

Research Article

# Novel pyroptosis-associated genes signature for predicting the prognosis of sarcoma and validation

Hao Wen<sup>1,\*</sup>, Dandan Guo<sup>2,\*</sup>, Zhenguo Zhao<sup>3</sup>, Xin Xin<sup>2</sup>, Qi Shi<sup>2</sup>, Jiachen Cao<sup>2</sup>, Lingxie Song<sup>2</sup>, Yuliang Jiang<sup>4</sup>, Chunxia Liu<sup>5</sup> and  Feng Li<sup>1,2</sup>

<sup>1</sup>Department of Pathology and Key Laboratory for Xinjiang Endemic and Ethnic Diseases, The First Affiliated Hospital, Shihezi University School of Medicine, Shihezi 832002, China; <sup>2</sup>Medical Research Center and Department of Pathology, Beijing Institute of Respiratory Medicine and Beijing Chao-Yang Hospital, Capital Medical University, Beijing 100020, China; <sup>3</sup>Department of Orthopaedics, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; <sup>4</sup>Department of Oncology, Beijing Chao-Yang Hospital, Capital Medical University, Beijing 100020, China; <sup>5</sup>Department of Pathology, The Second Affiliated Hospital of Guangzhou Medical University, Guangzhou 510260, China

**Correspondence:** Feng Li (lifeng7855@126.com) or Chunxia Liu (liuliu2239@sina.com) or Yuliang Jiang (yuliang.jiang@ccmu.edu.cn)



**Background:** Sarcoma is a rare mesenchymal malignant tumor. Recently, pyroptosis has been reported to be a mode of programmed cell death. Nonetheless, levels of pyroptosis-associated genes in sarcoma and its relevance to prognostic outcomes are yet to be elucidated. **Results:** Sarcoma cases were classified into two subtypes with regards to differentially expressed genes. We established a profile composed of seven genes and classified the sarcoma patients into low- and high-risk groups through least absolute shrinkage and selection operator Cox regression. Survival rate of low-risk sarcoma patients was markedly higher, relative to high-risk group ( $P < 0.001$ ). In combination with clinical features, the risk score was established to be an independent predictive factor for OS of sarcoma patients. Chemotherapeutic drug sensitivity response analysis found 65 drugs with higher drug sensitivity in low-risk, than in high-risk group and 14 drugs with higher drug sensitivity in the high-risk patient group, compared with low-risk patient group. In addition, functional enrichment, pathway and gene mutation of the two modules were analyzed. Finally, we used qRT-PCR to detect the expression of seven pyroptosis-related genes in tumor cells, and human skeletal muscle cells, compared with human skeletal muscle cells, PODXL2, LRRC17, GABRA3, SCUBE3 and RFLNB genes show high expression levels in tumor cells, while IGHG2 and hepatic leukemia factor show low expression levels in tumor cells. **Conclusions:** Our research suggest that pyroptosis is closely associated with sarcoma, and these findings confirm that pyroptosis-associated seven genes have a critical role in sarcoma and are potential prognostic factors for sarcoma.

\*These authors contributed equally to this work and should be considered co-first authors.

Received: 23 May 2022  
Revised: 10 September 2022  
Accepted: 13 September 2022

Accepted Manuscript online:  
26 September 2022  
Version of Record published:  
09 December 2022

## Introduction

Sarcoma is a kind of rare malignancy of mesenchymal tissues [1]. The disease tends to occur in mesenchymal tissue, which includes bones, fat, joints and muscles [2]. Sarcomas are grouped into two, osteosarcoma and soft-tissue sarcoma [3]. Sarcoma pathogenesis is associated with fusion genes, chromosome mutations and other factors, but its underlying pathogenic mechanisms are still unclear. Although the source, histological and molecular markers of sarcomas differ, their common feature is poor disease prognosis [4]. At present, the principal treatments are radiotherapy and chemotherapy in the clinical, and no targeted drugs have been proved to be effective [5]. Therefore, it is urgent to further study the molecular mechanism and signaling pathway of sarcoma.

Programmed cell death mechanisms, such as apoptosis, necroptosis and pyroptosis have distinct inflammatory outcomes. Among them, the inflammation of pyroptosis is more severe compared with apoptosis [6]. Pyroptosis is a programmed cell death also called inflammatory necrosis [7]. Pyrocytes are featured with cell swellings and various vesicular projections. Under electron microscopy, pyroptosis cells first appear in form of many vesicles. Upon the formation of these vesicles, they form holes in cell membranes, leading to rupture and release of their contents [8]. Gasdermin family is the major genes of pyroptosis including gasdermin-A, -E as well as peptidylarginine deiminase 5 (DFNB59 or P2RX7) [9]. The gasdermin protein family can be cleaved and polymerized, resulting in the splitting of N- and C-terminal linking domains and the release of activated N-terminal regions. These released regions bind membrane lipids, cardiolipin as well as phosphatidylinositol and are located in cell membrane pores [10,11]. The cell gasdermin protein family in the cell membranes are gradually released from pores of 10–20 nm, and the cell components are gradually released via membrane pores, leading to enhanced inflammatory reactions. The cells slowly flattened and produced 1–5 µm of apoptotic vesicle-like processes (blebbed vesicles). The cells slowly expanded to rupture of plasma membrane, which was characterized by nuclear condensation as well as chromatin DNA fragmentation [12,13]. Initially, pyroptosis was established to be a crucial pathomechanism in fighting infection and is also involved in tumor occurrence. Proinflammatory cytokines, inflammatory vesicles and gasdermin proteins are reported to be key components of pyroptosis and are relevant to tumor genesis, invasion and metastasis [14]. Dupaul-Chicoine et al. reported that when the inflammatory vesicle-associated genes (CASP1 and NLRP3) were knocked out in transgenic mice, they were predisposed to colon cancer development, relative to mice with the wild-type gene [15]. In addition, unlike apoptosis, various risk-related signaling molecules as well as cytokines are activated and thereafter released when pyroptosis occurs, accompanied by severe inflammatory reactions and immune system activation [16]. The strong proinflammatory result of pyroptosis is correlated with tumor immune microenvironment regulation. Expression deficiency of GSDMD was followed by a marked decrease in numbers and activities of CD8<sup>+</sup>T lymphocytes [17]. The crucial role of pyroptosis in NK cell antitumor function was also demonstrated in a recent study [18].

According to the literature, we conclude that pyroptosis is vital in tumor occurrence as well as anticancer processes. However, its precise roles in sarcomas are rarely studied. Therefore, this was a systematic analysis to assess the association between sarcomas and levels of pyroptosis-associated genes, and to provide new therapeutic targets and options for sarcomas.

## Materials and methods

### SARC datasets and preprocessing

Analyses of Cancer Genome Atlas (TCGA) datasets were downloaded in UCSC Xena data browser (<https://xenabrowser.net>). We converted the number of fragments per kilobase of non-overlapped exon per million fragments mapped (FPKM) to Gene expression levels which were then quantified as transcripts per million reads (TPM) values. The R software (version 3.6.1) and a collection of R/Bioconductor packages were used for analyses.

### Clustering genes of pyroptosis-associated genes

Thirty-three pyroptosis-associated genes were retrieved from prior reviews [19–22]. Clustering of sarcoma transcriptome profiles was performed using NMF implemented in ‘*nmf*’ in R (version NMF.0.20.5). R packages (PCA methods) were used for PCA analyses.

### Building and verifying the pyroptosis-related gene prognostic model

‘Limma’ in R was used for identifying differentially expressed genes (DEGs) with  $P < 0.05$ . Notation of DEGs was as: \* $P < 0.05$ , \*\* $P < 0.01$  and \*\*\* $P < 0.001$ . To evaluate the prognostic significance of pyroptosis-associated genes, we used Cox regression analyses to determine the association between survival status and every gene in TCGA cohort. To preclude omissions, 0.2 was the cut-off  $P$ -value, random survival forest analysis was used to filter variable. LassoCox regression model (‘GLMnet’ in R) was used to identify the candidate genes and establish prognosis models. Finally, seven genes with their coefficients were preserved, and the penalty parameter ( $\lambda$ ) was determined according to the minimum criterion. After standardizing TCGA expression data sets (applying the ‘scale’ function in R), calculate the risk score: Risk score =  $\sum_i X_i \times Y_i$  ( $Y_i$ : levels of gene expressions,  $X_i$ : coefficients). TCGA sarcoma patients were assigned into low-risk and high-risk subgroups based on median risk scores. Circos plot was created using modified functions from the R package ‘RCircos’. Kaplan–Meier analyses were used for comparisons of OS time between the two subgroups. Receiver operating characteristic (ROC) curve analysis was performed for 1–5 years using ‘Survival’, ‘Survivor’ and ‘timeROC’ R packages.

## GDSC cell line drug response data

The GDSC large-scale composite reaction screening dataset was obtained [23], which included 990 human cancer cell lines from 25 cell lineages and 255 chemical compounds. The 255 compounds came from a variety of sources, such as candidates for clinical drugs, FDA-permitted drugs, and previously reported chemical sensitivity analysis assays. For drug sensitivity quantification, IC<sub>50</sub> from GDSC were used [23]. In order to determine approved to treat various varieties of cancer drugs, a manual search of the usages of these drugs was performed in Wikipedia and the NCI database (<https://www.cancer.gov/about-cancer/treatment/drugs>). The drugs for every tumor type are shown in. We also downloaded the somatic mutation profiles and tissues of origin for the 990 cancer cell lines [24]. We applied the propagation algorithm based on ESP to divide every cancer cell line into two clusters, using only the sub-network data. We then used rank-based Wilcoxon-type statistics for comparisons of differences in drug reactions between the clusters.

## Immunocorrelation analyses and gene expressions

Meanwhile, we compared CIBERSORT [25,26], ESTIMATE [27], MCPcounter [28], single-sample gene set enrichment analysis (ssGSEA) [29] as well as TIMER [30] algorithms, to evaluate cell constituents or cellular immune reactions between high- and low-risk groups with a basis on the characteristics of pyroptosis-associated genes. Heatmaps were employed to reveal differences in immune reactions under various algorithms. Moreover, ssGSEA was used for quantifying subsets of tumor-infiltrating immune cells between the groups and for immune function evaluation. Immune checkpoints were acquired from literature. Cluster Profiler in R was used for gene set enrichment analysis (GSEA).

## Function and pathway enrichment analyses

GO offers regulated and structured vocabularies that model cell components (CC), biological process (BP) and molecular function (MF) [31]. KEGG is a commonly used pathway analysis method, which contains 16 major databases, assigned into systematic information, chemical and genomics information [32]. We combined GO and KEGG analyses to conduct DEGs. The cut-off was  $p < 0.05$ . All gene sets related to GO and KEGG were downloaded from MSigDB database.

## Analysis of tumor mutation in sarcoma

As a quantifiable biomarker, tumor mutation burden (TMB) can be used to reflect the counts of mutations contained in cancer cells. In our study, tumor mutational burden of sarcoma patients was calculated as follows:  $TMB = S_n \times 1,000,000/n$  ( $S_n$  is the absolute somatic mutation value, and  $n$  denotes the counts of exon base coverage depth  $\geq 100\times$ ) [33].

## Cell lines and culture conditions

Human skeletal muscle cells (HSKMC) and Rhabdomyosarcoma cells (RH30) were purchased from American type culture collection (ATCC, U.S.A.). HSKMC were cultured in Mesenchymal Stem Cell Basal Medium with Primary Skeletal Muscle Cell Growth Kit (ATCC, U.S.A.) at 37°C in a humidified atmosphere containing 5% CO<sub>2</sub>. RH30 were cultured in were cultured in RPMI 1640 medium with 10% fetal bovine serum (FBS) (Gibco, Grand Island, NY, U.S.A.) and 1% penicillin-streptomycin at 37 °C in a humidified atmosphere containing 5% CO<sub>2</sub>. Synovial sarcoma cells (SW982) were purchased from FUXIANG Biotechnology (Shanghai, China). SW982 were cultured in were cultured in Dulbecco's Modified Eagle Medium with 10% fetal bovine serum (FBS) (Gibco, Grand Island, NY, U.S.A.) and 1% penicillin-streptomycin at 37 °C in a humidified atmosphere containing 5% CO<sub>2</sub>.

## RNA extraction and real-time quantitative PCR (qRT-PCR)

Total RNA was extracted using Trizol (Invitrogen) according to manufacturer's protocol. One microgram of total RNA was used to synthesize cDNA using random hexamers and the Superscript III First-Strand Synthesis System for RT-PCR (Invitrogen). The real-time PCR was performed using the Fast SYBR Green Master Mix (Applied Biosystems) and run on a 7500 Fast Real-Time PCR System machine (Applied Biosystems) in the fast mode. The relative number of mRNAs was calculated by using the  $\Delta\Delta C_t$  ( $C_t$ , threshold cycle) method.

Primers sequence are as follows:

---

IGHG2 (94 bp):  
Forward 5'-AGGGACAACCTCCGCAAACAC-3'  
Reverse 5'-TCCCCGAATGTGCTTTTCGC-3'

PODXL2 (144 bp):  
Forward 5'-CTCCCTGCTAGACCTCCTG-3'  
Reverse 5'-TGCAGAATCCGAGACTTTCAT-3'

LRR17 (177 bp):  
Forward 5'-AGAAGCCGAGTGAATCATGGC-3'  
Reverse 5'-GTGCAGCAAATCCTGAGGC-3'

GABRA3 (145 bp):  
Forward 5'-CAAGGGGAATCAAGACGACAA-3'  
Reverse 5'-CGTCCAGAAGACGATCCAAGAT-3'

SCUBE3 (135 bp):  
Forward 5'-CAGAACACCCCGAGGTCATAC-3'  
Reverse 5'-GCCAGGGATGTTGACACAGTC-3'

HLF (103 bp):  
Forward 5'-CTGGGGCCTACCTTATGGGA-3'  
Reverse 5'-GGGGAATGCCATTTTCTGACA-3'

RFLNB (188 bp):  
Forward 5'-AGGCACCTTCATCGACGACG-3'  
Reverse 5'-TAGACGGCCTTGGGGTACTT-3'

GAPDH (197 bp)  
Forward 5'-GGAGCGAGATCCCTCCAAAT-3'  
Reverse 5'-GGCTGTTGTCATCTCTCATGG-3'

---

## Statistical analyses

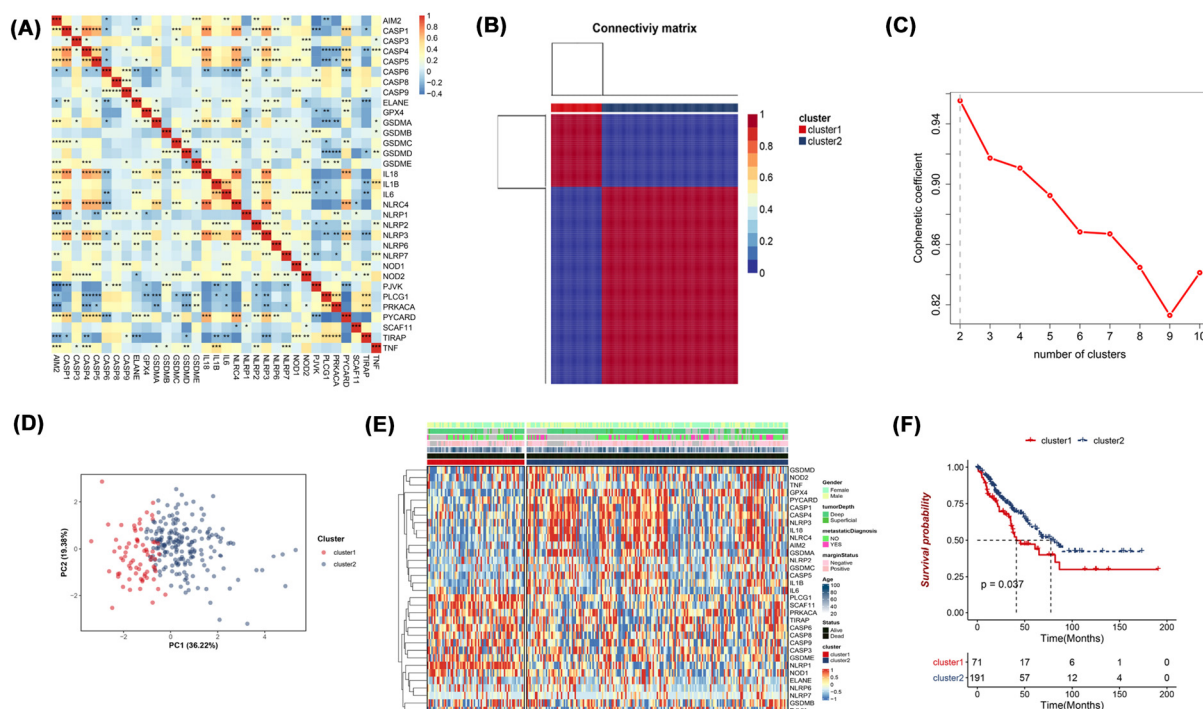
Verification of normality of variables was done by the Shapiro–Wilk normality test. We use the *T*-test to compare differences between two sets of normally distributed variables. Comparisons of non-normally distributed variables was done by Wilcoxon signed rank test. Pearson correlation coefficient is used to measure similarity, and Pearson correlation distance is used to estimate dissimilarity. ‘ggplot2’ in R was used for plot generation, while R package ‘randomForest’, in R, version 3.6.1 was used to build the Random Forest models. Survival-ROC curves were applied with ‘timeROC’ the package. Kaplan–Meier was applied to estimate survival probability and generation of survival curves. Statistical differences in survival data were evaluated by the log-rank test. Analysis and survival curves were by ‘survminer’ and survival in R. Heatmap was generated using pheatmap (v1.0.12). R version 3.6.1 (<https://www.r-project.org/>) was used for analyses. All the tests were two-sided with  $P < 0.05$  denoting significance. Mutated genes were assessed by the R package maftools v2.2.10. GSEA and enrichment analysis are implemented by R package clusterProfiler.

## Result

### Classification of tumors based on pyroptosis-associated genes

To evaluate associations among pyroptosis-associated genes, we calculated pairwise correlations among the expression of 33 pyroptosis-associated genes in sarcoma, which showed that pairs of these genes are commonly linked, and the negative correlations were more frequent than positive correlations (Figure 1A). And then, NMF algorithm was used to cluster sarcoma samples in the TCGA dataset. To investigate the association between levels of 33 pyroptosis-associated genes and sarcoma subtypes, consensus cluster analysis was performed on 262 sarcoma patients in the TCGA cohort. We performed NMF unsupervised clustering of pyroptosis-related genes in classes 2–10 and found that the best clustering was achieved when clustering two classes, when the clustering variable ( $k$ ) was increased from 2 to 10, it was established that at  $k = 2$ , intragroup and intergroup associations were highest and low, respectively, implying that the 262 sarcoma patients can be assigned into two clusters based on the 33 pyroptosis-associated genes (Figure 1B,C). Then, principal component analysis (PCA) was used to verify the clustering situation. PCA analysis indicated two clusters are two distinct groups (Figure 1D). The gene expression profiles as well as clinical characteristics, such as age, gender, tumor depth, metastatic diagnosis. Margin status are displayed in the heatmap. Differences





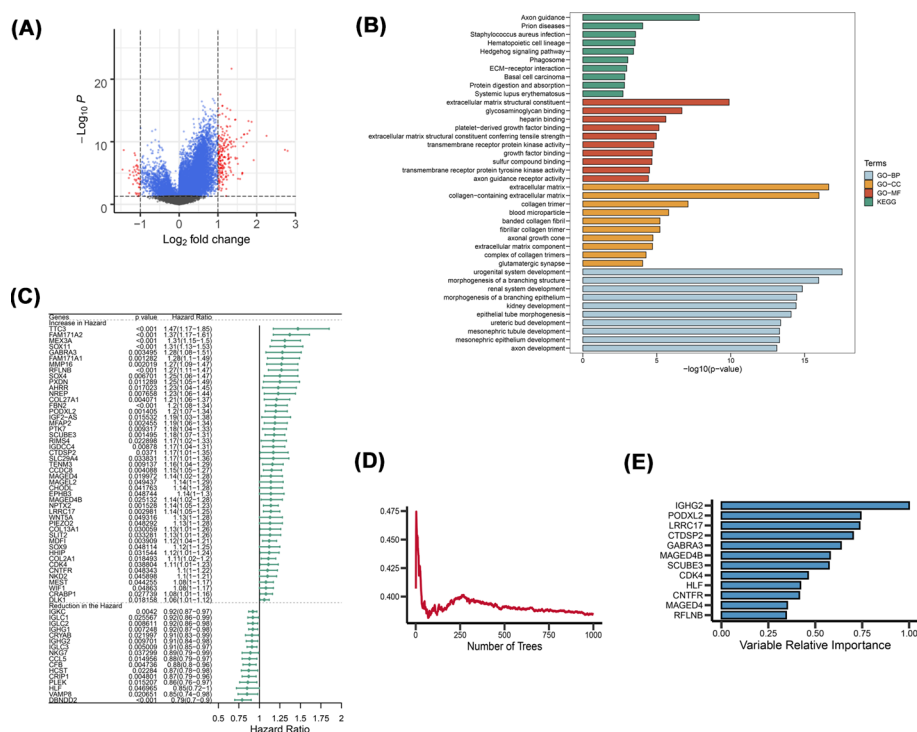
**Figure 1. Tumor classification based on the pyroptosis-related DEGs**

(A) Pairwise correlations among the levels of 33 pyroptosis-associated genes in sarcoma. (B) About 262 Sarcoma patients assigned into two clusters based on consensus clustering matrix ( $k = 2$ ). (C) When the clustering variable ( $k$ ) was increased from 2 to 10, ( $k = 2$ ) is the best clustering. (D) The PCA plot for sarcoma patients, which was based on the pyroptosis-related genes. (E) Heatmap and clinic-pathologic features of the two clusters, as classified by the DEGs. (F) Kaplan–Meier OS curves of the two clusters.

in clinical characteristics between the clusters were minimal (Figure 1E). In an assessment of overall survival rate in the two clusters, patients between the two clusters have a significantly different overall survival rate ( $P=0.037$ , Figure 1F).

## Establishment of prognostic gene models in TCGA

DEG analysis was performed to compare the differential gene expressions of the two clusters. The volcano map shows that 246 genes have been analyzed (Figure 2A). In addition, KEGG pathway enrichment analyses and functional evaluations were done using the Enrichr database, including three types of GO analysis: BP, CC and MF. These genes were highly enriched in Axon guidance in BP. In terms of CC, they were enriched in extracellular matrix structural constituent. As for MF, these genes are mainly associated with extracellular matrix and collagen-containing extracellular matrix. KEGG enrichment showed that 246 DEGs were associated with urogenital system development and other aspects (Figure 2B). In order to obtain more accurate genes related to cell pyroptosis, we reduced DEG genes to 63 by univariate Cox regression analysis dimension reduction (Figure 2C), and then 12 genes (IGHG2, PODXL2, LRRC17, CTDSP2, GABRA3, MAGED4B, SCUBE3, CDK4, hepatic leukemia factor [HLF], CNTFR, MAGED4 and RFLNB) were obtained by random survival forest (Figure 2D,E). A signature of seven genes was established using the optimal  $\lambda$  value through LASSO Cox regression (Figure 3A,B). The risk score formula was: risk score =  $(-0.0588 \times \text{IGHG2 exp.}) + (0.1123 \times \text{PODXL2 exp.}) + (0.0476 \times \text{LRRC17 exp.}) + (0.0905 \times \text{GABRA3 exp.}) + (0.1529 \times \text{SCUBE3 exp.}) + (-0.271 \times \text{HLF exp.}) + (0.1519 \times \text{RFLNB exp.})$ . With regards to the median score as determined by the risk score formula, 262 patients were assigned into low- and high-risk sub-groups (Figure 3C). High-risk group patients were correlated with a high death rate and shorter survival times, relative to low-risk patient group (Figure 3D, on right side of dotted line). Among the seven genes (IGHG2, PODXL2, LRRC17, GABRA3, SCUBE3, HLF and RFLNB), the expression levels of PODXL2, LRRC17, GABRA3, RFLNB and SCUBE3 in high-risk group were evidently higher, relative to low-risk group. Expressions of IGHG2 as well as HLF were markedly low in high-risk group than in the low-risk group (Figure 3E). Subsequently, we analyzed the relationship between these seven genes and the two clusters, as well as the correlation between these seven genes and clinical indicators (Figure 3F), IGHG2, PODXL2, LRRC17, GABRA3, SCUBE3,



**Figure 2. The screening and functional enrichment analyses of the DEGs**

(A) Volcano map of DEGs. (B) GO and pathway analyses of DEGs. (C) Sixty-three were established by univariate Cox Regression analyses for dimensionality reduction. (D) Error rate for the data as a function of the classification tree. (E) The importance values for 12 the predictors.

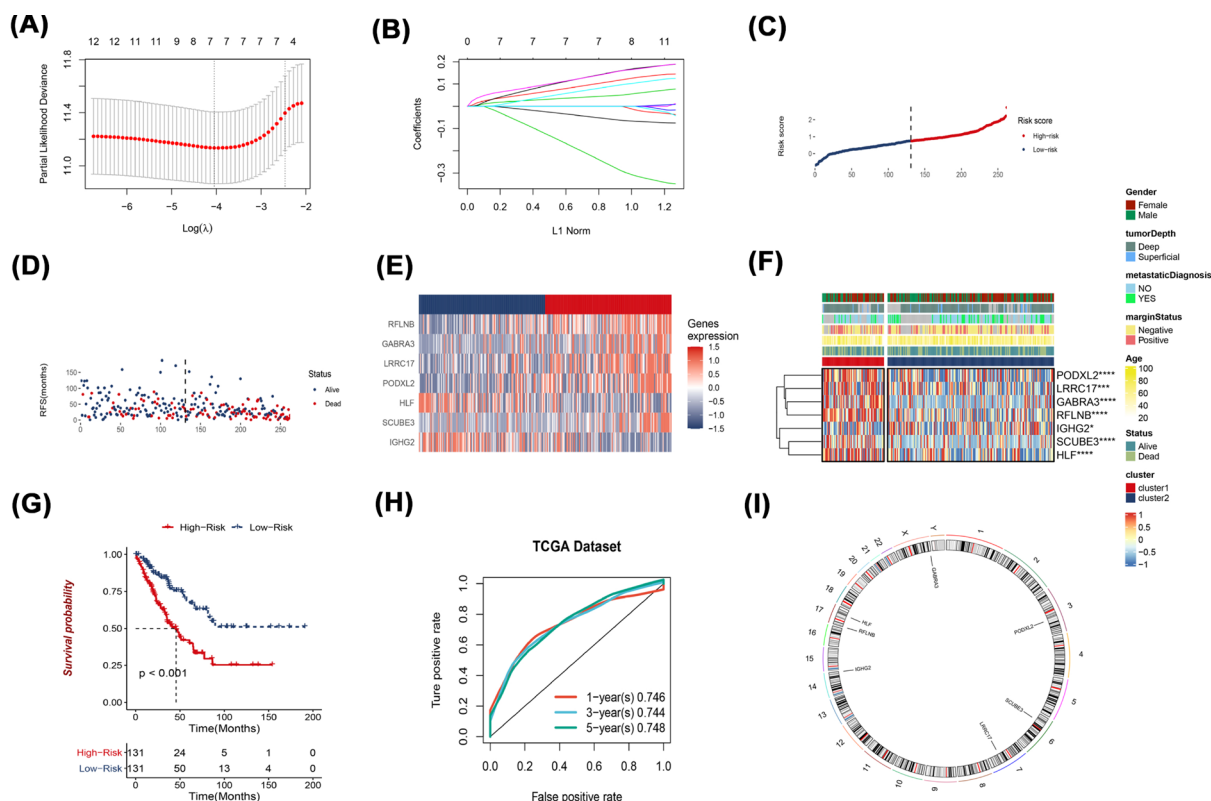
HLF and RFLNB were associated with gender, tumor depth, metastatic diagnosis, margin stage, age, and status. No-table differences in OS time were detected between low- and high-risk groups ( $P < 0.001$ , Figure 3G), patients with in high-risk groups have significantly worse overall survival. Time-dependent ROC analysis was used to assess the models' specificity and sensitivity. Notably, we the area under the ROC curve (AUC) for survival was respectively 0.746, 0.744, and 0.748 for 1, 3 and 5 years (Figure 3H). In addition, these seven genes, are distributed in randomly chromosome and shown in the Circos plot (Figure 3I).

## Prognostic significance of the risk model

The risk score was an independent factor that was a factor in predicting low survival by univariate Cox regression analyses in TCGA cohort (HR = 3.32, 95% CI: 2.35–4.7, Figure 4A). After adjustment for other con-founding factors, multivariate analysis implied that the risk score is a prognostic factor (HR = 3.65, 95% CI: 2–6.68, Figure 4A) for patients with Sarcoma in the cohort. Furthermore, the clinical characteristic heatmaps of the TCGA cohort are shown (Figure 4B) and found that PODXL2, LRRC17, GABRA3, SCUBE3 and RFLNB have positive correlation with risk score. On the contrary, IGHG2 and HLF have negative correlation with risk score.

## Identify drug candidates for risk models

We found that the IC50 values of 65 drugs in high-risk patient group were markedly low relative to those in low-risk patient group. At present, the main chemotherapy drugs for sarcoma include cisplatin, docetaxel, doxorubicin, gemcitabine and so on. IC50 values implying that high-risk group patients were sensitive to chemotherapy with cisplatin, docetaxel, doxorubicin and gemcitabine. Hence, these drugs may be more appropriate for high-risk patients. This may provide a new idea for clinical chemotherapy regimen (Figure 5). In addition, IC50 values of 14 drugs, including erlotinib, gefitinib, etc., were higher in high-risk group, relative to low-risk group, and the high-risk group may not be sensitive to the chemotherapy response of these 14 drugs. IC50 values of all 79 drugs are shown in Supplementary Figures.

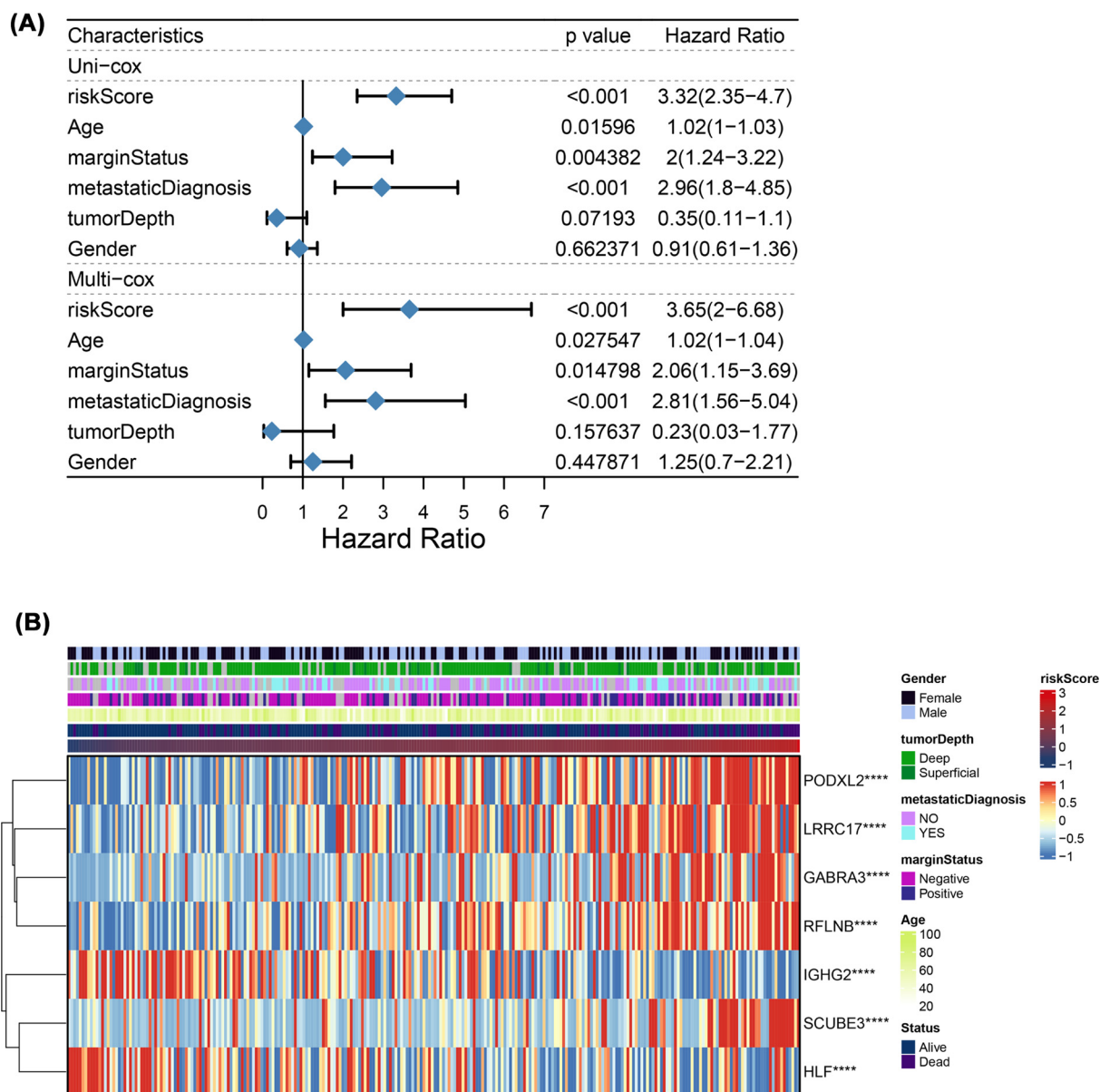


**Figure 3. Creation of the risk signature in TCGA cohort**

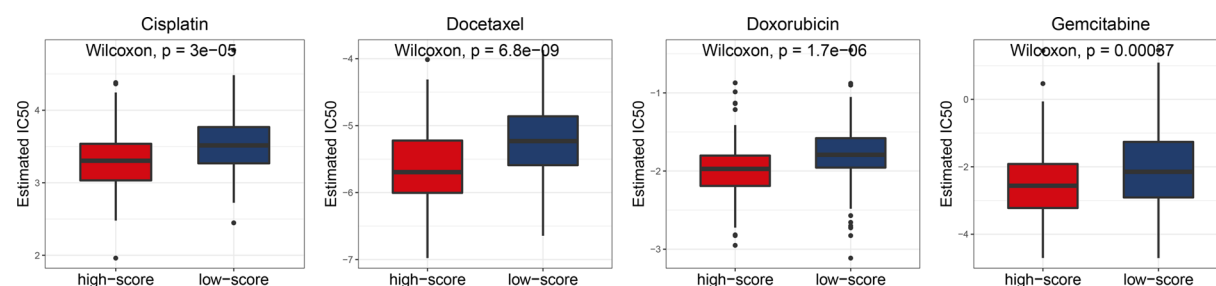
(A) Cross-validation for tuning parameter selections in LASSO regression. (B) LASSO regression of seven genes. (C–E) Risk score distribution, survival status, and expressions of seven prognostic pyroptosis-related genes in sarcoma. (F) Heatmap of connections between clinic-pathologic characteristics and seven genes. (G,H) Overall survival curves for high-/low-risk group sarcoma patients and ROC curve for evaluating the predictive value. (I) Circos plot shows the chromosomal distribution of the seven genes.

## Different functional pathways analyses of the risk models

To explore the reasons for different sensitivity of drugs in high- and low-risk groups, we conducted the GSEA of our models. GO term analyses revealed that molecular function of high-risk patient group was significant enrichment in the cytokine activity, cytokine secretion, fibroblast growth factor binding, muscle system process, muscle tissue development, positive regulation of immune effector processes, fibroblast growth factor receptor signaling pathway regulation, inflammatory reaction regulation, regulation of leukocyte-mediated immunity and skeletal muscle cell differentiation (Figure 6A). KEGG pathway analyses also showed the pathways enrichment of high-risk group in cell cycle, such as chemokine signaling pathway, cytokine–cytokine receptor associations, glioma, hedgehog and p53 signaling pathways, pathways in cancer, T-cell receptor signaling pathway, primary immunodeficiency and VEGF signaling pathway (Figure 6B). To further dig into the reasons for the differences of sensitivity of drugs, we analyzed the different modules in terms of immunity and mutation. The immune responses heatmap is presented. CIBERSORT, TIMER, ssGSEA and MCP counter showed that the levels of infiltration of immune cells were usually lower in high-risk patient group, particularly CD8+ T cells. ESTIMATE revealed that at a higher risk score, tumor purity was higher, and the immune score was lower (Figure 6C). These results are consistent with previous analyses. Moreover, Oncoplot shows the top 30 mutated genes of low- and high-risk groups, among them, mutations of TP53, RB1, MUC16 and some other genes between the groups were markedly different. These differences in immune response and genetic mutation may account for the differences in sensitivity to chemotherapeutic drugs between high- and low-risk groups (Figure 6D).

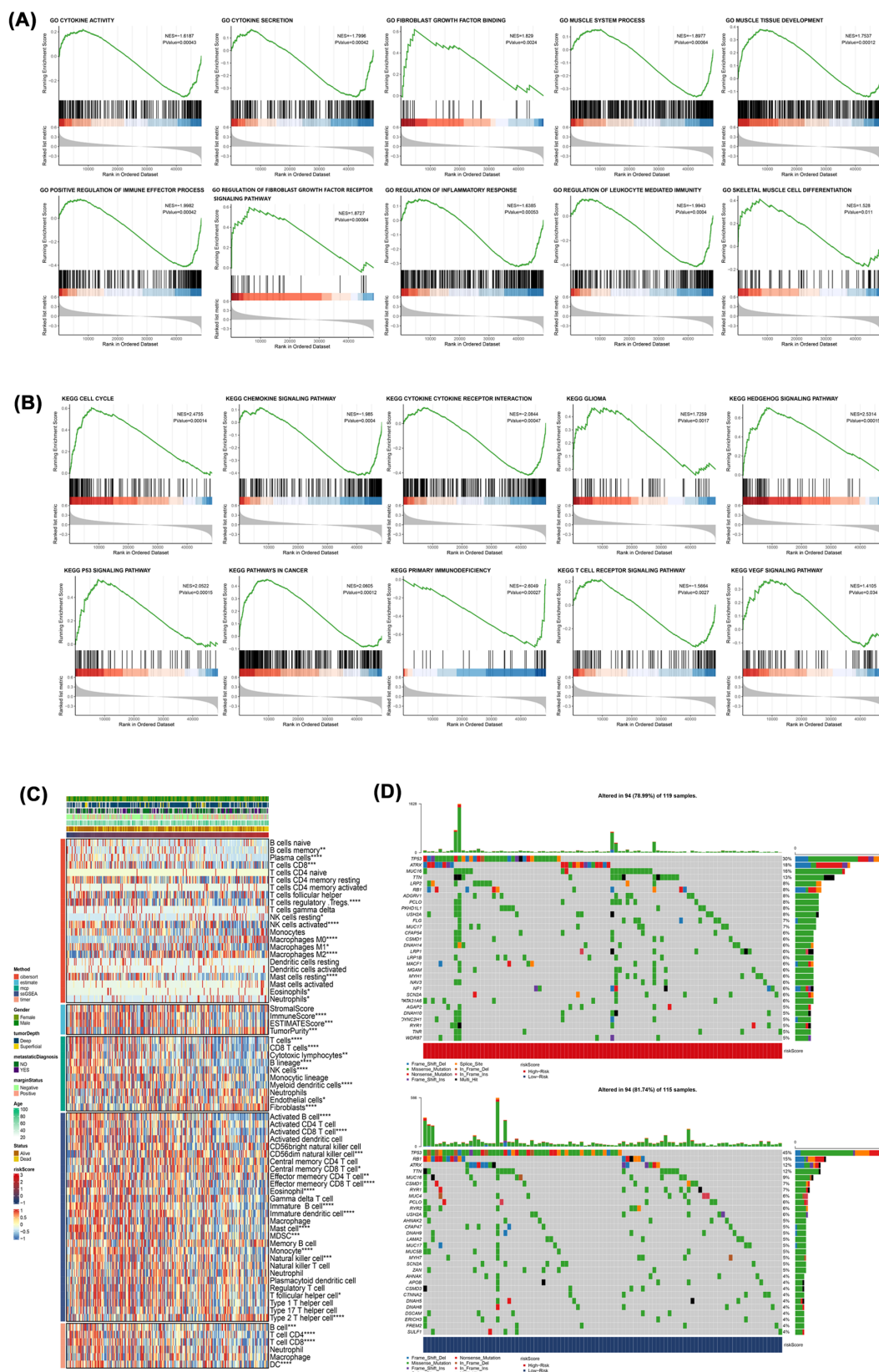


**Figure 4. Univariate as well as multivariate Cox regression analyses of the risk score**  
(A) Univariate and Multivariate analysis for the TCGA cohort. (B) Heatmap of connections between clinic-pathologic characteristics and risk groups.

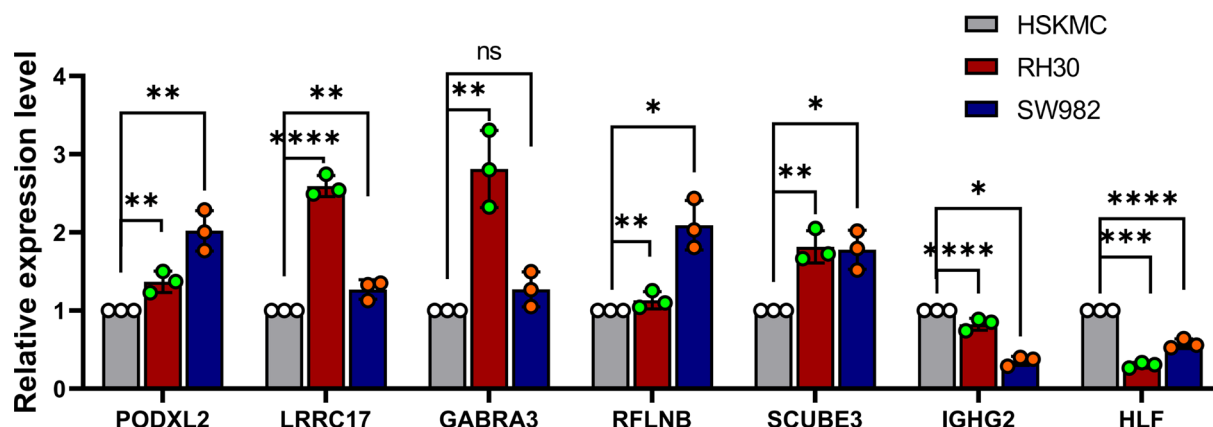


**Figure 5. Chemotherapeutic response for risk models**  
High-risk group patients were sensitive to chemotherapy with cisplatin, docetaxel, doxorubicin and gemcitabine.









**Figure 7. Validation of seven genes expression in RH30, SW982 and HSKMC by qRT-PCR**

Bars represent mean  $\pm$  SEM ( $n=3$ ), \* $P<0.05$ , \*\* $P<0.01$ , \*\*\* $P<0.001$ , \*\*\*\* $P<0.0001$ .

## Validation of pyroptosis-associated genes expression by qRT-PCR

To validate the expression levels of seven pyroptosis-associated genes in sarcoma, we used qRT-PCR to verify RNA expression levels in rhabdomyosarcoma cell line—RH30, synovial sarcoma cell line—SW982, and human normal skeletal muscle cell—HSKMC. The results showed that compared with HSKMC, PODXL2, LRRC17, GABRA3, SCUBE3, and RFLNB were highly expressed in tumor cells, while IGHG2 and HLF were low expressed (Figure 7). This result is consistent with our previous data analysis.

## Discussion

Sarcoma is a rare tumor with local invasion and distant metastasis. The prognosis is poor. Histological biopsy is the most authoritative standard for the diagnosis of Sarcoma [34,35]. The main therapeutic methods are radiation therapy and surgery [36], and there are still no proven target drugs for sarcomas. Up to now, the mechanism of sarcomas is still poorly understood, which is the main obstacle to the development of specific targeted drugs.

Pyroptosis, a newly discovered mode of programmed cell death, plays double roles in tumor occurrence and therapeutic mechanisms. Normal cells are activated by many inflammatory molecules secreted during pyroptosis, causing their transformation into tumor cells [19]. In addition, pyroptosis of tumor cells may be a new treatment target [37].

In sarcomas, the problem of how pyroptosis-associated genes relate with each other, and if they are correlated to patients' survival outcomes, remains unclear. Our study generated a profile characterized by seven pyroptosis-related genes (IGHG2, PODXL2, LRRC17, GABRA3, SCUBE3, HLF and RFLNB) and found that these genes can predict OS in patients. IGHG2 (immunoglobulin heavy constant  $\gamma 2$ ), a protein coding gene, is associated with some kinds of diseases, for example, immunoglobulin k light chain deficiency. And the related pathways of it are creation of C4 and C2 activators and innate immune system [38]. PODXL2 is a member of the CD34 protein family, which is found in endothelial cells and is a ligand for angiostatin, mediating the interaction of white blood cells with the surface of blood vessels [39]. LRRC17 (leucine-rich repeat containing 17), a gene that mainly regulate the osteoblastogenesis process, is primarily expressed in osteoblasts under physiological conditions [40].  $\alpha 3$ , GABRA3- $\gamma$ -aminobutyric acid type A receptor subunit is a protein-coding gene, which is related with diseases including periodic thyrotoxic paralysis and reflex sympathetic dystrophy. The related pathways include ligand-gated ion channel transport and trans-chemical synaptic transport [41]. SCUBE3 (signal peptide, CUB domain and EGF-like domain containing 3) is a protein coding gene. Diseases related with SCUBE3 include facial dysmorphism, short stature and skeletal anomalies with or without cardiac anomalies 2 and bone disease [42]. The gene encodes signal peptides, complement subcomponent C1r/C1s, Uegf, bone morphogenetic protein-1 as well as epidermal growth factor-like domains. Both full-length protein as well as the C-terminal fragment bind transforming growth factor TYPE II receptors to enhance epithelial-mesenchymal transformation and angiogenesis in tumors [43]. HLF, a transcriptional factor, plays an significant regulatory function in various tumors, particularly leukemia and participates in therapy-mediated immunogenic cell death [44]. RFLNB, a member of the Refilin family, is a novel Actin regulatory protein that functions as molecular switches for interconverting the Actin meshwork into bundles. One major trait of this regulatory protein is its short half-life, unique among Actin regulatory proteins [45]. So far, no studies have shown that these seven genes (IGHG2, PODXL2,

LRRC17, GABRA3, SCUBE3, HLF and RFLNB) are directly related to pyroptosis in tumor cells. We first proposed these seven genes and analyzed their function and significance in sarcoma.

We investigated the expressions of 33 known pyroptosis-associated genes in sarcomas, analyzed the relationship between each two genes, and investigated their correlation with clinicopathological features. To assess the association between expressions of 33 pyroptosis-related genes and sarcoma subtypes, the NMF algorithm was used to cluster 262 sarcoma samples in the TCGA dataset. About 262 sarcoma patients assigned into two clusters based on consensus clustering matrix. DEG analysis was performed to compare the differential gene expressions of the two clusters and 246 genes were screened. In order to obtain more accurate genes related to cell pyroptosis, we reduced DEG genes to seven genes by univariate Cox regression analysis, random survival forest and LASSO Cox regression dimension reduction. In addition to this, these seven genes, are distributed in randomly chromosome and shown in the Circos plot. With regards to the median score as determined by the risk score formula, 262 patients were assigned into low- and high-risk sub-groups. The overall survival rate in the two groups has obvious differences, patients with the pyroptosis-related genes exhibited a significantly poor overall survival rate, relative to those without pyroptosis-associated genes. In the present study, the IC50 of the 65 drugs in the high-risk group were markedly low, compared with low-risk group, indicating that these drugs may be suitable for high-risk patients, among them, including cisplatin, docetaxel, doxorubicin and gemcitabine. They are currently the main treatment of sarcoma chemotherapy drugs. To determine the underlying mechanisms of different sensitivities of drugs to the high- and low-risk groups, we applied GSEA to our model. Functional analyses revealed that DEGs between low- and high-risk groups was associated with immune and musculoskeletal pathways. We compared the pathways related to infiltration and activation of immune cells in low- and high-risk patient groups and found that the number of infiltrating immune cells as well as activation levels of immune-associated pathways were generally low in high-risk group, relative to the low-risk patient group. Moreover, we also compared the top 30 mutated genes in low- and high-risk patient groups. Among them, mutations of TP53, RB1 as well as MUC16 markedly differed between the groups. In order to further explore the potential target genes of these seven genes in sarcoma. Based on the TRANSFAC database, we obtained transcription factors as well as motifs for seven genes, and mined transcription factors that have been confirmed by studies in sarcoma based on the TF cancer database (Supplementary Table).

Until now, few studies have clearly shown the significance of pyroptosis, particularly its functional mechanisms in sarcomas. We identified seven genes that may have a critical impact on pyroptosis in sarcomas and may serve as regulatory factors. We evaluated the prognostic significance of these pyroptosis-related genes and gave a theoretical basis for subsequent studies. Taken together, our study suggests that pyroptosis were unable to confirm whether those previously reported regulatory factors also have similar function in the pyroptosis pathway of sarcoma for lack of data, which deserves further study and experimental verification.

Taken together, our research suggests that pyroptosis is closely associated with sarcoma, which is evidenced by the results that genes associated with or unrelated to pyroptosis are expressed differently in sarcoma tissues. In addition, our score based on risk signature generation of seven pyroptosis-associated genes is an independent risk factor for OS prediction in TCGA cohort. DEGs between low-risk and high-risk patient groups are correlated with tumor immunity and mutations, as well as skeletal muscle pathways. At last, we used qRT-PCR to detect the expression of seven pyroptosis-related genes in sarcoma cells (RH30 and SW982) and human skeletal muscle cells, compared with human skeletal muscle cells, PODXL2, LRRC17, GABRA3, SCUBE3 and RFLNB genes show high expression levels in tumor cells, while IGHG2 and HLF show low expression levels in tumor cells. This result confirms our data analysis.

Our study gives a new genetic feature for prognostic prediction of sarcoma patients, and forms a theoretical basis for the studies on of pyroptosis-associated genes in sarcoma patients in the future.

## Declaration of Humans and/or the use of Human Tissue Samples

Data sets used in this study are all from the public database TCGA, and humans and/or the use of human tissue samples are not involved. All methods were carried out in accordance with relevant guidelines and regulations.

### Data Availability

The following information was supplied regarding data availability: The datasets used in the present study are available from Analyses of Cancer Genome Atlas (TCGA) and the TCGA data in the supplement table

### Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

## Funding

This work was supported by the grants from the National Natural Science Foundation of China [grant number 82173057].

## CRediT Author Contribution

**Hao Wen:** Conceptualization, Resources, Data curation, Software, Formal analysis, Supervision, Funding acquisition, Validation, Investigation, Visualization, Methodology, Writing—original draft, Project administration, Writing—review & editing. **Dandan Guo:** Data curation, Writing—original draft. **Zhenguo Zhao:** Writing—review & editing. **Xin Xin:** Writing—original draft, Writing—review & editing. **Qi Shi:** Data curation, Methodology. **Jiachen Cao:** Writing—review & editing. **Lingxie Song:** Writing—review & editing. **Yuliang Jiang:** Validation, Visualization, Writing—review & editing. **Chunxia Liu:** Investigation, Project administration, Writing—review & editing. **Feng Li:** Writing—original draft, Project administration, Writing—review & editing.

## Abbreviations

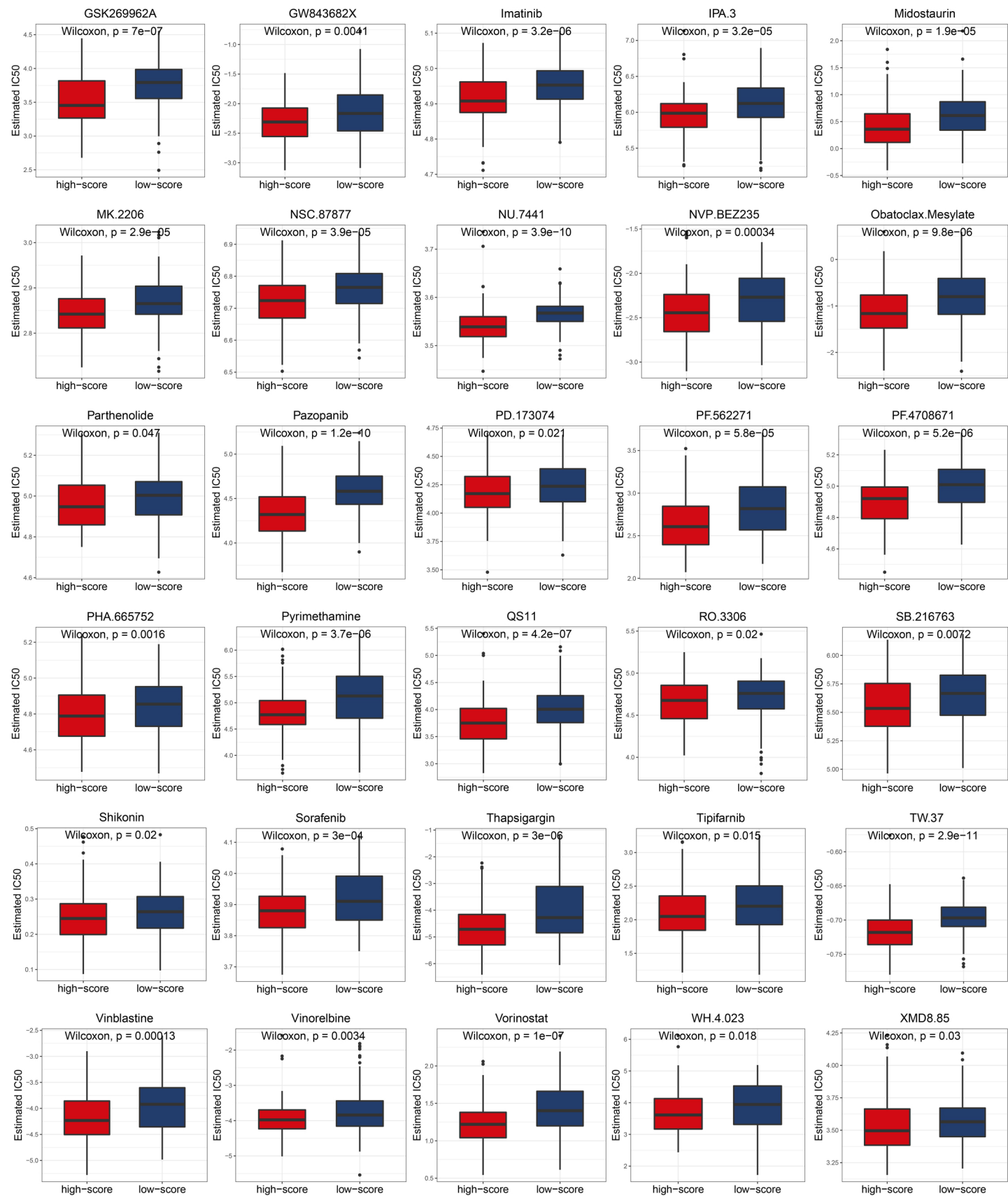
DEG, differentially expressed gene; GSEA, gene set enrichment analysis; HLF, hepatic leukemia factor; HSKMC, Human skeletal muscle cells; LRRC17, leucine-rich repeat containing 17; ROC, receiver operating characteristic; TMB, tumor mutation burden.

## References

- Birdi, H., Jirovec, A., Cortés-Kaplan, S., Werier, J., Nessim, C., Diallo, J. et al. (2021) Immunotherapy for sarcomas: new frontiers and unveiled opportunities. *J. Immunother. Cancer* **9** (2)
- Burningham, Z., Hashibe, M., Spector, L. and Schiffman, J. (2012) The epidemiology of sarcoma. *Clin. Sarcoma Res.* **2**, 14, <https://doi.org/10.1186/2045-3329-2-14>
- Booth, L., West, C., Hoff, D. and Dent, P. (2020) GZ17-6.02 and doxorubicin interact to kill sarcoma cells via autophagy and death receptor signaling. *Front. Oncol.* **10**, 1331, <https://doi.org/10.3389/fonc.2020.01331>
- JGeorge D Demetri 1, Scott Antonia, Robert S Benjamin, Marilyn M Bui, Ephraim S Casper, et al. (2010) Soft tissue sarcoma. *J Natl Compr Canc Netw* **8** (6), 630–674
- Liu, W., Yu, Z., Tang, H., Wang, X., Zhang, B., Zhao, J. et al. (2020) Silencing KIF18B enhances radiosensitivity: identification of a promising therapeutic target in sarcoma. *EBioMedicine* **61**, 103056, <https://doi.org/10.1016/j.ebiom.2020.103056>
- Nudel, K., Massari, P. and Genco, C. (2015) Neisseria gonorrhoeae Modulates Cell Death in Human Endocervical Epithelial Cells through Export of Exosome-Associated cIAP2. *Infect. Immun.* **83**, 3410–3417, <https://doi.org/10.1128/IAI.00732-15>
- Peng, F., Chang, W., Sun, Q., Xu, X., Xie, J., Qiu, H. et al. (2020) HGF alleviates septic endothelial injury by inhibiting pyroptosis via the mTOR signalling pathway. *Respir. Res.* **21**, 215, <https://doi.org/10.1186/s12931-020-01480-3>
- Miao, E., Rajan, J. and Aderem, A. (2011) Caspase-1-induced pyroptotic cell death. *Immunol. Rev.* **243**, 206–214, <https://doi.org/10.1111/j.1600-065X.2011.01044.x>
- Broz, P., Pelegrin, P. and Shao, F. (2020) The gasdermins, a protein family executing cell death and inflammation. *Nat. Rev. Immunol.* **20**, 143–157, <https://doi.org/10.1038/s41577-019-0228-2>
- Ding, J., Wang, K., Liu, W., She, Y., Sun, Q., Shi, J. et al. (2016) Pore-forming activity and structural autoinhibition of the gasdermin family. *Nature* **535**, 111–116, <https://doi.org/10.1038/nature18590>
- Feng, S., Fox, D. and Man, S. (2018) Mechanisms of gasdermin family members in inflammasome signaling and cell death. *J. Mol. Biol.* **430**, 3068–3080, <https://doi.org/10.1016/j.jmb.2018.07.002>
- Zhang, Y., Chen, X., Gueydan, C. and Han, J. (2018) Plasma membrane changes during programmed cell deaths. *Cell Res.* **28**, 9–21, <https://doi.org/10.1038/cr.2017.133>
- Frank, D. and Vince, J. (2019) Pyroptosis versus necroptosis: similarities, differences, and crosstalk. *Cell Death Differ.* **26**, 99–114, <https://doi.org/10.1038/s41418-018-0212-6>
- Kolb, R., Liu, G., Janowski, A., Sutterwala, F. and Zhang, W. (2014) Inflammasomes in cancer: a double-edged sword. *Protein Cell* **5**, 12–20, <https://doi.org/10.1007/s13238-013-0001-4>
- Dupaul-Chicoine, J., Yeretssian, G., Doiron, K., Bergstrom, K., McIntire, C., LeBlanc, P. et al. (2010) Control of intestinal homeostasis, colitis, and colitis-associated colorectal cancer by the inflammatory caspases. *Immunity* **32**, 367–378, <https://doi.org/10.1016/j.immuni.2010.02.012>
- Tang, R., Xu, J., Zhang, B., Liu, J., Liang, C., Hua, J. et al. (2020) Shi S: Ferroptosis, necroptosis, and pyroptosis in anticancer immunity. *J. Hematol. Oncol.* **13**, 110
- Xi, G., Gao, J., Wan, B., Zhan, P., Xu, W., Lv, T. et al. (2019) GSDMD is required for effector CD8 T cell responses to lung cancer cells. *Int. Immunopharmacol.* **74**, 105713, <https://doi.org/10.1016/j.intimp.2019.105713>
- Zhang, Z., Zhang, Y., Xia, S., Kong, Q., Li, S., Liu, X. et al. (2020) Gasdermin E suppresses tumour growth by activating anti-tumour immunity. *Nature* **579**, 415–420, <https://doi.org/10.1038/s41586-020-2071-9>
- Karki, R. and Kanneganti, T. (2019) Diverging inflammasome signals in tumorigenesis and potential targeting. *Nat. Rev. Cancer* **19**, 197–214, <https://doi.org/10.1038/s41568-019-0123-y>
- Xia, X., Wang, X., Cheng, Z., Qin, W., Lei, L., Jiang, J. et al. (2019) The role of pyroptosis in cancer: pro-cancer or pro-“host”? *Cell Death Dis.* **10**, 650, <https://doi.org/10.1038/s41419-019-1883-8>
- Wang, B. and Yin, Q. (2017) AIM2 inflammasome activation and regulation: a structural perspective. *J. Struct. Biol.* **200**, 279–282, <https://doi.org/10.1016/j.jsb.2017.08.001>

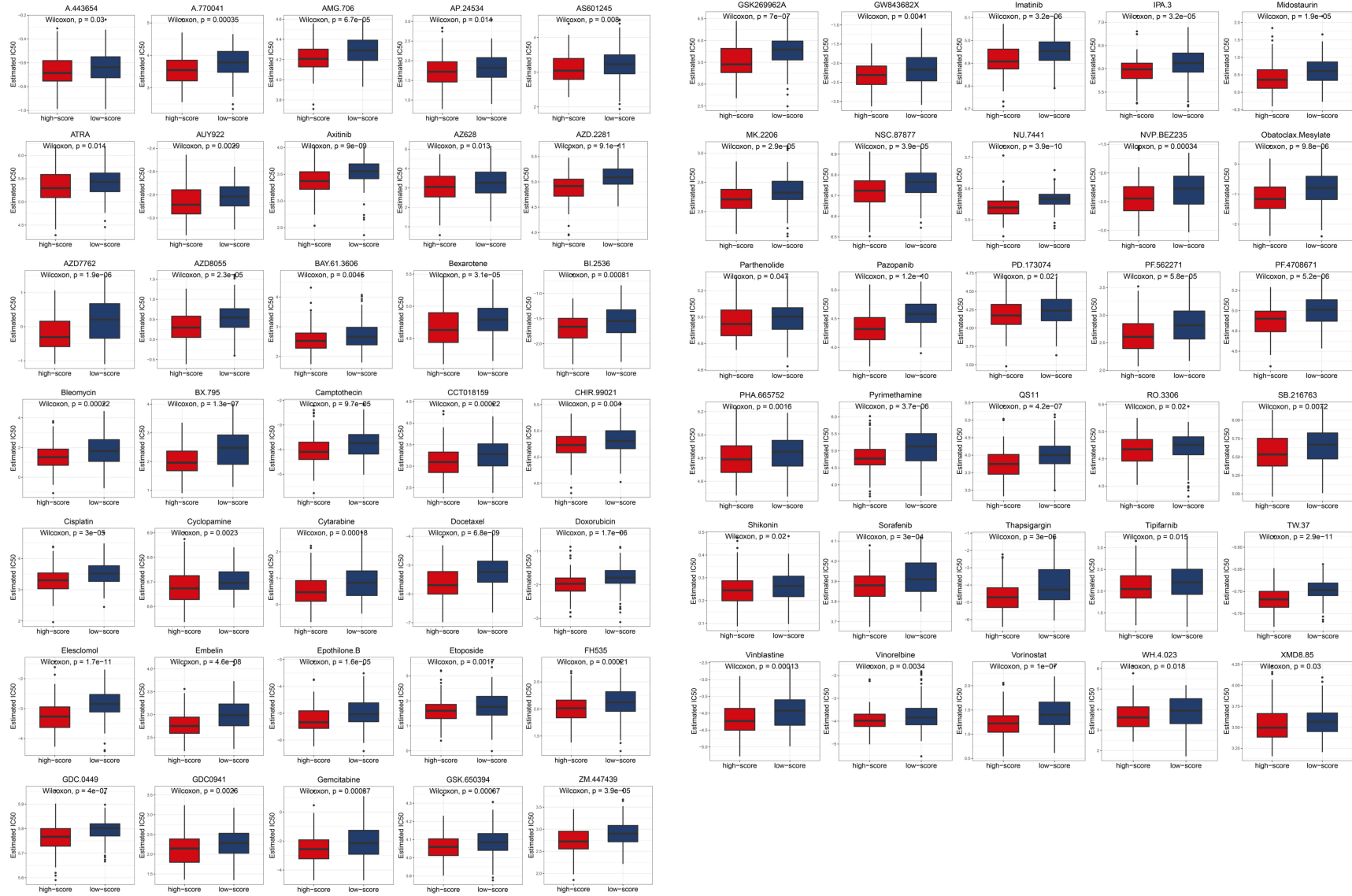
- 22 Ye, Y., Dai, Q. and Qi, H. (2021) A novel defined pyroptosis-related gene signature for predicting the prognosis of ovarian cancer. *Cell Death Discovery* **7**, 71, <https://doi.org/10.1038/s41420-021-00451-x>
- 23 Iorio, F., Knijnenburg, T., Vis, D., Bignell, G., Menden, M., Schubert, M. et al. (2016) A landscape of pharmacogenomic interactions in cancer. *Cell* **166**, 740–754, <https://doi.org/10.1016/j.cell.2016.06.017>
- 24 Yang, W., Soares, J., Greninger, P., Edelman, E., Lightfoot, H., Forbes, S. et al. (2013) Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, D955–D961, <https://doi.org/10.1093/nar/gks111>
- 25 Newman, A., Liu, C., Green, M., Gentles, A., Feng, W., Xu, Y. et al. (2015) Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457, <https://doi.org/10.1038/nmeth.3337>
- 26 Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D. et al. (2017) Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* **18**, 248–262, <https://doi.org/10.1016/j.celrep.2016.12.019>
- 27 Yoshihara, K., Shahmoradgol, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W. et al. (2013) Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612, <https://doi.org/10.1038/ncomms3612>
- 28 Shi, J., Jiang, D., Yang, S., Zhang, X., Wang, J., Liu, Y. et al. (2020) LPAR1, correlated with immune infiltrates, is a potential prognostic biomarker in prostate cancer. *Front. Oncol.* **10**, 846, <https://doi.org/10.3389/fonc.2020.00846>
- 29 Yi, M., Nissley, D., McCormick, F. and Stephens, R. (2020) ssGSEA score-based Ras dependency indexes derived from gene expression data reveal potential Ras addiction mechanisms with possible clinical implications. *Sci. Rep.* **10**, 10258, <https://doi.org/10.1038/s41598-020-66986-8>
- 30 Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. et al. (2017) TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* **77**, e108–e110, <https://doi.org/10.1158/0008-5472.CAN-17-0307>
- 31 Boyle, E., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J. et al. (2004) G0::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* **20**, 3710–3715, <https://doi.org/10.1093/bioinformatics/bth456>
- 32 Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. and Hirakawa, M. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38**, D355–D360, <https://doi.org/10.1093/nar/gkp896>
- 33 Bi, F., Chen, Y. and Yang, Q. (2020) Significance of tumor mutation burden combined with immune infiltrates in the progression and prognosis of ovarian cancer. *Cancer Cell Int.* **20**, 373, <https://doi.org/10.1186/s12935-020-01472-9>
- 34 Xin, Y., Li, S., Jiang, Q., Hu, F., He, Y. and Zhang, J. (2020) Establishment of a jaw fibrosarcoma patient-derived xenograft and evaluation of the tumor suppression efficacy of plumbagin against jaw fibrosarcoma. *Front. Oncol.* **10**, 1479, <https://doi.org/10.3389/fonc.2020.01479>
- 35 McNair, J., Damron, T., Landas, S., Ambrose, J. and Shrimpton, A. (2001) Inheritance of osteosarcoma and Paget's disease of bone: a familial loss of heterozygosity study. *J. Mol. Diagn.* **3**, 171–177, [https://doi.org/10.1016/S1525-1578\(10\)60669-1](https://doi.org/10.1016/S1525-1578(10)60669-1)
- 36 Zhai, Z., Hu, J., You, Y., Yang, X. and Song, Z. (2020) Tumor with slow-developing and recurring lumps of the external ear with skull base and lung metastasis: a case report. *Oncol. Lett.* **20**, 1567–1572, <https://doi.org/10.3892/ol.2020.11724>
- 37 Sollberger, G. (2022) Approaching neutrophil pyroptosis. *J. Mol. Biol.* **434** (4, ), 167335
- 38 Orwoll, E., Wiedrick, J., Nielson, C., Jacobs, J., Baker, E., Piehowski, P. et al. (2020) Proteomic assessment of serum biomarkers of longevity in older men. *Aging Cell* **19**, e13253, <https://doi.org/10.1111/accel.13253>
- 39 Olin, J., Di Narzo, A., d'Escamard, V., Kadian-Dodov, D., Cheng, H., Georges, A. et al. (2020) A plasma proteogenomic signature for fibromuscular dysplasia. *Cardiovasc. Res.* **116**, 63–77, <https://doi.org/10.1093/cvr/cvz219>
- 40 Brazert, M., Kranc, W., Celichowski, P., Ożegowska, K., Budna-Tukan, J., Jeseta, M. et al. (2019) [Corrigendum] Novel markers of human ovarian granulosa cell differentiation toward osteoblast lineage: a microarray approach. *Mol. Med. Rep.* **20** (5), 4403–4414, <https://doi.org/10.3892/mmr.2021.12308>
- 41 Zhang, S., Zhao, X., Xu, S., Yuan, J., Si, Z., Yang, Y. et al. (2019) Low free triiodothyronine predicts worsen neurological outcome of patients with acute ischemic stroke: a retrospective study with bioinformatics analysis. *BMC Neurol.* **19**, 272, <https://doi.org/10.1186/s12883-019-1509-x>
- 42 Rios, J., Denton, K., Yu, H., Manickam, K., Garner, S., Russell, J. et al. (2021) Saturation mutagenesis defines novel mouse models of severe spine deformity. *Dis. Models Mechanisms* **14** (6), <https://doi.org/10.1242/dmm.048901>
- 43 Brilha, S., Ong, C., Weksler, B., Romero, N., Couraud, P.O. and Friedland, J.S. (2017) Matrix metalloproteinase-9 activity and a downregulated Hedgehog pathway impair blood-brain barrier function in an in vitro model of CNS tuberculosis. *Sci. Rep.* **7**, 16031, <https://doi.org/10.1038/s41598-017-16250-3>
- 44 Li, J., Wang, H., Li, Z., Zhang, C., Zhang, C., Li, C. et al. (2020) A 5-gene signature is closely related to tumor immune microenvironment and predicts the prognosis of patients with non-small cell lung cancer. *BioMed Res. Int.* **2020**, 2147397
- 45 Gay, O., Gilquin, B., Assard, N., Stuelsatz, P., Delphin, C., Lachuer, J. et al. (2016) Refilins are short-lived actin-bundling proteins that regulate lamellipodium protrusion dynamics. *Biol. Open* **5**, 1351–1361, <https://doi.org/10.1242/bio.019588>

A





B



**Supplementary Figure :** **IC<sub>50</sub> values of all 79 drugs.** A. 14 drugs with low drug sensitivity of the risk group. B. 65 drugs with high drug sensitivity of the risk group.

pmid	TF	characteristics	gene	regulation_type	hallmark	original_text	title	Motif	Strand	Location	p-value	Match, Sequence	Gene
28643791	PAX7	high expression; targeted by	EWSR1 fusion protein	positive	Ewing sarcoma	Here, using analyses of published whole-genome gene expression microarray data, we identify PAX7 as a gene with significantly increased expression in Ewing sarcoma in comparison to CIC-DUX4 round cell sarcoma. Exploring the mechanism of PAX7 expression in Ewing sarcoma using curated RNA- and ChIP-sequencing data, we demonstrate that the EWSR1 fusion protein is required for PAX7 expression in Ewing sarcoma and identify a candidate EWSR1-FLI1-bound PAX7 enhancer that coincides with both a consensus GCGAA repeat-containing binding site and a peak of regulatory H3K27 acetylation.	EWSR1 fusion proteins moderate PAX7 expression in Ewing sarcoma.	V_PAX7_01 G_M01338	-	102557208-102557224	5.00E-06	TCAGAAATAGAAAAA	LRRC17
28008375	SMAD3	targeted by	EWSR1/FLI1	negative	N/A	SMAD3 were up-regulated and FLI1, MYB, E2F1, ETS2, WT1 were down-regulated with more than half of their targets were down-regulated after EWSR1/FLI1 knockdown.	Knockdown of EWSR1/FLI1 expression alters the transcriptome of Ewing sarcoma cells in vitro.	V_SMAD3_02 G_M011801	-	127348889-127348901	7.00E-06	AGCGACAGACAATC	PODNL2
28008375	MYB	targeted by	EWSR1/FLI1	positive	N/A	SMAD3 were up-regulated and FLI1, MYB, E2F1, ETS2, WT1 were down-regulated with more than half of their targets were down-regulated after EWSR1/FLI1 knockdown.	Knockdown of EWSR1/FLI1 expression alters the transcriptome of Ewing sarcoma cells in vitro.	V_MYB_03 G_M00773	+	102557023-102557033	2.00E-06	TTGTCTCAGATG	LRRC17
29773426	SOX2	correlate with	H3K27me3	N/A	synovial sarcoma	H3K27me3 immunohistochemistry of the synovial sarcoma cases revealed a high statistically significant correlation between SOX2 and H3K27me3 expression (p < .0.0005; Chi square test). Six SOX2 positive synovial sarcoma cases were analyzed by FISH using a SOX2/CEN3 dual color FISH probe. None of these cases revealed an amplification of the SOX2 gene.	Stem cell transcription factor SOX2 in synovial sarcoma and other soft tissue tumors.	V_SOX2_06 G_M01272	+	102553198-102553213	1.00E-05	ATTCTCTTTGTTTGTTG	LRRC17
7828148	CHOP	amplification; high expression	N/A	N/A	N/A	We have now found CHOP amplification in two sarcoma cell lines with previously reported amplification of the nearby GLI1 gene. High constitutive expression levels of CHOP were observed in tumors with gene amplification, but also in some other samples. MDM2 and CHOP were co-amplified in two of these, whereas the two osteosarcoma lines had amplified CHOP but not MDM2. CHOP was amplified in both cell lines with GLI1 amplification, and MDM2 only in one.	The proto-oncogene EWSR1/FLI1 is involved in growth arrest and DNA damage response, is amplified in a subset of human sarcomas.	V_CHOP_01 G_M00249	+	102557059-102557071	7.00E-06	TTGTGCAATACCTC	LRRC17
10574952	ATF1	high expression	N/A	N/A	cell viability; synovial sarcoma	The level of EWS/ATF1 expression was found to be significantly higher in primary tumor tissue than in SLCCS-1 cells or in 203T cells following induction of an EWS/ATF1 expression vector. These studies demonstrate a direct role for the EWS/ATF1 fusion protein in maintaining tumor cell viability of Clear Cell sarcoma and indicate that intracellular antibodies may be used to achieve a phenotypic knockout of tumor-related proteins as a method to explore their function.	Tumor cell viability in clear cell sarcoma requires DNA binding activity of the EWS/ATF1 fusion protein.	V_ATF1_06 G_M00691	+	53342311-53342321	1.00E-06	CCCTGACGTCAT	HLF
10574952	ATF1	high expression	N/A	N/A	cell viability; synovial sarcoma	The level of EWS/ATF1 expression was found to be significantly higher in primary tumor tissue than in SLCCS-1 cells or in 203T cells following induction of an EWS/ATF1 expression vector. These studies demonstrate a direct role for the EWS/ATF1 fusion protein in maintaining tumor cell viability of Clear Cell sarcoma and indicate that intracellular antibodies may be used to achieve a phenotypic knockout of tumor-related proteins as a method to explore their function.	Tumor cell viability in clear cell sarcoma requires DNA binding activity of the EWS/ATF1 fusion protein.	V_ATF1_04 G_M02618	+	102553731-102553741	7.00E-06	CTATGACAAAGAAA	LRRC17
20514024	EGR1	regulate	PTEN	positive; promoter binding	cell death; synovial sarcoma	Moreover, we find that under these conditions phosphatase and tensin homolog deleted in chromosome 10 (PTEN) is upregulated and this occurs through direct binding of EGR1 to an element upstream of the PTEN promoter. Using a combination of gain- and loss-of-function approaches, we show that EGR1 modulation of PTEN contributes to HDAC inhibitor-induced apoptosis in synovial sarcoma. Finally, restoration of EGR1 or PTEN expression is sufficient to induce synovial sarcoma cell death.	EGR1 reactivation by histone deacetylase inhibitors promotes synovial sarcoma cell death through the PTEN tumor suppressor.	V_EGR1_06 G_M01873	-	53342277-53342286	4.00E-06	CGCGGGGCGG	HLF
20514024	EGR1	regulate	PTEN	positive; promoter binding	cell death; synovial sarcoma	Moreover, we find that under these conditions phosphatase and tensin homolog deleted in chromosome 10 (PTEN) is upregulated and this occurs through direct binding of EGR1 to an element upstream of the PTEN promoter. Using a combination of gain- and loss-of-function approaches, we show that EGR1 modulation of PTEN contributes to HDAC inhibitor-induced apoptosis in synovial sarcoma. Finally, restoration of EGR1 or PTEN expression is sufficient to induce synovial sarcoma cell death.	EGR1 reactivation by histone deacetylase inhibitors promotes synovial sarcoma cell death through the PTEN tumor suppressor.	V_EGR1_06 G_M02744	+	53342276-53342289	1.00E-06	CCGCGCCCGCAGCG	HLF
24415532	REST	targeted by	EWS-FLI1	positive	tumor growth	Inhibition of EWS-FLI1 using small interfering RNA decreased REST expression in human Ewing sarcoma cells. Inhibition of REST did not affect EWS-FLI-1, but significantly suppressed tumor growth in vivo, reduced the tumor vessel pericyte markers - smooth muscle actin (SMA) and desmin, increased hypoxia and apoptosis in tumor tissues, and decreased the expression of delta-like ligand 4 (DLL4) and Hes1.	EWS-FLI1 regulates the neuronal repressor gene REST, which controls Ewing sarcoma growth and vascular morphology.	V_REST_01 G_M01256	+	35182303-35182324	1.00E-06	CCTGGGGGCTCTCTCTGCTGCT	SCUBE3
24415532	REST	targeted by	EWS-FLI1	positive	tumor growth	Inhibition of EWS-FLI1 using small interfering RNA decreased REST expression in human Ewing sarcoma cells. Inhibition of REST did not affect EWS-FLI-1, but significantly suppressed tumor growth in vivo, reduced the tumor vessel pericyte markers - smooth muscle actin (SMA) and desmin, increased hypoxia and apoptosis in tumor tissues, and decreased the expression of delta-like ligand 4 (DLL4) and Hes1.	EWS-FLI1 regulates the neuronal repressor gene REST, which controls Ewing sarcoma growth and vascular morphology.	V_REST_02 G_M02256	-	35182307-35182327	7.00E-06	CCAGCAGACCCAGCAGACGCCG	SCUBE3
24415532	REST	regulate	SMA; desmin; DLL4; Hes1	positive	N/A	Inhibition of EWS-FLI1 using small interfering RNA decreased REST expression in human Ewing sarcoma cells. Inhibition of REST did not affect EWS-FLI-1, but significantly suppressed tumor growth in vivo, reduced the tumor vessel pericyte markers - smooth muscle actin (SMA) and desmin, increased hypoxia and apoptosis in tumor tissues, and decreased the expression of delta-like ligand 4 (DLL4) and Hes1.	EWS-FLI1 regulates the neuronal repressor gene REST, which controls Ewing sarcoma growth and vascular morphology.	V_REST_01 G_M01256	+	35182303-35182324	1.00E-06	CCTGGGGGCTCTCTCTGCTGCT	SCUBE3
24415532	REST	regulate	SMA; desmin; DLL4; Hes1	positive	N/A	Inhibition of EWS-FLI1 using small interfering RNA decreased REST expression in human Ewing sarcoma cells. Inhibition of REST did not affect EWS-FLI-1, but significantly suppressed tumor growth in vivo, reduced the tumor vessel pericyte markers - smooth muscle actin (SMA) and desmin, increased hypoxia and apoptosis in tumor tissues, and decreased the expression of delta-like ligand 4 (DLL4) and Hes1.	EWS-FLI1 regulates the neuronal repressor gene REST, which controls Ewing sarcoma growth and vascular morphology.	V_REST_02 G_M02256	-	35182307-35182327	7.00E-06	CCAGCAGACCCAGCAGACGCCG	SCUBE3
24069508	REST	interact with	EWS	N/A	neuronal phenotype development; oncogenic transformation; Ewing sarcoma	Co-immunoprecipitation analysis demonstrated that EWS interacts directly with REST. Genome-wide binding analysis showed that EWS binds chromatin at or near NRSE. Furthermore, functional studies revealed that both EWS and REST inhibit neuronal phenotype development and oncogenic transformation in Ewing sarcoma cells.	EWS and REST: Silencing Transcription Factor Inhibit Neuronal Phenotype Development and Oncogenic Transformation in Ewing Sarcoma.	V_REST_01 G_M01256	+	35182303-35182324	1.00E-06	CCTGGGGGCTCTCTCTGCTGCT	SCUBE3
24069508	REST	interact with	EWS	N/A	neuronal phenotype development; oncogenic transformation; Ewing sarcoma	Co-immunoprecipitation analysis demonstrated that EWS interacts directly with REST. Genome-wide binding analysis showed that EWS binds chromatin at or near NRSE. Furthermore, functional studies revealed that both EWS and REST inhibit neuronal phenotype development and oncogenic transformation in Ewing sarcoma cells.	EWS and REST: Silencing Transcription Factor Inhibit Neuronal Phenotype Development and Oncogenic Transformation in Ewing Sarcoma.	V_REST_02 G_M02256	-	35182307-35182327	7.00E-06	CCAGCAGACCCAGCAGACGCCG	SCUBE3
24043308	GLI1	regulate	KRT17	positive	cellular adhesion; oncogenic transformation	In this work, we identify keratin 17 (KRT17) as a direct downstream target gene upregulated by GLI1. We demonstrate that KRT17 regulates cellular adhesion by activating AKT/PKB (protein kinase B) signaling. In addition, KRT17 is necessary for oncogenic transformation in Ewing sarcoma and accounts for much of the GLI1-mediated transformation function but via a mechanism independent of AKT signaling.	A novel role for keratin 17 in coordinating oncogenic transformation and cellular adhesion in Ewing sarcoma.	V_GLI1_01 G_M01702	-	102556953-102556963	6.00E-06	GGCGACCCAAAG	LRRC17
11973649	ETS1	targeted by	PARP-1	negative	N/A	Previously, we cloned the PARP gene promoter region from EWS cells, showed that it contains multiple ETS-binding sites and demonstrated a positive regulation of PARP by ETS1. Results show that stable down-regulation of ETS1 increases the resistance of EWS cells to various genotoxic agents, whereas down-regulation of EWS/FLI-1 has pro-apoptotic effects. Because down-regulation EWS/FLI-1 does not dramatically change PARP levels, these results suggest a direct effect for EWS/FLI-1 in the DNA damage response of EWS cells.	Differential regulation of ETS1 response to DNA damage in Ewing's sarcoma cells and EWS/FLI-1, ETS1 and EWS/FLI-1.	V_ETS1_B_0 G_M00339	-	102549332-102549346	0	CCAGGAAGTGTTATC	LRRC17
30219084	ATF1	EWSR1-ATF1 fusion	N/A	N/A	N/A	These cases most likely had EWSR1-ATF1 and EWSR1-CREB1 fusions, respectively. RT-PCR was performed in 8 available cases, including 6 CCSBs and 2 CCSLGTs. All CCSBs showed EWSR1-ATF1 fusions. Among the 2 CCSLGT cases, one had EWSR1-ATF1 fusion and the other had EWSR1-CREB1 fusion.	Detection of specific gene rearrangements by fluorescence in situ hybridization in 10 cases of clear cell sarcoma of soft tissue and 6 cases of clear cell sarcoma-like gastrointestinal tumor.	V_ATF1_06 G_M00691	+	53342311-53342321	1.00E-06	CCCTGACGTCAT	HLF
30219084	ATF1	EWSR1-ATF1 fusion	N/A	N/A	N/A	These cases most likely had EWSR1-ATF1 and EWSR1-CREB1 fusions, respectively. RT-PCR was performed in 8 available cases, including 6 CCSBs and 2 CCSLGTs. All CCSBs showed EWSR1-ATF1 fusions. Among the 2 CCSLGT cases, one had EWSR1-ATF1 fusion and the other had EWSR1-CREB1 fusion.	Detection of specific gene rearrangements by fluorescence in situ hybridization in 10 cases of clear cell sarcoma of soft tissue and 6 cases of clear cell sarcoma-like gastrointestinal tumor.	V_ATF1_04 G_M02842	+	102553731-102553744	7.00E-06	CTATGACAAAGAAA	LRRC17
16463269	STAT3	N/A	N/A	N/A	Ewing sarcoma family of tumours	In conclusion, STAT3 activation is present in approximately half of ESFT and correlates with clinical features. The role of STAT3 activation in ESFT pathogenesis seems to be independent of the type of EWS/ETS translocation.	STAT3 is activated in a subset of the Ewing sarcoma family of tumours.	V_STAT3_03 G_M01595	-	127351434-127351449	4.00E-06	CATTCCAGGAAGAAAA	PODNL2
23185447	ETS	regulate	KCNK2	negative	Ewing's sarcoma	Conversely, KCNK2 was found underexpressed in ESFT relative to AHRMS, suggesting that the EWSR1-ETS oncoprotein may have the opposite effect of ERG rearrangements in PCa.	Potential downstream target genes of aberrant ETS transcription factors are differentially affected in Ewing's sarcoma and prostate carcinoma.	V_ETS_B_M00340	-	102549334-102549347	6.00E-06	GGCAGGAAATGTGGTT	LRRC17
23185447	ETS	regulate	KCNK2	negative	Ewing's sarcoma	Conversely, KCNK2 was found underexpressed in ESFT relative to AHRMS, suggesting that the EWSR1-ETS oncoprotein may have the opposite effect of ERG rearrangements in PCa.	Potential downstream target genes of aberrant ETS transcription factors are differentially affected in Ewing's sarcoma and prostate carcinoma.	V_ETS_G4 G_M00771	+	102549334-102549345	0	AACCACTCTGTC	LRRC17
23995784	FOXO1	targeted by	EWS-FLI1	negative; EWS-FLI1-suppressed regulator	decrease in ES tumor growth	In addition to FOXO1 regulation by direct promoter binding of EWS-FLI1, its subcellular localization and activity is regulated by cyclin-dependent kinase 2- and AKT-mediated phosphorylation downstream of EWS-FLI1. Restoration of nuclear FOXO1 expression in ES cells impaired proliferation and significantly reduced clonogenicity. Gene-expression profiling revealed a significant overlap between EWS-FLI1-repressed and FOXO1-activated genes.	Suppression of FOXO1 is responsible for a growth regulatory repressive transcriptional sub-signature of EWS-FLI1 in Ewing sarcoma.	V_FOXO1_G3 G_M01216	-	102557196-102557204	6.00E-06	AAAAACAAAT	LRRC17
23995784	FOXO1	targeted by	EWS-FLI1	negative; EWS-FLI1-suppressed regulator	decrease in ES tumor growth	In addition to FOXO1 regulation by direct promoter binding of EWS-FLI1, its subcellular localization and activity is regulated by cyclin-dependent kinase 2- and AKT-mediated phosphorylation downstream of EWS-FLI1. Restoration of nuclear FOXO1 expression in ES cells impaired proliferation and significantly reduced clonogenicity. Gene-expression profiling revealed a significant overlap between EWS-FLI1-repressed and FOXO1-activated genes.	Suppression of FOXO1 is responsible for a growth regulatory repressive transcriptional sub-signature of EWS-FLI1 in Ewing sarcoma.	V_FOXO1_B G_M00473	-	35185862-35185891	3.00E-06	AAAAACAAAT	SCUBE3

**Supplementary Table :** transcription factor motifs of seven genes.