

88612_Auto_Edited-check.docx

Identification and validation of a pyroptosis-related prognostic model for colorectal cancer based on bulk and single-cell RNA sequencing data

Zhu LH *et al.* Pyroptosis in colorectal cancer

Li-Hua Zhu, Jun Yang, Yun-Fei Zhang, Li Yan, Wan-Rong Lin, Wei-Qing Liu

Abstract

BACKGROUND

Pyroptosis impacts the development of malignant tumors, yet its role in colorectal cancer (CRC) prognosis remains uncertain.

AIM

To assess the prognostic significance of pyroptosis-related genes and their association with CRC immune infiltration.

METHODS

Gene expression data were obtained from The Cancer Genome Atlas (TCGA) and single-cell RNA sequencing dataset GSE178341 from the Gene Expression Omnibus (GEO). Pyroptosis-related gene expression in cell clusters was analyzed, and enrichment analysis was conducted. A pyroptosis-related risk model was developed using the LASSO regression algorithm, with prediction accuracy assessed through K-M and receiver operating characteristic analyses. A nomogram predicting survival was created, and the correlation between the risk model and immune infiltration was analyzed using CIBERSORTx calculations. Finally, the differential expression of the 8 prognostic genes between CRC and normal samples was verified by analyzing TCGA-COADREAD data from the UCSC database.

RESULTS

An effective pyroptosis-related risk model was constructed using 8 genes-*CHMP2B*, *SDHB*, *BST2*, *UBE2D2*, *GJA1*, *AIM2*, *PDCD6IP*, and *SEZ6L2* ($P < 0.05$). Seven of these genes exhibited differential expression between CRC and normal samples based on TCGA database analysis ($P < 0.05$). Patients with higher risk scores demonstrated increased death risk and reduced overall survival ($P < 0.05$). Significant differences in immune infiltration were observed between low- and high-risk groups, correlating with pyroptosis-related gene expression.

CONCLUSION

We developed a pyroptosis-related prognostic model for CRC, affirming its correlation with immune infiltration. This model may prove useful for CRC prognostic evaluation.

Key Words: Colorectal cancer; Pyroptosis; Single-cell RNA sequencing; Immune infiltration; Prognostic model

Zhu LH, Yang J, Zhang YF, Yan L, Lin WR, Liu WQ. Identification and validation of a pyroptosis-related prognostic model for colorectal cancer based on bulk and single-cell RNA sequencing data. *World J Clin Oncol* 2024; In press

Core Tip: We constructed a prognostic model related to colorectal cancer (CRC) focal death and confirmed the correlation with immune infiltration. This model may be useful for prognostic assessment of CRC.

¹INTRODUCTION

Colorectal cancer (CRC) is the prevalent malignancy and consider the second most common cause of cancer deaths globally^[1,2]. Genetic, lifestyle, obesity and environmental factors are considered as main causative agents of CRC^[3]. Besides, changes in the microenvironment of cells also proved to affect the growth development of this disease^[4-8]. The prognosis for CRC is grim, with nearly 20% of patients

progressing to stage 4 and an additional 20%-50% of early-stage patients developing metastatic disease^[9]. While immunotherapy introduces a promising avenue for CRC treatment, its effectiveness hinges on the intricacies of the immune microenvironment^[10-12]. Although numerous biomarkers identified through traditional methods based on bulk RNA sequencing, such as methylation^[13], lncRNA^[14], and IGFBP-2^[15], their accuracy in predicting CRC prognosis and the association with the tumor microenvironment (TME) is insufficient. Hence, there is an urgent need to develop a novel prognostic model with advanced technology for effective risk stratification and prediction of immunotherapy outcomes in CRC.

Pyroptosis, defined as gasdermin-mediated programmed cell death, has been established to influence tumor development^[16-19]. By modulating the immune microenvironment, **pyroptosis plays a crucial role in the prognosis of various cancers, including CRC**. Notably, pyroptosis-related genes like *IL-18*, *CASP1*, *GSDMB*, and *GASP5* have been utilized to construct prognostic models for bladder, ovarian, and gastric cancers^[20-22]. However, many of these models relied on bulk RNA sequencing levels, and adequate pyroptosis-related prognostic models specifically tailored for CRC are lacking.

To date, single-cell RNA sequencing (scRNA-seq) has emerged as the optimal method for discovering, identifying, and validating new biomarkers, particularly in TME research^[23]. This technique offers genomic and transcriptomic insights into cancers at the single-cell RNA level, surpassing the limitations of bulk RNA sequencing^[24-27]. Leveraging scRNA-seq, we developed a pyroptosis-related prognostic model for CRC and explored potential correlations between pyroptosis and immune infiltration. This study contributes valuable insights for clinical management and immunization research in CRC.

1 MATERIALS AND METHODS

Data source

The scRNA-seq dataset GSE178341, obtained from the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) database^[28], encompasses 61 CRC and 27 non-malignant colorectal tissues. This dataset, encompassing diverse clinical conditions, facilitates comprehensive analyses, offering profound insights into the involvement of pyroptosis-related genes in CRC. Additionally, its widespread use allows for robust comparisons and validation with other studies, augmenting the reliability of our research outcomes. The “The Cancer Genome Atlas (TCGA) biolinks” package of R software (version 2.22.4)^[19] was employed to retrieve TCGA-colon adenocarcinoma (TCGA-COAD) and TCGA-rectum adenocarcinoma (TCGA-READ) raw counts expression data and clinical data, comprising 578 CRC tumor samples and 106 paracancer samples. Merging these two expression matrices resulted in a baseline fact sheet with 619 cases containing clinical information (Table 1). Prognostic analyses were conducted on the samples that contained COAD and READ data.

Quality control of the data by Seurat

The R software (<https://www.r-project.org/>, version 4.1)⁷ and the R package Seurat (version 4.0.5)⁵^[29] were installed, and the expression matrix of the GSE178341 dataset was created as a Seurat object. Cells with > 20% mitochondrial genes, potentially indicating a stressful state, were excluded. Cells with FEATURE < 200 or > 3000 were also filtered, resulting in 115489 cells.

Subsequently,⁷ the sequencing depth of the dataset was normalized using the “NormalizeData” function with the default “LogNormalize” standardization method. The “FindVariableFeatures” function, employing the “vst” method, identified 2000 variable features of the dataset. Data scaling,⁵ utilizing the “ScaleData” function, mitigated the impact of sequencing depth. Principal Component Analysis (PCA) identified significant PCs^[30], and the Elbowplot function visualized the *P* value distribution. For the Uniform Manifold Approximation and Projection (UMAP) analysis, 30 PCs were selected. The Louvain algorithm, through the “FindClusters” function, optimized class groups, resulting in 38 different clusters with a resolution of

0.8. Finally, the “RunUMAP” function enabled dimensionality reduction for dataset visualization and exploration. The “FindAllMarkers” function compared ⁵ gene expression of cell clusters with the gene expression of all other cell clusters.

Cell types annotating

The Blueprint Encode Dateset in SingleR (version 1.8.1)^[31] was employed to annotate cell types in the single-cell data. Identified cell types included T cells, NK cells, B cells, plasma cells, epithelial cells, myeloid cells (DC, Macrophage, Monocyte), stromal cells, mast cells, and endothelial cells. Differential genes between cell types were identified using the “FindAllMarkers” function.

Pyroptosis-related differently expressed genes among cell clusters

A total of 427 pyroptosis-related genes were obtained from the Gene Cards database (<https://www.genecards.org/>)^[32] (Supplementary Table 1). The genes were intersected with marker genes of cell clusters for obtaining the pyroptosis-related differently expressed genes (DEGs) among cell clusters, A heat map illustrating the expression of DEGs in cell clusters was generated using the “DoHeatmap” function.

Correlation analysis of pyroptosis-related DEGs among cell subclusters

Pyroptosis-related DEGs expressing in specific cell types were visualized using the “FeaturePlot” from Seurat. The Pearson correlation coefficient of pyroptosis-related DEGs between cells was calculated using the corr R package. The correlation network was plotted using the “network_plot” function.

CellChat analysis

The CellChat R package (version 1.1.3)^[33] was used to quantitatively infer and analyze the communication network between the identified 11 cell clusters. A circle diagram depicted the interaction between cell groups, while a bubble diagram counted all important ligand pairs during intercellular signaling.

GSVA

The “c2.cp.kegg.v7.5.1.symbols.gmt” geneset was downloaded ¹ from the Molecular Signatures Database (MSigDB, <https://www.gsea-msigdb.org/gsea/msigdb/>)^[34]. The “gsva” method ² of the R package GSVA (version 1.42.0) was employed for analyzing CRC single-cell data. Gene expression data from an expression matrix with individual genes as features were transformed into an expression matrix with specific genesets as features. The expression matrix was transformed into an enrichment score (ES) matrix for the pathway, obtaining a GSVA ES for each cell corresponding to each pathway. Using the limma R package (version 3.50.0)^[35], pathways with significant differences ($P < 0.05$) were analyzed, and pathway activity scores for each cell group were compared with all other cell groups. The top 3 pathways in each group, ranked from the largest to the smallest t -value, were selected for plotting the heat map.

Immune infiltration

The TCGA-COAD and TCGA-READ transcriptome data underwent quantitative conversion into absolute abundance of immune and stromal cells using the “CIBERSORTx” method^[36,37]. This method assessed changes in the proportion of ¹ immune cell subsets, including memory B cells, naive B cells, activated dendritic cells, resting dendritic cells, eosinophils, M0 macrophages, M1 macrophages, M2 macrophages, activated mast cells, resting mast cells, monocytes, neutrophils, activated NK cells, resting NK cells, plasma cells, activated memory CD4⁺ T cells, resting memory CD4⁺ T cells, naive CD4⁺ T cells, CD8⁺ T cells, T follicular helper cells, gamma delta T cells, and regulatory T cells (Tregs). Significant differences between groups with ¹² high and low risk were assessed using the t-test method, considering P values < 0.05 as significant.

Weighted co-expression network analysis

Weighted co-expression network analysis (WGCNA) was employed to construct co-expression networks and identify modules of highly correlated genes^[38,39]. The COAD and READ datasets from TCGA were selected as the trait data for WGCNA.

Differential expression of pyroptosis-related genes in TCGA data

DESeq2 (version 1.34.0)^[40] from the R software package was used for analyzing the differential expression of pyroptosis-related genes in TCGA-CRC data. Pyroptosis-related DEGs were identified with a screening threshold of P value < 0.05 and $|\log FC| > 0.5$. Clustered heat maps, volcano maps, and Gene Ontology (GO) functional enrichment maps for the relevant genes were generated.

GO enrichment analysis

In GO enrichment analyses, each term in biological process (BP), molecular function (MF), and cellular component was analyzed for enrichment significance^[41]. This method was applied to characterize the features of pyroptosis-related genes.

TCGA tumor sample typing

Consensus Clusterin, a tool for cancer subtype classification, was used for analysis on the 178 key genes derived from scRNA-seq and TCGA datasets using the ConsensusClusterPlus (Version 1.58.0) package of R software^[42]. The distance calculation method was Spearman, and the clustering algorithm was PAM (Partitioning Around Medoids). Consistent cumulative distribution function maps, Delta Area Plots, and consistency matrix heat maps were utilized for clustering analysis.

Model construction and evaluation for clinical prognosis

Initially, the pyroptosis-related DEGs underwent univariate Cox analysis to identify genes with significant prognostic value. Subsequently, CRC samples were randomly divided into two parts, with a ratio of 7:3 for training and validation of the prognostic model. The LASSO-COX regression algorithm was applied to establish the prognostic

model, and the risk score calculation formula was defined as: Where “coef (k)” represents the multivariate Cox regression coefficient; “x (k)” represents the expression value of each single gene, and “n” represents the number of genes.

Evaluation of prognostic models

Initially, the 606 CRC samples were divided into two groups based on high and low risk scores using the median risk score. Subsequently, Kaplan-Meier survival analysis and time-dependent receiver operating characteristic (ROC) analysis were conducted to assess the prognostic accuracy for OS. The risk scores were compared under different clinical feature groups, including age, gender, and TNM stage.

Construction and evaluation of clinical prediction model

To illustrate the predictive ability of risk scores combined with clinicopathologic characteristics for patient prognosis, both were incorporated into the model. A clinical predictive nomogram was constructed to predict risk, and its predictive ability was evaluated using calibration curves by comparing predicted values with actual survival rates. OS of the predicted scores was analyzed using Kaplan-Meier, and the prognostic accuracy of the model was tested using time ROC analysis.

Tumor mutational burden

Tumor mutational burden (TMB), reflecting the quantity of cancer mutations^[43], was calculated using the “Maftools” package of R software (version 2.10.0). The somatic mutation levels in TCGA CRC samples were assessed, and the top 10 high-frequency mutated genes were counted to generate a waterfall plot. Subsequently, the impact on survival was explored by grouping according to high or low TMB levels, and comparisons were made between TMB differences in the two groups.

Differential expression of the prognostic genes

To validate the expression of the 8 prognostic genes between CRC and normal samples, we obtained the uniformly normalized pan-cancer dataset of TCGA TARGET GTEx (PANCAN, $n = 19131$, $G = 60499$) from the UCSC (<https://xenabrowser.net/>) database. Expression data for ENSG00000083937 (CHMP2B), ENSG00000117118 (SDHB), ENSG00000130303 (BST2), ENSG00000131508 (UBE2D2), ENSG00000152661 (GJA1), ENSG00000163568 (AIM2), ENSG00000170248 (PDCD6IP), and ENSG00000174938 (SEZ6L2) in samples from solid tissue normal, primary solid tumor, primary tumor, normal tissue, primary blood derived cancer-bone marrow, and primary blood derived cancer-peripheral blood were downloaded. A $\log_2(x + 0.001)$ transformation was applied to each expression value, and the analysis was restricted to CRC. Expression differences between normal and tumor samples were calculated for each tumor using R software. The significance of differences was assessed using unpaired Wilcoxon Rank Sum and Signed Rank Tests. The expression data of the 8 genes in CRC were provided in Table 2.

Statistical analysis

All calculations and analyses were performed using the R programming language. The risk model was constructed using LASSO and COX regression analyses.

RESULTS

ScRNA-seq data revealed cellular heterogeneity in CRC

The scRNA-seq data from 88 CRC samples underwent analysis, resulting in the identification of 115489 cells after adherence to quality control standards. Standardization and normalization of the data facilitated the extraction of the top 2000 high-variant genes. Subsequently, the selected high-variant genes underwent downscaled by PCA algorithm, followed by clustering analysis using the SNN algorithm. Visualization of the PCA-based downscaling results was achieved through UMAP for single-cell clustering. The successful classification of 115489 cells into 38 independent clusters is depicted in Figure 1A, and differential marker genes for each

cluster are outlined in Table 3. Using Single R, 11 distinct cell subsets (epithelial cell, myeloid cells, macrophage, monocyte, mast cells, endothelial cells, stromal cell, plasma, B cells, NK cells, and T cells) were identified (Figure 1B). Subsequently, expression patterns of selected datasets corresponding to markers in published articles were visually represented through bubble plots (Figure 1C). Violin plots illustrated differentially marked genes for each cell subset (Figure 1D), and heat maps displayed the top 2 differentially marked genes for each cell type (Figure 1E). The distribution of cells in CRC tissues (T) and non-CRC tissues (N) across each cell type is illustrated in Figure 1F. Notably, T cells are more predominant in CRC tissues, while plasma/B cells are more prevalent in non-CRC tissues. A comparative UMAP plot in tumor and normal samples is presented in Supplementary Figure 1.

Pyroptosis-related genes differentially expressed between cell subsets

We intersected the differential genes between cell types and pyroptosis-related genes, resulting in 125 pyroptosis-related DEGs (Supplementary Table 2). Subsequently, we utilized a heat map to depict the expression of these DEGs in each of the 11 cell subsets (Figure 2A). Notably, among the DEGs, *GZMA* was specifically expressed in the cluster where NK cells and T cells are located (Figure 2B), while *IL-1B* was found to be specifically expressed in Monocytes (Figure 2C). The correlation among intersecting pyroptosis-related genes was visualized in Figure 2D. Notably, genes such as *APOE*, *VIM*, and *STAT3* exhibited a high degree of correlation.

CellChat and GSVA

We utilized CellChat to construct a graph displaying the total number of interactions among 11 cell subsets and their overall interaction intensity (Figure 3A). The statistical plot depicting cellular interactions identified by CellChat is presented in Supplementary Figure 2. For a clearer examination of interactions among cell subsets, we conducted subset analysis (Figure 3B), resulting in the division of subsets into 27 subclusters (Supplementary Table 3) [B cells: B01, 2138, 1.85%; B02, 1101, 0.95%; B03,

10239, 8.87%; B04, 42, 0.04%. Plasma 1, 8620, Plasma 2, 1491, 1.29%; Plasmablasts, 168, 0.15%. Dendritic cells (DC): DC1, 568, 0.49%; DC2, 891, 0.77%; pDC, 334, 0.29%. Endothelial cells (Endo): Endo, 2636, 2.28%. Epithelial cells: Epithelial Normal (EpiN), 10480, 9.07%. Epithelial tumor (EpiT): 21614, 18.72%. Fibroblasts: FB1, 1329, 1.15%; FB2, 855, 0.74%. Macrophages: M01, 4020, 3.48%; M02, 1593, 1.38%. Mast, 1286, 1.11%; Monocytes, 4804, 4.16%; Mural, 1162, 1.01%; NK, 6448, 5.58%; Schwann, 159, 0.14%; T cells: T01, 12099, 10.48%; T02, 5772, 5.00%; T03, 7925, 6.86%; T04, 6071, 5.26%; T05, 1644, 1.42%]. Based on the 27 cell subclusters, a CellChat heatmap analysis was performed (Figure 3C), identifying 5 key cell subclusters (M01 Macrophage, T05 T cell, FB2 Fibroblast, Plasmablasts, Schwann) for subsequent analysis. The SPP1 signaling pathway influences the effectiveness of immunotherapy in CRC^[44]. We individually aligned for CellChat visualization (Figure 3D). It is evident that the M01 subcluster of macrophages was highly active in the SPP1 signaling pathway. Interestingly, the M02 subcluster of macrophages exhibited minimal activity. Additionally, Schwann cells demonstrated significant activity in the SPP1 signaling pathway. Furthermore, $\gamma\delta$ T cells (T05) were correlated with the initiation and progression of immune responses. We analyzed the interaction of T05 Ligands and receptors with other cells (Figure 3E and F). Subsequently, CellChat analysis of the 5 key cell clusters mentioned above was performed with tumor cell clusters (Figure 3G). The close interlinking of the key subclusters was observed. We also noted the enrichment of different metabolic pathways among the cell clusters. Gamma-delta T-cells (T05) were enriched in the cell cycle, DNA replication, and base excision repair pathways, aligning with their function in initiating immune responses. Notably, SPP1-macrophage (M01) was enriched in the toll-like receptor signaling pathway and cytokine-cytokine receptor interaction, while M02 was enriched in the pathways of retinol metabolism and linoleic acid metabolism (Figure 3H).

Immune infiltration analysis

We derived the abundance values of immune cells by utilizing the CIBERSORT online tool to analyze TCGA-COAD and TCGA-READ data. The boxplot visually presents the percentage differences in predicted results among various cell subsets (Figure 4A). Notably, immune cells such as M0 Macrophages, M2 Macrophages, and naïve B cells exhibited significant percentage differences. Subsequently, we eliminated immune cells with 0 abundance in more than half of the samples and constructed a Pearson correlation heatmap depicting relationships among 14 immune cell types (Figure 4B). The correlation analysis revealed strong associations between T cell subtypes, monocytes, and macrophage subtypes. For instance, negative correlations were observed between CD8+ T cells and M0 macrophages ($R = -0.4$), CD8+ T cells and resting memory CD4+ T cells ($R = -0.38$), Monocytes and M0 macrophages ($R = -0.37$), as well as resting memory CD4+ T cells and M0 macrophages ($R = -0.28$). Conversely, positive correlations were identified between CD8+ T cells and M1 macrophages ($r = 0.21$) and resting memory CD4+ T cells and Monocytes ($r = 0.23$).

WGCNA

The soft threshold value β was determined to be 16 (Figure 5A and B). Subsequently, we identified 16 modules for further analysis. Hierarchical cluster plots and module correlation heatmaps were generated to visualize the modules (Figure 5C and D). Notably, a significant correlation was observed between the MEcyan module and the M1 Macrophages feature, the MEpurple module and Monocytes and M2 Macrophages features, the MERed module and activated Mast cells, and the Megreen module with M0 Macrophages (Figure 6). From each module, we selected the top 30 genes, resulting in the identification of 120 genes forming the co-expressed gene list (Supplementary Table 4).

Differential expression, correlation analysis and enrichment analysis of pyroptosis-related genes

The non-CRC samples from TCGA served as the control group, while the CRC samples were designated as the experimental group for differential analysis (Supplementary Table 5). Among the 125 DEGs, 71 core pyroptosis-related DEGs exhibited significant differential expression in the TCGA dataset ($|\log_2 \text{FC}| > 0.5$, $P < 0.05$) (Table 4). The top 14 genes were selected for heatmap display (Figure 7A), and the differential analysis volcano plot provided a visual representation of pyroptosis-related DEGs (Figure 7B), including genes such as *CDKN2B-AS1*, *CTSG*, *MPEG1*, *GZMB*, and *DPEP1*, etc., between normal and tumor samples. The Pyroptosis-Related Genes PPI network diagram is presented (Figure 7C, Supplementary Figure 3), with the gene *CXCL8* also identified in the differential analysis. Additionally, GO enrichment analysis revealed that the 71 core pyroptosis-related genes were significantly enriched in functions such as the regulation of inflammatory response, mitotic cytokinesis, etc. (Figure 7D) (Table 5).

Consistent clustering and single-sample gene set enrichment analysis

Tumor samples TCGA were subjected to typing through the consensus clustering method. After a thorough evaluation considering the consistency matrix heatmap, cumulative distribution curve, and delta area curve, we determined the cluster number to be 2 (Figure 8A-C). Subsequently, the tSNE algorithm was employed for cluster visualization (Figure 8D). Finally, we conducted single-sample gene set enrichment analysis (ssGSEA) with a focus on immune cells (Figure 8E) and immune function (Figure 8F).

Construction and validation of a pyroptosis-related prognostic model

We conducted survival analysis utilizing both single-cell data and TCGA-CRC data, which comprises survival information for 606 samples. In addition to the 71 core genes, we incorporated differential genes specific to single-cell and bulk transcriptomes, resulting in a final set of 178 genes present in the TCGA expression matrix (Supplementary Table 6). Univariate Cox analysis assessed the correlation between

these 178 genes and the prognosis of CRC patients, revealing 10 genes significantly correlated with prognosis (P value < 0.05) (Supplementary Table 7). Subsequently, we randomly divided the diseased samples into training and validation sets at a ratio of 7:3. The training set was employed for constructing a prognostic model using LASSO-Cox regression (Figure 9A and B), yielding a risk model composed of 8 genes (*CHMP2B*, *SDHB*, *BST2*, *BE2D2*, *GJA1*, *AIM2*, *PDCD6IP*, and *SEZ6L2*) (Table 6). Based on the median value of the risk score, patients were classified into low-risk and high-risk groups. Risk maps and survival states for the training and test sets illustrated an increase in the risk score corresponding to an elevated risk of death and decreased survival time (Figure 9C and D).

To assess prediction accuracy, we performed a ROC analysis. The results indicated a favorable predictive ability of the risk score for OS in CRC patients, with area under curve (AUC) values of 63.8% and 63.6% for the training and validation sets, respectively (Figure 9E and F). Kaplan-Meier curves demonstrated worse OS in patients with high-risk scores compared to those with low-risk scores ($P < 0.05$, Figure 9G and H). The 1-, 3-, and 5-year AUCs of risk scores based on the prognostic models were all above 0.6 (Figure 9I and J).

Construction of a prediction nomogram

The forest plot (Figure 10A) highlighted strong correlations with clinicopathological features, particularly tumor stage. Additionally, by leveraging clinical data within the dataset, we observed a correlation between pyroptosis-related genes and the age of tumors, with no significant association with gender. Subsequently, by integrating clinicopathological characteristics, a nomogram (Figure 10B) was developed to predict survival probability. The calibration curve indicated accurate results (Figure 10C).

Immune infiltration and the prognostic model

Given the significantly lower survival rate in the high-risk group based on the previous results, we explored potential differences in immune infiltration between the two risk

groups. The CIBERSORTx algorithm was employed to calculate immune infiltration in CRC samples from TCGA. The scatterplot depicted correlations between the expression of prognostic genes and immune infiltration in CRC. *AIM2* expression showed a positive correlation with the cellular abundance of activated memory CD4⁺ T cells, while *UBE2D2* expression exhibited a negative correlation with the cellular abundance of Tregs (Figure 11A and B). The risk score demonstrated a negative correlation with the cellular abundance of resting memory CD4⁺ T cells (Figure 11C). Additionally, a significant difference in the abundance of Tregs was observed between the high and low-risk groups ($P < 0.05$) (Figure 11D).

TMB

TMB serves as a predictor of immunotherapy response, and we calculated TMB using “maf” files, investigating the relationship between the model groupings and TMB. Waterfall charts were generated for the top 10 frequently mutated genes, revealing common somatic mutation genes such as *APC* and *TP53* in CRC (Figure 12A). TMB was calculated and visualized, with a median TMB of 1.78/Mb (Figure 12B). Subsequently, we explored the impact of TMB on survival (Figure 12C), revealing that TMB had minimal effect on survival in this dataset. Furthermore, no significant difference was observed in TMB between the high- and low-risk groups in the prognostic model (Figure 12D), suggesting that incorporating TMB into the prognostic model for this dataset may not be necessary. Additional studies on the effect of TMB on prognosis may be warranted.

Differential expression of prognostic-related genes in CRC samples

We determined the expression of 8 prognostic genes in CRC samples using TCGA-COADREAD data in the UCSC database. Among them, 7 genes were shown differentially expressed in CRC samples and normal samples. The result showed that *CHMP2B*, *SDHB*, *UBE2D2*, *AIM2*, *PDCD6IP*, and *SEZ6L2* were significantly up-regulated in CRC samples while *GJA1* was significantly down-regulated. No significant

expression difference was found between normal and tumor samples for *BST2* (Figure 13).

DISCUSSION

Several studies have shown that pyroptosis plays a crucial role in tumor growth^[17-19,45]. It affects prognosis by changing the immune microenvironment and is linked to the effectiveness of immunotherapy^[46]. Consequently, it has been utilized in building prognostic models for various cancers^[21,47,48]. However, most of these studies utilized bulk RNA sequencing, whereas scRNA-seq is more advantageous for investigating cancer prognostic models and the immune microenvironment at a single-cell resolution level^[49-51]. Recognizing the pivotal role of pyroptosis in cancers and the unfavorable prognosis of CRC, we developed a pyroptosis-related prognostic model for CRC using the scRNA-seq method. Notably, this is the initial study applying scRNA-seq technology to identify pyroptosis-related genes for constructing CRC prognostic prediction models and exploring the relationship between pyroptosis-related genes and immune infiltration.

By integrating single-cell transcriptome and bulk transcriptome data, we identified 178 pyroptosis-related DEGs from CRC samples. Subsequently, utilizing univariate COX analysis and the LASSO-Cox regression algorithm, we established a risk model comprising 8 pyroptosis-related genes: *CHMP2B*, *SDHB*, *BST2*, *UBE2D2*, *GJA1*, *AIM2*, *PDCD6IP*, and *SEZ6L2*. The model was then validated. ¹³ Based on the median risk scores, patients from the TCGA cohort were stratified into high- and low-risk groups, revealing an elevated risk of death and reduced OS among high-risk group patients. The model exhibited high predictive accuracy for CRC survival, as confirmed by ROC analysis and a nomogram, while also demonstrating a strong correlation with clinicopathological characteristics, especially tumor stage.

Among the 8 genes, *SDHB* serves as the catalytic core component of succinate dehydrogenase (SDH), a mitochondrial metabolic enzyme^[52]. Mutations in *SDHB* result in enzyme dysfunction associated with cancer development^[52-54]. Wang *et al*^[52] observed

that SDHB influences CRC invasion and metastasis through the TGF- β pathway. BST2 (bone marrow stromal antigen 2) is a protein-coding gene overexpressed in several cancers^[55]. Chiang *et al*^[56] identified *BST2* as a biomarker and prognosticator for CRC^[56]. UBE2D2 (ubiquitin-conjugating enzymes E2), associated with hypoxia, prevents the degradation of HIF1 α and 2 α by proteasome systems. Lee *et al*^[57] reported that *UBE2D2* could predict the OS of CRC. GJA1 (gap junction alpha-1), a member of the GJ family, is the predominant one expressed in epithelial tissues. Hu *et al*^[58] demonstrated that *GJA1* serves as a prognostic biomarker for CRC. AIM2, an inflammasome sensor, provides cytokine-independent protection, influencing CRC^[59]. SEZ6L2 (seizure-related 6 homolog/mouse-like 2) of the SEZ6 family is identified as a potential prognosis biomarker and therapy target for CRC^[60]. The six pyroptosis-related genes above have demonstrated potential impacts on CRC prognosis, aligning with our findings. Notably, the existing prognostic models for these genes relied on bulk RNA sequencing and focused solely on whole tumor cells. In our study, risk models were validated at both bulk RNA and single-cell levels. Regarding *CHMP2B* and *PDCD6IP*, their roles in pyroptosis have not been fully explored. We are the first to identify these two genes as potential prognostic biomarkers for CRC.

Immune cells within the TME play a pivotal role in influencing the tumor process^[61]. Pyroptosis has been demonstrated to actively participate in regulating the immune microenvironment in various tumors^[62,63]. This study specifically investigated the regulatory function of pyroptosis-related genes on immune infiltration. CD4⁺ T cells are key participants in anti-tumor immune responses and significantly impact CRC prognosis^[64-67]. Previous studies have highlighted that activated memory CD4⁺ T cells exhibit infiltrative and antitumor effects during the early stages of CRC progression^[68], while infiltration of memory resting CD4⁺ T cells is associated with a favorable prognosis^[69]. The metabolic features and function of intra-tumoral Tregs in CRC remain unclear. To address this, we employed the CIBERSORTx algorithm to analyze immune infiltration results in the high- and low-risk groups. The findings revealed differences in immune infiltration between these groups, correlating with prognostic genes. *AIM2*

expression positively correlated with activated memory CD4⁺ T cell abundance, while *UBE2D2* expression negatively correlated with Tregs cell abundance. This suggests that pyroptosis-related genes may impact prognosis by influencing immune infiltration in CRC. This study is potentially the first to establish a connection between pyroptosis-related genes and immune infiltration in CRC, offering insights that may contribute to advancements in immunotherapy.

We explored the relationship between the expression of pyroptosis-related genes and clinical data in CRC using the TCGA dataset. Our findings revealed an association between the expression of these genes and patient age and tumor stage, while no correlation was observed with gender.

Furthermore, we delved into the functional roles of pyroptosis in CRC. Functional analyses indicated significant enrichment of pyroptosis-related genes in the regulation of inflammatory responses. Notably, key intermediate factors such as *GSDMD*, *IL-1 β* , and *IL-18*, known for their involvement in the pyroptosis process, were identified as contributors to the regulation of inflammatory responses^[70]. This underscores our study's demonstration of the regulatory role of pyroptosis in inflammatory responses, thereby impacting tumor progression.

Finally, we validated the differential expression of the eight prognostic genes in CRC and normal samples using TCGA-COADREAD data from the UCSC database. Out of these, seven genes—*CHMP2B*, *SDHB*, *UBE2D2*, *GJA1*, *AIM2*, *PDCD6IP*, and *SEZ6L2*—showed significant differential expression, with six genes up-regulated and one gene down-regulated. This outcome suggests that these seven pyroptosis-related genes could be potential targets for clinical treatment in CRC. However, further data and validation from clinical trials are required. To advance research in this area, additional experiments involving patients, as well as *in vitro* and *in vivo* studies, are currently underway in our laboratory.

CONCLUSION

Leveraging scRNA-seq analysis, we formulated a pyroptosis-related prognostic model for CRC. This model demonstrates efficacy in predicting prognosis, survival OS, and effectively stratifying risk. The eight pyroptosis-related genes comprising the risk score play crucial roles in regulating inflammatory responses, modulating immune infiltration, and influencing the onset and progression of CRC. The insights derived from this study hold promise for enhancing clinical management and immune therapy strategies for CRC patients.

ARTICLE HIGHLIGHTS

Research background

Pyroptosis impacts the development of malignant tumors, yet its role in colorectal cancer (CRC) prognosis remains uncertain.

Research motivation

To explore the role of pyroptosis in CRC prognosis.

Research objectives

To assess the prognostic significance of pyroptosis-related genes and their association with CRC immune infiltration.

Research methods

Single-cell sequencing combined with Gene Expression Omnibus database and The Cancer Genome Atlas database.

Research results

We constructed a prognostic model and demonstration that pyroptosis is associated with immune infiltration in CRC.

Research conclusions

We developed a pyroptosis-related prognostic model for CRC, affirming its correlation with immune infiltration.

Research perspectives

This model may prove useful for CRC prognostic evaluation.

6

ACKNOWLEDGEMENTS

This study is a joint effort of many investigators and staff members, and their contribution is gratefully acknowledged. We especially thank all patients who participated in this study.

REFERENCES

- 1 **Ciardiello F**, Ciardiello D, Martini G, Napolitano S, Tabernero J, Cervantes A. Clinical management of metastatic colorectal cancer in the era of precision medicine. *CA Cancer J Clin* 2022; **72**: 372-401 [PMID: 35472088 DOI: 10.3322/caac.21728]
- 2 **Ye C**, Deng Y, Sun W, Yuan Y. Editorial: Recent advances of drug resistance research in colorectal cancer therapy. *Front Oncol* 2022; **12**: 1059223 [PMID: 36483046 DOI: 10.3389/fonc.2022.1059223]
- 3 **Arthur JC**, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan TJ, Campbell BJ, Abujamel T, Dogan B, Rogers AB, Rhodes JM, Stintzi A, Simpson KW, Hansen JJ, Keku TO, Fodor AA, Jobin C. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* 2012; **338**: 120-123 [PMID: 22903521 DOI: 10.1126/science.1224820]
- 4 **Khan S**. Potential role of Escherichia coli DNA mismatch repair proteins in colon cancer. *Crit Rev Oncol Hematol* 2015; **96**: 475-482 [PMID: 26014615 DOI: 10.1016/j.critrevonc.2015.05.002]
- 5 **Khan S**, Imran A, Khan AA, Abul Kalam M, Alshamsan A. Systems Biology Approaches for the Prediction of Possible Role of Chlamydia pneumoniae Proteins in

the Etiology of Lung Cancer. *PLoS One* 2016; **11**: e0148530 [PMID: 26871581 DOI: 10.1371/journal.pone.0148530]

6 **Khan S**, Zaidi S, Alouffi AS, Hassan I, Imran A, Khan RA. Computational Proteome-Wide Study for the Prediction of Escherichia coli Protein Targeting in Host Cell Organelles and Their Implication in Development of Colon Cancer. *ACS Omega* 2020; **5**: 7254-7261 [PMID: 32280866 DOI: 10.1021/acsomega.9b04042]

7 **Khan S**, Zakariah M, Palaniappan S. Computational prediction of Mycoplasma hominis proteins targeting in nucleus of host cell and their implication in prostate cancer etiology. *Tumour Biol* 2016; **37**: 10805-10813 [PMID: 26874727 DOI: 10.1007/s13277-016-4970-9]

8 **Khan S**, Zakariah M, Rolfo C, Robrecht L, Palaniappan S. Prediction of mycoplasma hominis proteins targeting in mitochondria and cytoplasm of host cells and their implication in prostate cancer etiology. *Oncotarget* 2017; **8**: 30830-30843 [PMID: 27027344 DOI: 10.18632/oncotarget.8306]

9 **Siegel R**, Desantis C, Jemal A. Colorectal cancer statistics, 2014. *CA Cancer J Clin* 2014; **64**: 104-117 [PMID: 24639052 DOI: 10.3322/caac.21220]

10 **Ganesh K**, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, Diaz LA Jr. Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nat Rev Gastroenterol Hepatol* 2019; **16**: 361-375 [PMID: 30886395 DOI: 10.1038/s41575-019-0126-x]

11 **Sahu A**, Kose K, Kraehenbuehl L, Byers C, Holland A, Tembo T, Santella A, Alfonso A, Li M, Cordova M, Gill M, Fox C, Gonzalez S, Kumar P, Wang AW, Kurtansky N, Chandrani P, Yin S, Mehta P, Navarrete-Dechent C, Peterson G, King K, Dusza S, Yang N, Liu S, Phillips W, Guitera P, Rossi A, Halpern A, Deng L, Pulitzer M, Marghoob A, Chen CJ, Merghoub T, Rajadhyaksha M. In vivo tumor immune microenvironment phenotypes correlate with inflammation and vasculature to predict immunotherapy response. *Nat Commun* 2022; **13**: 5312 [PMID: 36085288 DOI: 10.1038/s41467-022-32738-7]

- 12 **Xu J**, Dai S, Jiang K, Xiao Q, Yuan Y, Ding K. Combining single-cell sequencing to identify key immune genes and construct the prognostic evaluation model for colon cancer patients. *Clin Transl Med* 2021; **11**: e465 [PMID: 34323426 DOI: 10.1002/ctm2.465]
- 13 **Tepus M**, Yau TO. Non-Invasive Colorectal Cancer Screening: An Overview. *Gastrointest Tumors* 2020; **7**: 62-73 [PMID: 32903904 DOI: 10.1159/000507701]
- 14 **Xu MD**, Qi P, Du X. Long non-coding RNAs in colorectal cancer: implications for pathogenesis and clinical application. *Mod Pathol* 2014; **27**: 1310-1320 [PMID: 24603586 DOI: 10.1038/modpathol.2014.33]
- 15 **Diehl D**, Hessel E, Oesterle D, Renner-Müller I, Elmlinger M, Langhammer M, Göttlicher M, Wolf E, Lahm H, Hoefflich A. IGFBP-2 overexpression reduces the appearance of dysplastic aberrant crypt foci and inhibits growth of adenomas in chemically induced colorectal carcinogenesis. *Int J Cancer* 2009; **124**: 2220-2225 [PMID: 19142966 DOI: 10.1002/ijc.24193]
- 16 **Yu P**, Zhang X, Liu N, Tang L, Peng C, Chen X. Pyroptosis: mechanisms and diseases. *Signal Transduct Target Ther* 2021; **6**: 128 [PMID: 33776057 DOI: 10.1038/s41392-021-00507-5]
- 17 **Zhou B**, Zhang JY, Liu XS, Chen HZ, Ai YL, Cheng K, Sun RY, Zhou D, Han J, Wu Q. Tom20 senses iron-activated ROS signaling to promote melanoma cell pyroptosis. *Cell Res* 2018; **28**: 1171-1185 [PMID: 30287942 DOI: 10.1038/s41422-018-0090-y]
- 18 **Hu J**, Pei W, Jiang M, Huang Y, Dong F, Jiang Z, Xu Y, Li Z. DFNA5 regulates immune cells infiltration and exhaustion. *Cancer Cell Int* 2022; **22**: 107 [PMID: 35248047 DOI: 10.1186/s12935-022-02487-0]
- 19 **Molina-Crespo Á**, Cadete A, Sarrio D, Gámez-Chiachio M, Martinez L, Chao K, Olivera A, Gonella A, Díaz E, Palacios J, Dhal PK, Besev M, Rodríguez-Serrano M, García Bermejo ML, Triviño JC, Cano A, García-Fuentes M, Herzberg O, Torres D, Alonso MJ, Moreno-Bueno G. Intracellular Delivery of an Antibody Targeting Gasdermin-B Reduces HER2 Breast Cancer Aggressiveness. *Clin Cancer Res* 2019; **25**: 4846-4858 [PMID: 31064780 DOI: 10.1158/1078-0432.CCR-18-2381]

- 20 **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 [PMID: 33828074 DOI: 10.1038/s41420-021-00451-x]
- 21 **Shao W**, Yang Z, Fu Y, Zheng L, Liu F, Chai L, Jia J. The Pyroptosis-Related Signature Predicts Prognosis and Indicates Immune Microenvironment Infiltration in Gastric Cancer. *Front Cell Dev Biol* 2021; **9**: 676485 [PMID: 34179006 DOI: 10.3389/fcell.2021.676485]
- 22 **Zhang Q**, Tan Y, Zhang J, Shi Y, Qi J, Zou D, Ci W. Pyroptosis-Related Signature Predicts Prognosis and Immunotherapy Efficacy in Muscle-Invasive Bladder Cancer. *Front Immunol* 2022; **13**: 782982 [PMID: 35479097 DOI: 10.3389/fimmu.2022.782982]
- 23 **Lei Y**, Tang R, Xu J, Wang W, Zhang B, Liu J, Yu X, Shi S. Applications of single-cell sequencing in cancer research: progress and perspectives. *J Hematol Oncol* 2021; **14**: 91 [PMID: 34108022 DOI: 10.1186/s13045-021-01105-2]
- 24 **Abrams D**, Kumar P, Karuturi RKM, George J. A computational method to aid the design and analysis of single cell RNA-seq experiments for cell type identification. *BMC Bioinformatics* 2019; **20**: 275 [PMID: 31167661 DOI: 10.1186/s12859-019-2817-2]
- 25 **Bartoschek M**, Oskolkov N, Bocci M, Lötvot J, Larsson C, Sommarin M, Madsen CD, Lindgren D, Pekar G, Karlsson G, Ringnér M, Bergh J, Björklund Å, Pietras K. Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. *Nat Commun* 2018; **9**: 5150 [PMID: 30514914 DOI: 10.1038/s41467-018-07582-3]
- 26 **Kim C**, Gao R, Sei E, Brandt R, Hartman J, Hatschek T, Crosetto N, Foukakis T, Navin NE. Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell* 2018; **173**: 879-893.e13 [PMID: 29681456 DOI: 10.1016/j.cell.2018.03.041]
- 27 **Zhang L**, Li Z, Skrzypczynska KM, Fang Q, Zhang W, O'Brien SA, He Y, Wang L, Zhang Q, Kim A, Gao R, Orf J, Wang T, Sawant D, Kang J, Bhatt D, Lu D, Li CM, Rapaport AS, Perez K, Ye Y, Wang S, Hu X, Ren X, Ouyang W, Shen Z, Egen JG, Zhang Z, Yu X. Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in

Colon Cancer. *Cell* 2020; **181**: 442-459.e29 [PMID: 32302573 DOI: 10.1016/j.cell.2020.03.048]

28 **Pelka K**, Hofree M, Chen JH, Sarkizova S, Pirl JD, Jorgji V, Bejnood A, Dionne D, Ge WH, Xu KH, Chao SX, Zollinger DR, Lieb DJ, Reeves JW, Fuhrman CA, Hoang ML, Delorey T, Nguyen LT, Waldman J, Klapholz M, Wakiro I, Cohen O, Albers J, Smillie CS, Cuoco MS, Wu J, Su MJ, Yeung J, Vijaykumar B, Magnuson AM, Asinovski N, Moll T, Goder-Reiser MN, Applebaum AS, Brais LK, DelloStritto LK, Denning SL, Phillips ST, Hill EK, Meehan JK, Frederick DT, Sharova T, Kanodia A, Todres EZ, Jané-Valbuena J, Biton M, Izar B, Lambden CD, Clancy TE, Bleday R, Melnitchouk N, Irani J, Kunitake H, Berger DL, Srivastava A, Hornick JL, Ogino S, Rotem A, Vigneau S, Johnson BE, Corcoran RB, Sharpe AH, Kuchroo VK, Ng K, Giannakis M, Nieman LT, Boland GM, Aguirre AJ, Anderson AC, Rozenblatt-Rosen O, Regev A, Hacohen N. Spatially organized multicellular immune hubs in human colorectal cancer. *Cell* 2021; **184**: 4734-4752.e20 [PMID: 34450029 DOI: 10.1016/j.cell.2021.08.003]

29 **Satija R**, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 2015; **33**: 495-502 [PMID: 25867923 DOI: 10.1038/nbt.3192]

30 **Kim S**, Kang D, Huo Z, Park Y, Tseng GC. Meta-analytic principal component analysis in integrative omics application. *Bioinformatics* 2018; **34**: 1321-1328 [PMID: 29186328 DOI: 10.1093/bioinformatics/btx765]

31 **Aran D**, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, Butte AJ, Bhattacharya M. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* 2019; **20**: 163-172 [PMID: 30643263 DOI: 10.1038/s41590-018-0276-y]

32 **Safran M**, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, Nativ N, Bahir I, Doniger T, Krug H, Sirota-Madi A, Olender T, Golan Y, Stelzer G, Harel A, Lancet D. GeneCards Version 3: the human gene integrator. *Database (Oxford)* 2010; **2010**: baq020 [PMID: 20689021 DOI: 10.1093/database/baq020]

- 33 **Jin S**, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, Myung P, Plikus MV, Nie Q. Inference and analysis of cell-cell communication using CellChat. *Nat Commun* 2021; **12**: 1088 [PMID: 33597522 DOI: 10.1038/s41467-021-21246-9]
- 34 **Liberzon A**, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 2015; **1**: 417-425 [PMID: 26771021 DOI: 10.1016/j.cels.2015.12.004]
- 35 **Ritchie ME**, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; **43**: e47 [PMID: 25605792 DOI: 10.1093/nar/gkv007]
- 36 **Chen B**, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol* 2018; **1711**: 243-259 [PMID: 29344893 DOI: 10.1007/978-1-4939-7493-1_12]
- 37 **Steen CB**, Liu CL, Alizadeh AA, Newman AM. Profiling Cell Type Abundance and Expression in Bulk Tissues with CIBERSORTx. *Methods Mol Biol* 2020; **2117**: 135-157 [PMID: 31960376 DOI: 10.1007/978-1-0716-0301-7_7]
- 38 **Langfelder P**, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; **9**: 559 [PMID: 19114008 DOI: 10.1186/1471-2105-9-559]
- 39 **Yin Y**, Wu S, Niu L, Huang S. A ZFP42/MARK2 regulatory network reduces the damage of retinal ganglion cells in glaucoma: a study based on GEO dataset and in vitro experiments. *Apoptosis* 2022; **27**: 1049-1059 [PMID: 36131186 DOI: 10.1007/s10495-022-01746-9]
- 40 **Love MI**, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; **15**: 550 [PMID: 25516281 DOI: 10.1186/s13059-014-0550-8]
- 41 **The Gene Ontology Consortium**. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* 2019; **47**: D330-D338 [PMID: 30395331 DOI: 10.1093/nar/gky1055]

- 42 **Wilkerson MD**, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010; **26**: 1572-1573 [PMID: 20427518 DOI: 10.1093/bioinformatics/btq170]
- 43 **Jardim DL**, Goodman A, de Melo Gagliato D, Kurzrock R. The Challenges of Tumor Mutational Burden as an Immunotherapy Biomarker. *Cancer Cell* 2021; **39**: 154-173 [PMID: 33125859 DOI: 10.1016/j.ccell.2020.10.001]
- 44 **Qi J**, Sun H, Zhang Y, Wang Z, Xun Z, Li Z, Ding X, Bao R, Hong L, Jia W, Fang F, Liu H, Chen L, Zhong J, Zou D, Liu L, Han L, Ginhoux F, Liu Y, Ye Y, Su B. Single-cell and spatial analysis reveal interaction of FAP(+) fibroblasts and SPP1(+) macrophages in colorectal cancer. *Nat Commun* 2022; **13**: 1742 [PMID: 35365629 DOI: 10.1038/s41467-022-29366-6]
- 45 **Cheng Z**, Han J, Jiang F, Chen W, Ma X. Prognostic pyroptosis-related lncRNA signature predicts the efficacy of immunotherapy in hepatocellular carcinoma. *Biochem Biophys Rep* 2022; **32**: 101389 [PMID: 36438599 DOI: 10.1016/j.bbrep.2022.101389]
- 46 **Li L**, Jiang M, Qi L, Wu Y, Song D, Gan J, Li Y, Bai Y. Pyroptosis, a new bridge to tumor immunity. *Cancer Sci* 2021; **112**: 3979-3994 [PMID: 34252266 DOI: 10.1111/cas.15059]
- 47 **Song J**, Sun Y, Cao H, Liu Z, Xi L, Dong C, Yang R, Shi Y. A novel pyroptosis-related lncRNA signature for prognostic prediction in patients with lung adenocarcinoma. *Bioengineered* 2021; **12**: 5932-5949 [PMID: 34488540 DOI: 10.1080/21655979.2021.1972078]
- 48 **Fu XW**, Song CQ. Identification and Validation of Pyroptosis-Related Gene Signature to Predict Prognosis and Reveal Immune Infiltration in Hepatocellular Carcinoma. *Front Cell Dev Biol* 2021; **9**: 748039 [PMID: 34820376 DOI: 10.3389/fcell.2021.748039]
- 49 **Wang Z**, Guo X, Gao L, Wang Y, Ma W, Xing B. Glioblastoma cell differentiation trajectory predicts the immunotherapy response and overall survival of patients. *Aging (Albany NY)* 2020; **12**: 18297-18321 [PMID: 32957084 DOI: 10.18632/aging.103695]
- 50 **Zheng H**, Liu H, Ge Y, Wang X. Integrated single-cell and bulk RNA sequencing analysis identifies a cancer associated fibroblast-related signature for predicting

prognosis and therapeutic responses in colorectal cancer. *Cancer Cell Int* 2021; **21**: 552 [PMID: 34670584 DOI: 10.1186/s12935-021-02252-9]

51 **Lim B**, Lin Y, Navin N. Advancing Cancer Research and Medicine with Single-Cell Genomics. *Cancer Cell* 2020; **37**: 456-470 [PMID: 32289270 DOI: 10.1016/j.ccell.2020.03.008]

52 **Wang H**, Chen Y, Wu G. SDHB deficiency promotes TGF β -mediated invasion and metastasis of colorectal cancer through transcriptional repression complex SNAIL1-SMAD3/4. *Transl Oncol* 2016; **9**: 512-520 [PMID: 27816688 DOI: 10.1016/j.tranon.2016.09.009]

53 **Hoekstra AS**, de Graaff MA, Briaire-de Bruijn IH, Ras C, Seifar RM, van Minderhout I, Cornelisse CJ, Hogendoorn PC, Breuning MH, Suijker J, Korpershoek E, Kunst HP, Frizzell N, Devilee P, Bayley JP, Bovée JV. Inactivation of SDH and FH cause loss of 5hmC and increased H3K9me3 in paraganglioma/pheochromocytoma and smooth muscle tumors. *Oncotarget* 2015; **6**: 38777-38788 [PMID: 26472283 DOI: 10.18632/oncotarget.6091]

54 **Hao HX**, Khalimonchuk O, Schraders M, Dephoure N, Bayley JP, Kunst H, Devilee P, Cremers CW, Schiffman JD, Bentz BG, Gygi SP, Winge DR, Kremer H, Rutter J. SDH5, a gene required for flavination of succinate dehydrogenase, is mutated in paraganglioma. *Science* 2009; **325**: 1139-1142 [PMID: 19628817 DOI: 10.1126/science.1175689]

55 **Cai D**, Cao J, Li Z, Zheng X, Yao Y, Li W, Yuan Z. Up-regulation of bone marrow stromal protein 2 (BST2) in breast cancer with bone metastasis. *BMC Cancer* 2009; **9**: 102 [PMID: 19338666 DOI: 10.1186/1471-2407-9-102]

56 **Chiang SF**, Kan CY, Hsiao YC, Tang R, Hsieh LL, Chiang JM, Tsai WS, Yeh CY, Hsieh PS, Liang Y, Chen JS, Yu JS. Bone Marrow Stromal Antigen 2 Is a Novel Plasma Biomarker and Prognosticator for Colorectal Carcinoma: A Secretome-Based Verification Study. *Dis Markers* 2015; **2015**: 874054 [PMID: 26494939 DOI: 10.1155/2015/874054]

57 **Lee JH**, Jung S, Park WS, Choe EK, Kim E, Shin R, Heo SC, Lee JH, Kim K, Chai YJ. Prognostic nomogram of hypoxia-related genes predicting overall survival of colorectal

cancer-Analysis of TCGA database. *Sci Rep* 2019; **9**: 1803 [PMID: 30755640 DOI: 10.1038/s41598-018-38116-y]

58 **Hu W**, Li S, Zhang S, Xie B, Zheng M, Sun J, Yang X, Zang L. GJA1 is a Prognostic Biomarker and Correlated with Immune Infiltrates in Colorectal Cancer. *Cancer Manag Res* 2020; **12**: 11649-11661 [PMID: 33235496 DOI: 10.2147/CMAR.S235500]

59 **Sharma BR**, Kanneganti TD. Inflammasome signaling in colorectal cancer. *Transl Res* 2023; **252**: 45-52 [PMID: 36150688 DOI: 10.1016/j.trsl.2022.09.002]

60 **An N**, Zhao Y, Lan H, Zhang M, Yin Y, Yi C. SEZ6L2 knockdown impairs tumour growth by promoting caspase-dependent apoptosis in colorectal cancer. *J Cell Mol Med* 2020; **24**: 4223-4232 [PMID: 32105413 DOI: 10.1111/jcmm.15082]

61 **Lei X**, Lei Y, Li JK, Du WX, Li RG, Yang J, Li J, Li F, Tan HB. Immune cells within the tumor microenvironment: Biological functions and roles in cancer immunotherapy. *Cancer Lett* 2020; **470**: 126-133 [PMID: 31730903 DOI: 10.1016/j.canlet.2019.11.009]

62 **Chen X**, Chen H, Yao H, Zhao K, Zhang Y, He D, Zhu Y, Cheng Y, Liu R, Xu R, Cao K. Turning up the heat on non-immunoreactive tumors: pyroptosis influences the tumor immune microenvironment in bladder cancer. *Oncogene* 2021; **40**: 6381-6393 [PMID: 34588621 DOI: 10.1038/s41388-021-02024-9]

63 **Erkes DA**, Cai W, Sanchez IM, Purwin TJ, Rogers C, Field CO, Berger AC, Hartsough EJ, Rodeck U, Alnemri ES, Aplin AE. Mutant BRAF and MEK Inhibitors Regulate the Tumor Immune Microenvironment via Pyroptosis. *Cancer Discov* 2020; **10**: 254-269 [PMID: 31796433 DOI: 10.1158/2159-8290.CD-19-0672]

64 **Galon J**, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pagès C, Tosolini M, Camus M, Berger A, Wind P, Zinzindohoué F, Bruneval P, Cugnenc PH, Trajanoski Z, Fridman WH, Pagès F. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* 2006; **313**: 1960-1964 [PMID: 17008531 DOI: 10.1126/science.1129139]

65 **Prall F**, Dührkop T, Weirich V, Ostwald C, Lenz P, Nizze H, Barten M. Prognostic role of CD8+ tumor-infiltrating lymphocytes in stage III colorectal cancer with and

without microsatellite instability. *Hum Pathol* 2004; **35**: 808-816 [PMID: 15257543 DOI: 10.1016/j.humpath.2004.01.022]

66 **Laghi L**, Bianchi P, Miranda E, Balladore E, Pacetti V, Grizzi F, Allavena P, Torri V, Repici A, Santoro A, Mantovani A, Roncalli M, Malesci A. CD3+ cells at the invasive margin of deeply invading (pT3-T4) colorectal cancer and risk of post-surgical metastasis: a longitudinal study. *Lancet Oncol* 2009; **10**: 877-884 [PMID: 19656725 DOI: 10.1016/S1470-2045(09)70186-X]

67 **Lu H**, Sun Y, Zhu Z, Yao J, Xu H, Huang R, Huang B. Pyroptosis is related to immune infiltration and predictive for survival of colon adenocarcinoma patients. *Sci Rep* 2022; **12**: 9233 [PMID: 35655081 DOI: 10.1038/s41598-022-13212-2]

68 **Ge P**, Wang W, Li L, Zhang G, Gao Z, Tang Z, Dang X, Wu Y. Profiles of immune cell infiltration and immune-related genes in the tumor microenvironment of colorectal cancer. *Biomed Pharmacother* 2019; **118**: 109228 [PMID: 31351430 DOI: 10.1016/j.biopha.2019.109228]

69 **Deng D**, Luo X, Zhang S, Xu Z. Immune cell infiltration-associated signature in colon cancer and its prognostic implications. *Aging (Albany NY)* 2021; **13**: 19696-19709 [PMID: 34349038 DOI: 10.18632/aging.203380]

70 **Van Opdenbosch N**, Lamkanfi M. Caspases in Cell Death, Inflammation, and Disease. *Immunity* 2019; **50**: 1352-1364 [PMID: 31216460 DOI: 10.1016/j.immuni.2019.05.020]

Figure 1 Identifies 11 cell clusters with different annotations based on single-cell RNA sequencing-seq data, revealing a high degree of cellular heterogeneity in colorectal cancer cells. A: Selection of 88 samples from the GSE178341 dataset, followed by quality control, resulted in the inclusion of 115489 cells in the analysis, which were classified into 38 independent clusters. Different colors denote distinct clusters; B: Uniform Manifold Approximation and Projection distribution highlighting different cell types; C: Dot plot depicting cell type marker genes. Circle size corresponds to the proportion of gene expression in the cell cluster, with darker colors indicating higher average expression; D: Violin plot illustrating differential genes for each cell type; E: Heatmap showcasing the top 2 differential genes for each cell type; F: Proportion of each cell population in different samples, including epithelial cells (Epi, 27.79%), myeloid cells (DC, 1.55%; Macrophage, Macro, 4.86%; Monocyte, Mono, 4.16%), mast cells (Mast, 1.11%), endothelial cells (Endo, 2.28%), Stromal cells (Stroma, 3.03%), plasma (Plasma, 8.75%), B cells (B, 11.85%), NK cells (NK, 5.58%), and T cells (T, 29.02%).

Figure 2 Heat map of 125 pyroptosis-related genes in cell types. The heat map depicts the expression of 125 pyroptosis-related genes across 11 cell types: T cells, NK cells, B cells, Plasma, Epithelial cells, Myeloid cells (DC, Macrophage, Monocyte), Stromal cells, Mast cells, and Endothelial cells. A: The color gradient from blue to red represents the gradual increase in gene expression; B: Specific expression of *GZMA* in the cluster where NK cells and T cells are located; C: Specific expression of *IL-1B* in Monocytes; D: Correlation analysis between pyroptosis-related genes in intersecting cells. Blue represents a negative correlation, while red represents a positive correlation.

Figure 3 CellChat and GSVA. A: Graph illustrating the quantity and strength of interactions among primary cell clusters; B: Uniform Manifold Approximation and Projection plots displaying 27 subsets. The accompanying legend identifies the subgroups; C: Analysis of cell communication within the 27 subsets; D: Examination of the SPP1 signaling pathway interaction within each cluster; E: Interactions originating from a subset of gamma-delta T cells (T05). The X-axis represents the cell pair, and the Y-axis represents the receptor-ligand pair; F: Interactions of other cell subsets with gamma-delta T cells (T05) subsets; G: Interactions involving key cell subsets (M01, T05, FB2, Plasmablasts, Schwann, and EpiT); H: Presentation of significantly distinct signaling pathways in each cell subset, with the cell subset on the X-axis and the pathway name on the Y-axis. Colors ranging from blue to red indicate higher enrichment of the cell subset.

Figure 4 Immune cell prediction from The Cancer Genome Atlas Dataset. A: Disparities in different immune cell types between tumor and normal samples in the The Cancer Genome Atlas (TCGA) dataset. Normal samples are denoted in green, and tumor samples are denoted in red. Significance levels are indicated as follows: *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$; B: Heatmap illustrating the correlation among highly expressing immune cells in the TCGA dataset. The color scale of blue, white, and red denotes the strength of correlation, with darker colors signifying stronger correlations. Red indicates a positive correlation, while blue indicates a negative correlation.

Figure 5 Weighted co-expression network analysis co-expression identifies cell type-dependent modules. A: Analysis of the Scale-free Fit Index for soft threshold power (β) in the range of 1-34; B: Assessment of the Average Connectivity across soft-threshold powers (β) ranging from 1 to 34; C: Hierarchical clustering tree depicting genes based on their topological overlap; D: Correlation plot illustrating relationships between modules.

Figure 6 Weighted co-expression network analysis Co-expression Modules and Cell Types. The X-axis represents cell types provided by CIBERSORTx, while the Y-axis represents the Weighted co-expression network analysis co-expression modules.

Figure 7 Differential analysis, correlation, and enrichment analysis of pyroptosis-related genes. A: Heatmap illustrating the expression profiles of the top 14 differential pyroptosis-related genes. Colors range from blue to red, indicating a gradual increase in expression levels. The color bar above distinguishes (non-colorectal cancer) non-CRC tissues (N) in blue and CRC tissues (T) in red; B: Volcano plot depicting the results of CRC *vs* non-CRC differential analysis; C: Correlation network diagram highlighting highly connected pyroptotic genes; D: Results of Gene Ontology enrichment analysis for 125 pyroptosis-related genes. The bubble plot displays the top 10 most significant enriched functions. The X-axis represents Gene Ratio, and the color of the bubbles ranges from blue to red, with red indicating more significant enrichment. The Y-axis denotes the name of the pathway.

Figure 8 Exploring typing by consensus clustering and single-sample gene set enrichment analysis. A: Heatmap depicting the concordance clustering matrix, with values ranging from 0 (impossible to cluster together) to 1 (always clustered together). Shades of white to dark blue represent the scale of concordance; B and C: Consistent CDF plot and Delta Area Plot; D: Cluster analysis using tSNE algorithm; E: Single-sample gene set enrichment analysis (ssGSEA) of immune cells. Legend includes tumor stage, gender, and age; F: ssGSEA of immune function. Legend includes tumor stage, gender, and age.

Figure 9 Prognostic model based on pyroptosis-related genes. A: Construction of a fitting model using LASSO regression, illustrating changes in the lambda value of 10 pyroptosis-related genes significantly associated with prognosis. The X-axis represents the Log λ value, and the Y-axis represents the coefficient; B: Cross-validation analysis

determining the optimal lambda value for the fitted model. The X-axis represents the logized lambda value, the Y-axis represents the error of the model, and the dashed line on the left signifies the lambda value minimizing the error and the number of screened features; C: Risk map of the training set, where red dots represent high-risk patients, and light blue represents low-risk patients; D: Risk map of the test set; E: Survival curve of the training set ($P = 0.002$), where a smaller P value indicates higher accuracy; F: Survival curve of the test set ($P = 0.009$); G: Receiver operating characteristic (ROC) curve of the training set [area under the curve (AUC) = 63.8%], where a higher AUC signifies greater accuracy; H: ROC curve for the test set (AUC = 63.6%); I: ROC curve for 1-, 3-, and 5-year calculated from the risk score in the training set; J: ROC curve for 1-, 3-, and 5-year calculated from the risk score in the test set.

Figure 10 Construction of the Nomogram. A: Forest plot illustrating the influence of clinicopathological features; B: Nomogram integrating multi-omics data with clinicopathological features; C: Calibration curve of the overall survival nomogram, where the diagonal dashed line represents the ideal nomogram. *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$.

Figure 11 Immunoassays of prognostic models. A: Positive correlation between AIM2 gene expression and the abundance of activated memory CD4⁺ T cells; B: Inverse correlation between UBE2D2 gene expression and the cellular abundance of regulatory T cells (Tregs); C: Correlation of immune cell infiltration with high and low-risk groups; D: Significant differences observed in the abundance of Tregs between the high and low-risk groups.

Figure 12 Tumor mutation burden analysis. A: Waterfall plot illustrating the top 10 frequently mutated genes; B: Dot plot presenting the results of tumor mutation burden analysis, with the median tumor mutational burden at 1.78/Mb; C: Survival analysis of high and low tumor mutation burden groups; D: Comparison of tumor mutation burden between high and low prognostic risk groups.

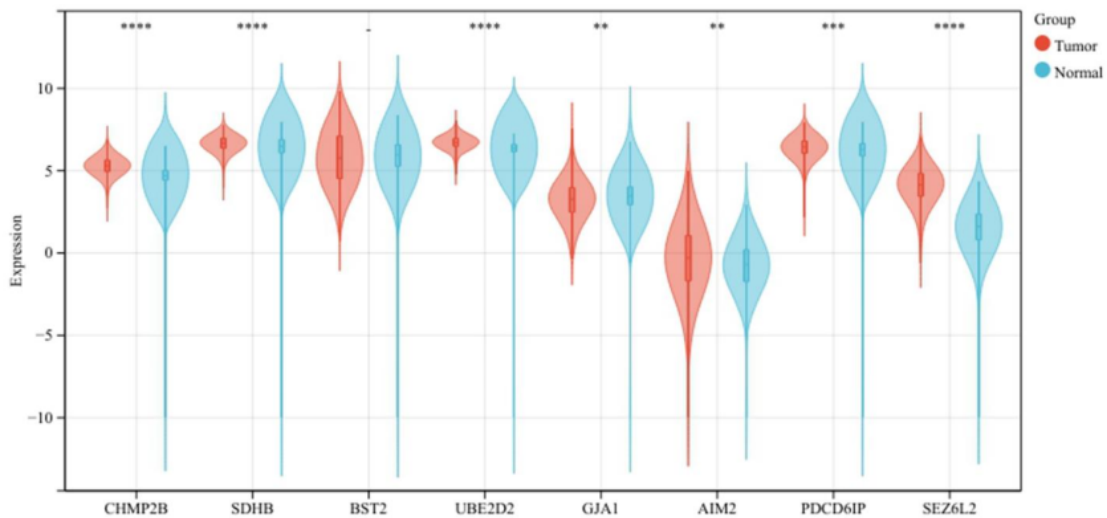


Figure 13 Expression of 8 prognosis-related genes in normal and tumor samples from The Cancer Genome Atlas-COADREAD: *: $p \leq 0.05$; **: $p \leq 0.01$; *: $p \leq 0.001$; ****: $p \leq 0.0001$. *CHMP2B*, *SDHB*, *UBE2D2*, *GJA1*, *AIM2*, *PDCD6IP*, and *SEZ6L2* exhibited significant differential expression between colorectal cancer and normal samples. No significant expression difference was found for *BST2*.**

Table 1

Table 2

Table 3

Table 4

Table 5

10%

SIMILARITY INDEX

PRIMARY SOURCES

- 1

www.frontiersin.org
Internet

206 words — 4%
- 2

[Qingqing Zhou, Yamei Wang, Chenqi Xin, XiaoMing Wei, Yuan Yao, Liang Xia. "Identification of telomere-associated gene signatures to predict prognosis and drug sensitivity in glioma", Computers in Biology and Medicine, 2023](#)
Crossref

41 words — 1%
- 3

www.researchsquare.com
Internet

38 words — 1%
- 4

[Jumin Xie, Ruxi Chen, Qingzhi Wang, Hui Mao. "Exploration and validation of Taraxacum mongolicum anti-cancer effect", Computers in Biology and Medicine, 2022](#)
Crossref

34 words — 1%
- 5

[Ganggang Wang, Jian Li, Linggang Zhu, Zhijie Zhou, Zenghui Ma, Hao Zhang, Yulong Yang, Qiang Niu, Xiaoliang Wang. "Identification of hepatocellular carcinoma-related subtypes and development of a prognostic model: a study based on ferritinophagy-related genes", Discover Oncology, 2023](#)
Crossref

33 words — 1%
- 6

bmccancer.biomedcentral.com
Internet

29 words — 1%

7

www.biorxiv.org

Internet

27 words — < 1%

8

Wei Song, Jun Ren, Rensheng Xiang, Can Kong, Tao Fu. "Identification of pyroptosis-related subtypes, the development of a prognosis model, and characterization of tumor microenvironment infiltration in colorectal cancer", OncoImmunology, 2021

Crossref

26 words — < 1%

9

assets.researchsquare.com

Internet

21 words — < 1%

10

Yujun Luo, Jinghua wang, Xinqi Qiu, Jiayu Huang et al. "Development and Validation of a Novel Mitophagy-Related Gene Prognostic Signature for Glioblastoma multiforme", Research Square Platform LLC, 2022

Crossref Posted Content

18 words — < 1%

11

Chi Zhang, Jiangnan Xia, Xiujuan Liu, Zexing Li, Tangke Gao, Tian Zhou, Kaiwen Hu. "Identifying prognostic genes related PANoptosis in lung adenocarcinoma and developing prediction model based on bioinformatics analysis", Scientific Reports, 2023

Crossref

16 words — < 1%

12

peerj.com

Internet

16 words — < 1%

13

www.aging-us.com

Internet

16 words — < 1%

14

Yuxuan Wang, Weikang Chen, Minqi Zhu, Lei Xian. "Ferroptosis-Related Gene Signature and Patterns of Immune Infiltration Predict the Overall Survival in Patients With Lung Adenocarcinoma", Frontiers in Molecular Biosciences, 2021

15 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	< 15 WORDS
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 10 WORDS