A Prognostic Model for the Overall Survival of Patients with Ewing's Sarcoma

Xiaofei Feng (✉ fengxf17@lzu.edu.cn)  
Lanzhou University First Affiliated Hospital  https://orcid.org/0000-0003-1322-4102

Yao Ma  
Clinical Laboratory Center, Gansu Provincial Maternity and Child-care Hospital

Hai Lei  
Tanzhou Hospital of Zhongshan City

Wenji Wang  
Lanzhou University First Affiliated Hospital

Research

Keywords: Ewing's sarcoma, WGCNA, prognosis, biomarker

DOI: https://doi.org/10.21203/rs.3.rs-117483/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. 
Read Full License
Abstract

Background: Ewing's sarcoma (ES) is the second most common primary malignant bone tumor. Although the disease has been studied on a molecular basis, its prognosis has not improved. Therefore, the goal of this study is to screen effective biomarkers for predicting the prognosis and progression of ES.

Methods: In this study, the gene expression profile of Ewing's sarcoma was downloaded from public databases. Candidate genes were screened by two methods of weighted gene co-expression network analysis (WGCNA) and differential genes expression analysis, and the Hub gene were determined by the protein–protein interaction (PPI) network. The univariate cox regression and least absolute shrinkage and the selection operator (LASSO) Cox regression were performed on the Hub gene to construct a prognostic model. The receiver operating characteristic (ROC) curve tests the predictive power of the prognostic model. Use the clinical data in the International Cancer Genome Consortium (ICGC) database for verification. Finally, Gene Set Enrichment Analysis (GSEA) was performed to obtain the biological role of prognostic genes in ES.

Results: A total of 542 differentially expressed genes (DEGs) were obtained. Through WGCNA, the obtained turquoise module is significantly related to ES. PPI network analysis of candidate genes identified 10 Hub genes. The univariate cox regression results showed that CACNB1, IDH2, ATP1B4, PRKAG3, STAC3 are risk factors affecting the prognosis of ES. A prognostic model was constructed using IDH2 and PRKAG3, and patients were divided into two risk groups. The survival time of patients in the high-risk group was significantly shorter than that in the low-risk group. The ROC curve confirms that this model has certain accuracy. ICGC cohort verification yielded consistent results. GSEA results showed that IDH2 in the development of ES may promote the progression of the disease through G2M CHECKPOINT, GLYCOLYSIS, MITOTIC SPINDLE, MTORC1 SIGNALING, PEROXISOME, PI3K-AKT-MTOR SIGNALING pathways, while PRKAG3 through MYOGENESIS pathway.

Conclusions: This study shows that IDH2 and PRKAG3 may play important roles in the progression of ES, and can be used as therapeutic targets and prognostic evaluation biomarkers.

Background

ES is an aggressive tumor originating from mesenchymal stem cells, with the highest incidence in adolescents and young adults[1–3]. In the past 30 years, chemotherapy, surgery, and/or radiotherapy have had significant effects on improving the survival of patients[1, 4], but at this stage, even if intensive treatment has been taken to improve the prognosis of patients, the effect was still not significantly improved[5, 6]. ES has a certain background in molecular research, and its therapeutic and prognostic effects are still worth expecting. Studies have shown that ES is characterized by balanced chromosomal translocation, which leads to the expression of fusion oncoproteins. The most common one is EWS/FLI to drive tumor occurrence and development[7, 8]. However, studies have pointed out that targeted inhibitors of EWS/FLI are not clinically feasible[9]. Also, insulin-like growth factor 1 and its receptor play
key roles in promoting the progression of Ewing's sarcoma[10, 11]. However, the results of clinical trials of inhibitors of this pathway have been disappointing[12, 13]. Nevertheless, the results still have important guiding significance for us to research the treatment of the disease on a molecular basis. Therefore, the identification of new genes and pathways related to ES occurrence and patient prognosis is crucial.

In recent years, sequencing technology and bioinformatics have shown indispensable roles in the study of disease molecular mechanisms and specific biomarkers. WGCNA is a systematic biological method that can obtain gene function and gene association from the expression of the whole genome. The identification of genes that promote key roles in disease provides profound insights[14]. This method has been used to explore the molecular pathology of diseases, such as liver cancer, Laryngeal Squamous Cell Carcinoma, etc. [15, 16]. However, there is no report on using the WGCNA method to study Ewing's sarcoma related genes and discover new prognostic markers.

In this study, we constructed a prognostic model in the GSE17679 cohort and verified it in the ICGC cohort. We further performed a functional enrichment analysis of IDH2 and PRKAG3 to explore the underlying mechanism. This study provides an explanation for the pathogenesis and potential molecular biological processes of ES, and the identified genes are expected to become potential biomarkers and targets for diagnosis and treatment.

**Materials And Methods**

**Data source**

In this study, gene expression data sets (GSE12102, GSE17679, and GSE34620) were obtained from the GEO (https://www.ncbi.nlm.nih.gov/geo/), and these data sets are all based on the GPL570 platform. The GSE12102 data set includes 37 cases of ES. The GSE17679 data set contains 88 cases of ES tissue and 18 cases of non-tumor tissue, and obtains the survival information of 88 patients. The GSE34620 data set contains 117 cases of ES tissue. The gene expression profile and survival information of 57 patients with ES were obtained from the ICGC (https://icgc.org/).

**Data processing and differential expression analysis**

Use the affy package to read in the raw data of the three data sets for background processing, quality control, and normalization processing[17]. The probe is annotated through the platform annotation file. Combine the three data sets, use the Sva package to remove the batch effects between the three data sets[18], and then use the limma package to analyze differentially expressed genes[19], the false discovery rate (FDR) <0.05 and |log 2 FC| > 3 is used as the cut-off criterion.

**Weighted gene co-expression network analysis**

Select genes with mutation rate> 15% from the expression profile data for analysis. Use WGCNA R software package to cluster analysis of samples. Calculate the Pearson correlation coefficient of any two genes, and determine the soft threshold ability $\beta$ value to screen the co-expression module. To test
whether the $\beta$ value conforms to the scale-free network, the logarithm of nodes with k connectivity ($\log(k)$) should be negatively related to the logarithm of the occurrence probability of a specific node ($\log(P(k))$), and the correlation coefficient should be greater than 0.85. The number of genes in the gene network module is set to at least 50.

**ES candidate genes and their function analysis**

Select the module with the highest correlation with the disease, take $\text{cor.geneModuleMembership} > 0.8$ (correlation between genes and certain clinical phenotypes) as the cut-off criterion, and the overlapping genes between the acquired genes and DEGs as ES Related candidate genes. In order to clarify the potential biological process of candidate genes, the candidate genes were analyzed by GO enrichment and KEGG pathways through clusterProfiler software package[20].

**Construction of PPI and selection of Hub gene**

Search tool used to search for interacting genes (STRING) Online tools can be used to predict the PPI and construct a PPI network of genes with a confidence level of $\geq 0.7$[21]. The Cytoscape software visualizes the PPI network[22], and the cytoHubba plug-in obtains the top 10 genes, which are regarded as Hub genes.

**Prognosis analysis**

In order to verify the prognostic risk of the hub gene, the hub gene was analyzed by univariate Cox regression, LASSO COX regression, multivariate Cox regression and Kaplan-Meier survival curve. Validation is done through the ICGC database. The above analysis is through the R language survival package[23], the survminer package (https://CRAN.R-project.org/package=survminer), and the glmnet package[24]. $P<0.05$ was considered statistically significant.

**Gene Set Enrichment Analysis**

In order to explore the biological processes involved in prognostic genes, according to the median value of each gene expression, it is divided into two groups for GSEA analysis[25], and the GSEA software (version 4.1) is used for analysis, MsigDB gene set “h.all.v7.2.symbols.gmt” as a reference[26], the number of random combinations is set to 1000, and $p<0.05$ is considered statistically significant.

**Results**

**Differentially expressed genes in ES**

After preprocessing the data and removing the batch effect, the principal component analysis (PCA) method was used to display the results before and after processing (Figure 1A, B). The results show that the distribution pattern between tumor samples and normal samples is different, which can be used for
further analysis. Compared with normal tissues, tumor tissues have 119 up-regulated genes and 423 down-regulated genes. The volcano map (Figure 1C) and heat map (Figure 1D) show the DEGs.

Figure 1. Data processing and DEGs identification. (A) before batch effect removal; (B) after batch effect removal; (C) Volcanogram of DEGs; (D) Heat map of DEGs;

**Construction of co-expression network and screening of candidate genes**

Cluster analysis included 18 normal tissues and 242 ES samples, did not exclude any samples (Figure 2A). Choose $\beta = 3$ as the optimal soft threshold (Figure 2B, C), and scale-free R2 = 0.88 (Figure 2D, E) to ensure a scale-free network. 5 modules were identified, namely turquoise, blue, brown, green, and yellow modules (Figure 2F). We mapped the correlation between the module and the clinic. Among these modules, the turquoise module has the highest correlation with ES (correlation coefficient = -0.97, p<0.001) (Figure 2G). Select the genes with cor.geneModuleMembership> 0.8 in the turquoise module, there are 381 in total. Then take the intersection with the DEGs. There are 305 overlapping genes as candidate genes for ES (Figure 2H).

Figure 2. Construction of weighted co-expression network and screening of candidate genes. (A) Clustering dendrogram of 250 samples; (B) Determination of soft thresholds $\beta=3$; (C) Average connectivity of soft threshold powers; (D) When $\beta = 3$, the histogram of the connection distribution; (E) When $\beta = 3$, Check the scale-free topology. (F) The Cluster dendrogram of co-expression network modules was ordered by a hierarchical clustering of genes based on the 1-TOM matrix; (G) Heatmap of the correlation between module eigengenes and clinical traits of ES; (H) Venn diagram of candidate genes.

**Functional enrichment analysis of candidate genes**

In order to further explore the function of the candidate genes, we conducted Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. The GO analysis results show (Figure 3A) that candidate genes include muscle system process, muscle contraction, muscle organ development, muscle cell differentiation and other aspects in the biological process (BP), and the cellular component (CC) includes contractile fiber, myofibril, sarcomere, I band, etc., molecular function (MF) includes actin binding, actin filament binding, structural constituent of muscle, etc.. The results of the KEGG enrichment pathway show (Figure 3B) that candidate genes function mainly through the following pathways, such as Calcium signaling pathway, Oxytocin signaling pathway, cGMP−PKG signaling pathway, cAMP signaling pathway, Glucagon signaling pathway, etc.

Figure 3. GO and KEGG enrichment analyses. (A) Top 10 GO term enrichment analysis of candidate genes, including BP, CC, and MF; (B)KEGG pathway enrichment analysis of candidate genes.

**PPI network construction and Hub gene identification**
The PPI network between candidate genes is established using the STRING database. The clustering coefficient algorithm of the CytoHubba plug-in is used to select the top 10 genes from the PPI network as Hub genes, including STAC3, ATP1B4, TAS2R38, SSTR4, SORBS1, TNNC1, FBXO27, CACNB1, IDH2, PRKAG3 (Figure 4).

Figure 4. PPI network and hub genes.

**The GSE17679 cohort is used to build a prognostic model**

Using the clinical data of GSE17679, the univariate cox regression analysis was performed on 10 Hub genes (Figure 5A), and 5 genes were obtained through the threshold of $P<0.05$. 5 genes were subjected to LASSO Cox regression analysis, and the penalty parameter lambda was selected by the cross-validation method to obtain 2 relatively independent characteristic genes for constructing a prognostic model (Figure 5B, C). The results showed that high expression of IDH2 and high expression of PRKAG3 are associated with poor prognosis of ES. Kaplan-Meier curve also showed that the high expression of IDH2 ($P<0.001$) and PRKAG3 ($P<0.004$) were associated with poor prognosis (Figure 6A, B). Risk Score = $(0.749 \times \text{IDH2}) + (1.499 \times \text{PRKAG3})$, according to the median value of risk score, patients are divided into high-risk group and low-risk group. PCA analysis in the GSE17679 cohort showed that the distribution patterns of patients in different risk groups were different (Figure 6C). The Kaplan-Meier curve shows that the higher-risk group has a worse prognosis than the low-risk group (log-rank $p<0.001$) (Figure 6D). The ROC was used to curve to evaluate the performance of the risk score, and the area under the curve (AUC) for 3 years, 5 years, and 10 years are 0.811, 0.838, and 0.844, respectively (Figure 6E).

Figure 5. Identification of prognostic genes in ES patients. (A) Univariate Cox regression analysis of 10 Hub genes; (B) LASSO coefficients; (C) LASSO regression with tenfold cross-validation obtained 14 prognostic genes using minimum lambda value.

Figure 6. Survival curves of prognostic genes and prognostic models in GEO cohort. (A) OS curves of IDH2; (B) OS curves of PRKAG3; (C) PCA plot of the GEO cohort; (D) Kaplan-Meier curves for the OS of patients in the high-risk group and low-risk group in the GEO cohort; (E) AUC of time-dependent ROC curves verified the prognostic performance of the risk score in the GEO cohort.

**Validation of the prognostic model in the ICGC cohort**

The 57 samples from the International Cancer Genome Consortium (ICGC) database were verified, and the results showed that the high expression of IDH2 ($P=0.005$) and PRKAG3 ($P<0.034$) was associated with poor prognosis (Figure 7A, B). In order to test the prognosis model constructed by the GSE17679 cohort, patients in the ICGC cohort were divided into a high-risk group or a low-risk group after the median risk score calculated by the same formula as the GSE17679 cohort. Similar to the results obtained in the GSE17679 cohort, PCA analysis confirmed that the distribution patterns of patients in the high-risk group and the low-risk group were different (Figure 7C). Similarly, the high-risk group had a shorter survival time
than the low-risk group (log-rank p<0.019) (Figure 8D). In addition, the area under the curve (AUC) of the prognostic model at 3, 5, and 10 years are 0.659, 0.662, and 0.607, respectively (Figure 7E).

Figure 7. Validation of prognostic genes and prognostic models in the ICGC cohort. (A) OS curves of IDH2; (B) OS curves of PRKAG3; (C) PCA plot of the ICGC cohort; (D) Kaplan-Meier curves for the OS of patients in the high-risk group and low-risk group in the ICGC cohort; (E) AUC of time-dependent ROC curves verified the prognostic performance of the risk score in the ICGC cohort.

**Analysis of the role of prognostic genes in diseases**

Finally, GSEA is used to further determine the potential biological processes and pathways of IDH2 and PRKAG3. The results show that the high expression of IDH2 may be related to the following pathways, such as G2M CHECKPOINT, GLYCOLYSIS, MITOTIC SPINDLE, MTORC1 SIGNALING, PEROXISOME, PI3K-AKT-MTOR SIGNALING, UV RESPONSE UP (Figure 8A). Highly expressed PRKAG3 is related to the MYOGENESIS pathway (Figure 8B).

Figure 8. GSEA of prognostic genes. (A) IDH2 high expression enrichment pathway; (B) PRKAG3 high expression enrichment pathway.

**Discussion**

ES seriously harms the health of children and young people. Although the treatment of ES has improved in the past few decades, the ability to treat the disease is still limited due to the lack of precise molecular targets for ES. Many studies have shown that genetic abnormal changes promote the occurrence and development of the disease, but the molecular mechanism is still unclear[27]. Therefore, further exploration of molecular markers and potential mechanisms related to the progression and prognosis of ES can help the diagnosis and treatment of the disease. In this study, 542 DEGs between tumor and normal tissues were identified based on 3 ES microarray data of Gene Expression Omnibus (GEO) database. The turquoise module was found to be significantly related to ES, and candidate genes were screened out. Construct a PPI network to detect the interaction between the proteins encoded by candidate genes, and select Hub genes. Univariate COX regression, LASSO COX regression, multivariate COX regression, and Kaplan-Meier curve were used to analyze the survival of the Hub gene. The results showed that the high expression levels of IDH2 and PRKAG3 in ES are closely related to the poor prognosis.

Isocitrate dehydrogenase 2 (IDH2) catalyzes the oxidative decarboxylation of isocitrate to α-ketoglutarate in the mitochondria. Recent advances in cancer genetics have determined that the IDH2 gene is mutated in glioma, acute myeloid leukemia and other cancers[28]. Current studies have shown that inhibitors of IDH2 gene mutations may be clinically effective[29]. However, IDH2 has very few mutations in ES[30, 31]. Wild-type IDH2 is down-regulated in liver cancer and gastric cancer, and low-expression IDH2 has a low five-year survival rate[32, 33]. Studies have found that high expression of IDH2 is associated with poor prognosis in esophageal squamous cell carcinoma[34], colon cancer[35], lung cancer[36], and non-small
cell lung cancer[37]. The clinical correlation between wild-type IDH2 and ES has not been fully studied. Our results show that IDH2 is associated with the development and poor prognosis of ES.

PRKAG3, a protein kinase, is activated by AMP, the non-catalytic subunit of γ3, which may play a role in regulating energy metabolism of skeletal muscle[38, 39]. PRKAG3 and overall breast cancer risk are considered as potential targets for cancer treatment[40, 41]. Another study showed that PRKAG3 is associated with the recurrence of triple-negative breast cancer in the Chinese population[42]. We believe that PRKAG3 promotes the development of ES and serves as a new molecular marker.

Our research has constructed a new prognostic model, which has been well validated in the GSE17679 cohort and the ICGC cohort. IDH2 and PRKAG3 may be potential prognostic molecular markers for poor survival of ES, and provide potential targets for pathogenesis and treatment. In addition, G2M CHECKPOINT, GLYCOLYSIS, MITOTIC SPINDLE, MTORC1 SIGNALING, PEROXISOME, PI3K-AKT-MTOR SIGNALING, etc. may be the key pathways of IDH2 regulation in ES. PRKAG3 may promote the development of ES through the MYOGENESIS pathway. In future research, further experimental verification is needed to prove the biological role of IDH2 and PRKAG3 in ES.

**Conclusion**

In summary, we have studied the gene expression profiles of Ewing's sarcoma in GEO and ICGC databases through bioinformatics methods. We provide a prognostic model that has a good predictive effect on the prognosis of patients with Ewing's sarcoma. The model includes two genes (IDH2 and PRKAG3), and using GESA to find that these genes may play important roles through some signaling pathways. These results may serve as potential prognostic molecular biomarkers in Ewing's sarcoma, and may contribute to the molecular targeted therapy of Ewing's sarcoma.

**Abbreviations**

ES  ewing's sarcoma  
WGCNA  weighted gene co-expression network analysis  
PPI  protein–protein interaction  
LASSO  least absolute shrinkage and the selection operator  
ROC  receiver operating characteristic  
ICGC  International Cancer Genome Consortium  
GSEA
Gene Set Enrichment Analysis
DEGs
differentially expressed genes
STRING
Search tool used to search for interacting genes
PCA
principal component analysis
GO
Gene Ontology
KEGG
Kyoto Encyclopedia of Genes and Genomes
BP
biological process
CC
cellular component
MF
molecular function
AUC
area under the curve
OS
overall survival

Declarations

Ethics declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the present study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding
This study was funded by Gansu Provincial Youth Science and Technology Fund Program (Grant No. 17JR5RA196), The First Hospital of Lanzhou University Project (Grant No. Idyyyn2018-40), Lanzhou Talent Innovation and Entrepreneurship Project (Grant No. 2016-RC-57), Science and Technology Development Project of Chengguan (Grant No. 2017/7/5), and Natural Science Foundation of Gansu Province (Grant No. 1606RJZA126).

Acknowledgments

None.

Contributions

Conception and design: Xiaofei Feng; Administrative support: Wenji Wang; Collection and assembly of data: Xiaofei Feng, Hai Lei; Data analysis and interpretation: Xiaofei Feng, Yao Ma, Hai Lei; Manuscript writing: All authors. All authors read and approved the final manuscript.

References

1. Gaspar N, Hawkins DS, Dirksen U, Lewis IJ, Ferrari S, Le Deley MC, et al. Ewing Sarcoma: Current Management and Future Approaches Through Collaboration. J Clin Oncol. 2015;33:3036–46.
2. Grünewald TGP, Cidre-Aranaz F, Surdez D, Tomazou EM, de Álava E, Kovar H, et al. Ewing sarcoma. Nat Rev Dis Primers. 2018;4:5.
3. Tirode F, Laud-Duval K, Prieur A, Delorme B, Charbord P, Delattre O. Mesenchymal stem cell features of Ewing tumors. Cancer Cell. 2007;11:421–9.
4. Damron TA, Ward WG, Stewart A. Osteosarcoma, chondrosarcoma, and Ewing's sarcoma: National Cancer Data Base Report. Clin Orthop Relat Res. 2007;459:40–7.
5. Womer RB, West DC, Krailo MD, Dickman PS, Pawel BR, Grier HE, et al. Randomized controlled trial of interval-compressed chemotherapy for the treatment of localized Ewing sarcoma: a report from the Children's Oncology Group. J Clin Oncol. 2012;30:4148–54.
6. Grier HE, Krailo MD, Tarbell NJ, Link MP, Fryer CJ, Pritchard DJ, et al. Addition of ifosfamide and etoposide to standard chemotherapy for Ewing's sarcoma and primitive neuroectodermal tumor of bone. N Engl J Med. 2003;348:694–701.
7. Grünewald TG, Bernard V, Gilardi-Hebenstreit P, Raynal V, Surdez D, Aynaud MM, et al. Chimeric EWSR1-FLI1 regulates the Ewing sarcoma susceptibility gene EGR2 via a GGAA microsatellite. Nat Genet. 2015;47:1073–8.
8. Delattre O, Zucman J, Plougastel B, Desmaze C, Melot T, Peter M, et al. Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. Nature. 1992;359:162–5.
9. Wang S, Hwang EE, Guha R, O'Neill AF, Melong N, Veinotte CJ, et al. High-throughput Chemical Screening Identifies Focal Adhesion Kinase and Aurora Kinase B Inhibition as a Synergistic Treatment Combination in Ewing Sarcoma. Clin Cancer Res. 2019;25:4552–66.
10. Prieur A, Tirode F, Cohen P, Delattre O. EWS/FLI-1 silencing and gene profiling of Ewing cells reveal downstream oncogenic pathways and a crucial role for repression of insulin-like growth factor binding protein 3. Mol Cell Biol. 2004;24:7275–83.

11. Herrero-Martín D, Osuna D, Ordóñez JL, Sevillano V, Martins AS, Mackintosh C, et al. Stable interference of EWS-FLI1 in an Ewing sarcoma cell line impairs IGF-1/IGF-1R signalling and reveals TOPK as a new target. Br J Cancer. 2009;101:80–90.

12. Olmos D, Postel-Vinay S, Molife LR, Okuno SH, Schuetze SM, Paccagnella ML, et al. Safety, pharmacokinetics, and preliminary activity of the anti-IGF-1R antibody fitumumab (CP-751,871) in patients with sarcoma and Ewing's sarcoma: a phase 1 expansion cohort study. Lancet Oncol. 2010;11:129–35.

13. Basu B, Olmos D, de Bono JS. Targeting IGF-1R: throwing out the baby with the bathwater? Br J Cancer. 2011;104:1–3.

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

15. Zhang Q, Wang J, Liu M, Zhu Q, Li Q, Xie C, et al. Weighted correlation gene network analysis reveals a new stemness index-related survival model for prognostic prediction in hepatocellular carcinoma. Aging. 2020;12:13502–17.

16. Gong S, Xu M, Zhang Y, Shan Y, Zhang H. The Prognostic Signature and Potential Target Genes of Six Long Non-coding RNA in Laryngeal Squamous Cell Carcinoma. Front Genet. 2020;11:413.

17. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics. 2004;20:307–15.

18. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28:882–3.

19. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.

20. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012;16:284–7.

21. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47:D607-d13.

22. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.

23. Durisová M, Dedík L. SURVIVAL—an integrated software package for survival curve estimation and statistical comparison of survival rates of two groups of patients or experimental animals. Methods Find Exp Clin Pharmacol. 1993;15:535–40.

24. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33:1–22.
25. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545–50.

26. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27:1739–40.

27. Smith R, Owen LA, Trem DJ, Wong JS, Whangbo JS, Golub TR, et al. Expression profiling of EWS/FLI identifies NKX2.2 as a critical target gene in Ewing's sarcoma. Cancer Cell. 2006;9:405–16.

28. Yang H, Ye D, Guan KL, Xiong Y. IDH1 and IDH2 mutations in tumorigenesis: mechanistic insights and clinical perspectives. Clin Cancer Res. 2012;18:5562–71.

29. Yen K, Travins J, Wang F, David MD, Artin E, Straley K, et al. AG-221, a First-in-Class Therapy Targeting Acute Myeloid Leukemia Harboring Oncogenic IDH2 Mutations. Cancer Discov. 2017;7:478–93.

30. Amary MF, Bacsi K, Maggiani F, Damato S, Halai D, Berisha F, et al. IDH1 and IDH2 mutations are frequent events in central chondrosarcoma and central and periosteal chondromas but not in other mesenchymal tumours. J Pathol. 2011;224:334–43.

31. Na KY, Noh BJ, Sung JY, Kim YW, Santini Araujo E, Park YK. IDH Mutation Analysis in Ewing Sarcoma Family Tumors. J Pathol Transl Med. 2015;49:257–61.

32. Tian GY, Zang SF, Wang L, Luo Y, Shi JP, Lou GQ. Isocitrate Dehydrogenase 2 Suppresses the Invasion of Hepatocellular Carcinoma Cells via Matrix Metalloproteinase 9. Cell Physiol Biochem. 2015;37:2405–14.

33. Wu D. Isocitrate dehydrogenase 2 inhibits gastric cancer cell invasion via matrix metalloproteinase 7. Tumour Biol. 2016;37:5225–30.

34. Chen X, Xu W, Wang C, Liu F, Guan S, Sun Y, et al. The clinical significance of isocitrate dehydrogenase 2 in esophageal squamous cell carcinoma. Am J Cancer Res. 2017;7:700–14.

35. Lv Q, Xing S, Li Z, Li J, Gong P, Xu X, et al. Altered expression levels of IDH2 are involved in the development of colon cancer. Exp Ther Med. 2012;4:801–6.

36. Li J, He Y, Tan Z, Lu J, Li L, Song X, et al. Wild-type IDH2 promotes the Warburg effect and tumor growth through HIF1α in lung cancer. Theranostics. 2018;8:4050–61.

37. Li JJ, Li R, Wang W, Zhang B, Song X, Zhang C, et al. IDH2 is a novel diagnostic and prognostic serum biomarker for non-small-cell lung cancer. Mol Oncol. 2018;12:602–10.

38. Milan D, Jeon JT, Looft C, Amarger V, Robic A, Thelander M, et al. A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. Science. 2000;288:1248–51.

39. Barnes BR, Marklund S, Steiler TL, Walter M, Hjälm G, Amarger V, et al. The 5’-AMP-activated protein kinase gamma3 isoform has a key role in carbohydrate and lipid metabolism in glycolytic skeletal muscle. J Biol Chem. 2004;279:38441–7.

40. Rivera Rivera A, Castillo-Pichardo L, Gerena Y, Dharmawardhane S. Anti-Breast Cancer Potential of Quercetin via the Akt/AMPK/Mammalian Target of Rapamycin (mTOR) Signaling Cascade. PLoS
41. Wang S, Huo D, Ogundiran TO, Ojengbede O, Zheng W, Nathanson KL, et al. Association of breast cancer risk and the mTOR pathway in women of African ancestry in 'The Root' Consortium. Carcinogenesis. 2017;38:789–96.

42. Chen LH, Kuo WH, Tsai MH, Chen PC, Hsiao CK, Chuang EY, et al. Identification of prognostic genes for recurrent risk prediction in triple negative breast cancer patients in Taiwan. PLoS One. 2011;6:e28222.

**Figures**
Figure 1

Data processing and DEGs identification. (A) before batch effect removal; (B) after batch effect removal; (C) Volcanogram of DEGs; (D) Heat map of DEGs;
Figure 2

Construction of weighted co-expression network and screening of candidate genes. (A) Clustering dendrogram of 250 samples; (B) Determination of soft thresholds $\beta = 3$; (C) Average connectivity of soft threshold powers; (D) When $\beta = 3$, the histogram of the connection distribution; (E) When $\beta = 3$, Check the scale-free topology. (F) The Cluster dendrogram of co-expression network modules was ordered by a hierarchical clustering of genes based on the 1-TOM matrix; (G) Heatmap of the correlation between module eigengenes and clinical traits of ES; (H) Venn diagram of candidate genes.
Figure 2

Construction of weighted co-expression network and screening of candidate genes. (A) Clustering dendrogram of 250 samples; (B) Determination of soft thresholds $\beta = 3$; (C) Average connectivity of soft threshold powers; (D) When $\beta = 3$, the histogram of the connection distribution; (E) When $\beta = 3$, Check the scale-free topology. (F) The Cluster dendrogram of co-expression network modules was ordered by a hierarchical clustering of genes based on the 1-TOM matrix; (G) Heatmap of the correlation between module eigengenes and clinical traits of ES; (H) Venn diagram of candidate genes.
Figure 3

GO and KEGG enrichment analyses. (A) Top 10 GO term enrichment analysis of candidate genes, including BP, CC, and MF; (B) KEGG pathway enrichment analysis of candidate genes.
Figure 4

PPI network and hub genes.
Figure 4

PPI network and hub genes.
### Figure 5

Identification of prognostic genes in ES patients. (A) Univariate Cox regression analysis of 10 Hub genes; (B) LASSO coefficients; (C) LASSO regression with tenfold cross-validation obtained 14 prognostic genes using minimum lambda value.

| Gene     | p-value | Hazard ratio |
|----------|---------|--------------|
| CACNB1   | <0.001  | 2.217(1.429–3.441) |
| IDH2     | <0.001  | 2.447(1.588–3.771) |
| TAS2R38  | 0.991   | 0.993(0.254–3.886)  |
| SSTR4    | 0.112   | 2.358(0.818–6.795)  |
| TNNC1    | 0.280   | 1.074(0.943–1.223)  |
| SORBS1   | 0.471   | 1.262(0.644–2.590)  |
| ATP1B4   | 0.022   | 3.371(1.195–9.509)  |
| PRKAG3   | <0.001  | 6.163(2.326–16.33) |
| FBXO27   | 0.050   | 2.737(0.999–7.500)  |
| STAC3    | 0.002   | 1.806(1.235–2.639)  |
### Table A

| Gene      | p-value | Hazard ratio         |
|-----------|---------|----------------------|
| CACNB1    | <0.001  | 2.217(1.429–3.441)   |
| IDH2      | <0.001  | 2.447(1.588–3.771)   |
| TAS2R38   | 0.991   | 0.993(0.254–3.886)   |
| SSTR4     | 0.112   | 2.358(0.818–6.795)   |
| TNNC1     | 0.280   | 1.074(0.943–1.223)   |
| SORBS1    | 0.471   | 1.262(0.644–2.590)   |
| ATP1B4    | 0.022   | 3.371(1.195–9.509)   |
| PRKAG3    | <0.001  | 6.163(2.326–16.33)   |
| FBXO27    | 0.050   | 2.737(0.999–7.500)   |
| STAC3     | 0.002   | 1.806(1.235–2.639)   |

### Figure C

Identification of prognostic genes in ES patients. (A) Univariate Cox regression analysis of 10 Hub genes; (B) LASSO coefficients; (C) LASSO regression with tenfold cross-validation obtained 14 prognostic genes using minimum lambda value.
Survival curves of prognostic genes and prognostic models in GEO cohort. (A) OS curves of IDH2; (B) OS curves of PRKAG3; (C) PCA plot of the GEO cohort; (D) Kaplan-Meier curves for the OS of patients in the high-risk group and low-risk group in the GEO cohort; (E) AUC of time-dependent ROC curves verified the prognostic performance of the risk score in the GEO cohort.
Figure 6

Survival curves of prognostic genes and prognostic models in GEO cohort. (A) OS curves of IDH2; (B) OS curves of PRKAG3; (C) PCA plot of the GEO cohort; (D) Kaplan-Meier curves for the OS of patients in the high-risk group and low-risk group in the GEO cohort; (E) AUC of time-dependent ROC curves verified the prognostic performance of the risk score in the GEO cohort.
Figure 7

Validation of prognostic genes and prognostic models in the ICGC cohort. (A) OS curves of IDH2; (B) OS curves of PRKAG3; (C) PCA plot of the ICGC cohort; (D) Kaplan-Meier curves for the OS of patients in the high-risk group and low-risk group in the ICGC cohort; (E) AUC of time-dependent ROC curves verified the prognostic performance of the risk score in the ICGC cohort.
Figure 7
Validation of prognostic genes and prognostic models in the ICGC cohort. (A) OS curves of IDH2; (B) OS curves of PRKAG3; (C) PCA plot of the ICGC cohort; (D) Kaplan-Meier curves for the OS of patients in the high-risk group and low-risk group in the ICGC cohort; (E) AUC of time-dependent ROC curves verified the prognostic performance of the risk score in the ICGC cohort.
Figure 8

GSEA of prognostic genes. (A) IDH2 high expression enrichment pathway; (B) PRKAG3 high expression enrichment pathway.
Figure 8

GSEA of prognostic genes. (A) IDH2 high expression enrichment pathway; (B) PRKAG3 high expression enrichment pathway.