text{RPS26P47 exp.}) + (0.0409 \times \text{RPS29P14 exp.}) + (-0.0275 \times \text{SLIT3 exp.}). The 159 individuals were categorized equally as high- and low-risk groups based on the cut-off value.
Figure 3. Risk profiles and model establishment for the TCGA cohort. A, Univariate Cox regression analysis was implemented on OS of PRGs and 45 genes with \( p < 0.2 \). B, LASSO regression of 45 genes linked to OS. C, Cross-validation to align regressions for parameter choice in the LASSO. D, Distribution of EC patients on the basis of risk score. E, PCA plots of patients with EC based on risk scores. F, Vital status for each patient (Low-risk groups: blue to the left of the dotted line; high-risk groups: red to the right of the dotted line). G, Kaplan-Meier curves of OS in both groups. H, ROC curves for predictive efficiency of survival based on risk score.
low-risk subgroups based on the median derived from the risk score formula (Figure 3D). PCA revealed that patients with varying risks were well grouped into two clusters (Figure 3E). In comparison to patients in the low-risk group, the high-risk group presented shorter survival time and more deaths (Figure 3F, on the right side of the dotted line). A pronounced difference in OS time was observed between both groups ($p < 0.001$, Figure 3G). Employing time-dependent ROC analysis to assess the specificity and sensitivity of the prognostic model, we revealed that the AUC was 0.756, 0.705, and 0.696 for 1-, 2-, and 3-year survival (Figure 3H).

**Independent Prognostic Value of Risk Model**

We utilized multivariate and univariate Cox regression analyses to examine whether the risk score obtained from the genetic trait model could be taken as an independent prognostic factor. Univariate Cox regression analyses revealed that M stage, N stage, and risk score were independent predictors of poor survival in the TCGA cohort ($p < 0.001$, HR = 6.931, 95% CI: 3.584-13.406; HR = 1.825, 95% CI: 1.300-2.560 and HR = 4.879, 95% CI: 2.224-10.706, Figure 4A). Multivariate analysis after adjustment for other confounding factors also indicated that risk score is still a prognostic factor (HR = 4.643, 95% CI: 2.341-9.207, Figure 4B) for EC individuals in the TCGA cohort. Besides, we produced a heat map of clinical characteristics for the TCGA cohort (Figure 4C) and showed that there were significant differences in distant metastasis (M stage) scores between the two subgroups ($p < 0.05$).

**Functional Analysis on the Basis of Risk Model**

With the aim of assessing the differences in pathways and gene function between subgroups classified according to the risk model, we derived DEGs using the “limma” R package applying the criteria of $|\log_2 FC| \geq 1$ together with FDR < 0.05. 412 DEGs were determined in the TCGA cohort for both risk groups. Of these, 153 genes were up, while 259 were down (the data are presented in Supplementary Table I). KEGG pathway analysis and GO enrichment analysis were subsequently implemented based on such DEGs. The findings showed that DEGs were primarily linked to the proliferation of epithelial cells, cell-cell signaling by Wnt signaling pathways, and focal adhesion (Figure 5A-B).

**Discussion**

Esophageal cancer is a tumor of the digestive system that is highly malignant, with a poor prognosis and high mortality rate and is the sixth major cause of cancer death around the world\(^26\). Although there have been significant advancements in the treatment and diagnosis of EC, many patients can still enhance their quality of life through surgical procedures or chemoradiotherapy. Several factors can affect the therapy and prognosis of EC patients, and the current evaluation methods are not yet perfect, which can hinder the proper care of patients. Pyroptosis is a form of PCD, which is a new hotspot in cancer research in recent years. At the same time, the effect of pyroptosis in cancer is becoming clearer, which provides another possibility for treating the patients. In the current work, we built a PRG signature consisting of 13 genes from gene expression data and clinical data of individuals with EC in TCGA database. With this model, patients with EC could be categorized into a low- and high-risk group, with marked differences in OS between both groups (Figure 3D). This suggested that the model has excellent prognostic assessment.

Pyroptosis is a new form of PCD that has been discovered\(^27\) in the last few years to play a dual role in the mechanisms of tumor regulation. Moreover, releasing inflammatory factors after pyroptosis transforms normal cells into tumor cells. Pyroptosis can also facilitate the death of tumor cells, which may be a new therapeutic direction\(^28\). With more and more studies\(^29,30\) on pyroptosis in the last few years, the number of PRGs has also increased year by year. In this study, a total of 614 PRGs, including atypical pyroptosis pathways, were identified. 32 of these genes were differentially expressed markedly in EC samples compared to normal samples. 13 genes were closely linked to survival outcomes. These findings indicate that PRGs are involved in modulating the progress and prognosis of EC. Among these 13 genes, each plays a different role and has some tumor specificity. For instance, *CLDN12*, a member of our gene signature, is involved in cell junction, maintaining cell polarity and permeability. Abnormal expression of *CLDN12* affects cell-cell junction, which is one of the important causes of peripheral invasion and distant metastasis of tumor cells. It can facilitate the proliferation, metastasis, and epithelial-mesenchymal transi-
Characterization of PRGs in esophageal cancer and construction of a prognostic model

Pyroptosis has become a hot research topic in recent years. Several studies have used bioinformatics methods to elucidate the role of PRGs in various cancers, for instance, lung adenocarcinoma, ovarian cancer, glioblastoma, and gastric cancer. Compared to these studies, our study has some advantages that cannot be ignored. Firstly, since the pyroptosis-related genes have been rarely studied in esophageal cancer, we have created a complete pyroptosis-associated genome (n = 614), which considerably enhanced the completeness of the former pyroptosis-associated genome (n = 45) by including the currently known participants in the pyroptosis bypass pathway. At the same time, we initially analyzed the prognostic value of differential PRGs, which offered support and theory for future studies. It is worth noting that the differentially expressed genes screened by our pyroptosis-related model are mainly related to proliferation and intercellular junction, which is different from previous reports that regulate pyroptosis and influence tumorigenesis and progression. This may be related to the own function of the gene in the model. CDH19, as a cadherin, is involved in intercellular contact and adhesion, and in cancer, loss of intercellular adhesion will lead to malignant migratory properties of tumor cells. Therefore, the two common pathways regulating pyroptosis in this study may not apply to esophageal cancer, and the promotion of tumor cell metastasis by affecting cell-cell junctions may be the main regulatory pathway in esophageal cancer.

Limitations

However, this study also has some limitations. TCGA data is limited, and it is unclear whether this result will be changed in a larger data set. Moreover, the risk characteristics of esophageal cancer pyroptosis genes lack external clinical cohort validation and animal experimental data. In the future, we will address the following research and problems: first, what factors contribute to the pro- or anti-tumor properties of pyroptotic genes? In addition, whether the selected regulators also play a corresponding role in the classical pyroptosis pathway of EC deserves further research.

Overall, our study indicated that pyrexia is strongly associated with EC and that there is a substantial number of differentially expressed
Figure 4. Multivariate and univariate Cox regression analysis of the risk scores. A. Univariate analysis of TCGA cohort (M: presence or absence of distant metastasis; N: lymph node staging; T: depth of tumor invasion). B. Multivariate analysis of TCGA cohort. C. Heat map of relations between risk groups and clinicopathological characteristics (red: high expression; blue: low expression; * p < 0.05).
Figure 5. Functional analysis based on DEGs between both risk groups in TCGA cohort. 

A. Bar plot of GO enrichment (longer bars denote more enriched genes, red depth represents more marked differences). 

B. The bubble map of KEGG pathway (the larger the bubble, the more genes are enriched, and the deeper the red color, the more evident differences; Q-value: adjusted p-value).
PRGs in normal tissue and EC tissue. We utilized these data to present a risk model containing 13 PRGs that can be employed to explore the prognosis of EC, with the risk score being an independent risk factor for EC prognosis. This study gives a basis for future studies on the role of pyroptosis in EC, and the risk score can be used to guide postoperative chemotherapy on the basis of the original staging.

Conclusions
To evaluate the prognostic value of PRGs in EC, this work utilized RNA-seq data from the esophagus in the TCGA database. DEGs between tumor and non-tumor samples were compared. A risk model for gene features was established using the LASSO regression method, and patients were classified into different risk groups based on the model results. Differences in survival rates and clinical characteristics among the groups were explored. The results showed that the risk score was an independent prognostic factor for predicting OS in EC patients. PRGs were enriched in the high-risk group. This suggested that PRGs play a crucial role in the occurrence and development of EC and can be used to predict the prognosis of EC patients. It was important to note that this work utilized data from the TCGA database and was limited to the information and sample size provided by the database. There may be biological limitations and missing information associated with this method. Therefore, further validation of these results is necessary in real clinical practice. Overall, the findings of this work provided some reference value for the prognosis of EC patients.

Conflict of Interest
The Authors declare that they have no conflict of interests.

Ethics Approval
This study is based on data analysis from The Cancer Genome Atlas (TCGA) database and does not involve research on human subjects, thus ethical approval is not applicable.

Data Availability
The data used in this study are derived from The Cancer Genome Atlas (TCGA) project, which can be freely accessed from the official website (https://tcga-data.nci.nih.gov/tcga/). As this study only utilized publicly available and published data from a public database, no special permissions or licenses are required. We are committed to open science and data transparency, and all the raw data, code, and results related to the analyses will be submitted to a publicly accessible data repository for peer review and future researchers’ use after publication.

Funding
None.

Authors’ Contributions
All the authors of this study made substantial and important contributions in the whole process of experimental design, data collection, data analysis, and manuscript writing.

ORCID ID
Yanjun Feng: 0009-0009-6823-6937.

References


37) Li XY, Zhang LY, Li XY, Yang XT, Su LX. A Pyroptosis-Related Gene Signature for Predicting Survival in Glioblastoma. Front Oncol 2021; 11: 697198.