

Development and validation of novel prognostic models for zinc finger proteins-related genes in soft tissue sarcoma

Junqing Li^{1,*}, Quan Zhou^{2,*}, Changsheng Zhang¹, Huimin Zhu¹, Jie Yao^{1,&}, Meng Zhang^{2,&}

¹Minimally Invasive Spinal Surgery Center, Luoyang Orthopedic-Traumatological Hospital of Henan Province (Henan Provincial Orthopedic Hospital), Zhengzhou, China

²Department of Orthopedics, Henan Provincial People's Hospital, Zhengzhou University People's Hospital, Henan University People's Hospital, Zhengzhou, China

*Equal contribution

Correspondence to: Jie Yao, Meng Zhang; **email:** yaojie110120@163.com, <https://orcid.org/0009-0004-4835-9650>; zhangmeng.lh@163.com, <https://orcid.org/0000-0002-1300-935X>

Keywords: soft tissue sarcomas, zinc finger proteins, molecular subtype, immune infiltration, differentially expressed genes

Received: December 1, 2022

Accepted: April 15, 2023

Published: April 26, 2023

Copyright: © 2023 Li et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

As the most common transcriptional regulators, zinc finger proteins (ZNFs) play vital roles in occurrence and progression of malignant tumors. Whereas, information regarding the roles of ZNFs in soft tissue sarcomas (STS) remains scarce. In this study, a comprehensive bioinformatics analysis investigating roles of ZNFs in STS was performed. Initially, we extracted raw datasets of differentially expressed ZNFs from GSE2719. Using a sequence of bioinformatics methods, we then investigated the prognostic significance, function, and molecular subtype of these differentially expressed ZNFs. In addition, CCK8 and plate clone formation assays were used to explore the effect of ZNF141 on STS cells. A total of 110 differentially expressed ZNFs were identified. Nine ZNFs (HLTF, ZNF292, ZNF141, LDB3, PHF14, ZNF322, PDLIM1, NR3C2, and LIMS2) were selected to establish an overall survival (OS) prediction model, and seven ZNFs (ZIC1, ZNF141, ZHX2, ZNF281, ZNHIT2, NR3C2, and LIMS2) were used to develop a progression-free survival (PFS) prediction model. Compared with patients with low-risk in the TCGA training and testing cohorts, as well as the GEO validation cohorts, patients with high-risk had poorer OS and PFS. Using nomograms constructed with the identified ZNFs predicting OS and PFS, we established a clinically useful model. Four distinct molecular subtypes with different prognostic and immune infiltration characteristics were identified. *In vitro* experiments showed that ZNF141 promoted the proliferation and viability of STS cells. In conclusion, ZNF-related models are useful as prognostic biomarkers, suggesting their potentials as therapeutic targets in STS. These findings will enable us to develop novel strategies treating STS, which will potentially improve outcomes of patients with STS.

INTRODUCTION

Derived from embryonic mesoderm, soft tissue sarcoma (STS) is a group of rare and heterogeneous tumors [1]. STS encompasses more than 100 histological and molecular subtypes, and of these many kinds of tumors, undifferentiated polymorphic sarcoma, leiomyosarcoma, dedifferentiated liposarcoma and liposarcoma are the most common [2]. Prognoses of

patients with STS (especially those with advanced STS) remain not so favorable due to the relatively inefficient treatment methods. Specially, five-year overall survival (OS) of patients with progressive STS is less than 20% [3]. Additionally, unlike in other tumors, little progress improving survival of patients with advanced STS has been made despite the fact that radiotherapy, chemotherapy, immunotherapy and targeted therapy have been tried to treat patients with advanced STS [4, 5].

According to previously published studies, a good deal of molecular biomarkers (such as PD1 and PDL1) play key roles in predicting prognosis and determining optimal treatment choices for various tumors [6]. Therefore, identifying novel molecular biomarkers and subsequent elucidation of the specific mechanisms through which these biomarkers affect STS may help us develop novel therapeutic strategies and improve prognosis.

Belonging to a large and diverse family of proteins that own at least one zinc finger domain, zinc finger proteins (ZNFs) have been identified as one of the most abundant proteins in eukaryotic cells [7]. Biological functions of ZNFs have been proven extraordinary and functions of ZNFs include the following: RNA packaging, DNA recognition, regulation of apoptosis, transcriptional activation, lipid binding and protein folding and assembly [8]. According to recent studies, apart from the aforementioned functions, ZNFs also play vital roles in occurrence and progression of malignant tumors [9]. For example, ZFX [10], ZEB1 [11], and ZNF24 [12] have been shown to promote malignant behavior in various cancers. In addition, several ZNFs, such as ZNF545 [13], ZNF331 [14], and ZNF668 [15] have been demonstrated to be able to suppress tumor progression. However, by far, studies on roles played by ZNFs in STS are still scarce. Therefore, a comprehensive study investigating ZNFs and their roles in STS is necessary since it will enable us to better understand the specific mechanisms which promote progression of STS.

Raw data of STS were initially retrieved from TCGA and GEO databases, based on which differentially expressed ZNFs were identified and simultaneous investigation of potential functions and mechanisms of ZNFs was accomplished. Given the finding that some ZNFs could be used as prognostic biomarkers, we then established and validated a prognostic model based on ZNFs. Ultimately, we also used non-negative matrix factorization-based consensus clustering to define the STS subtype based on ZNFs and explored the immune characteristics of the subtypes.

MATERIALS AND METHODS

Data processing

Raw data of 1555 zinc finger genes (Supplementary Table 1) were initially obtained from the HUGO Gene Nomenclature Committee database (<https://www.genenames.org/>). Differentially expressed genes (DEGs) were identified from expression profilings of STS and normal tissue samples obtained from GSE2719 using the “limma” package [16] of R

4.0.0 software. The following two criteria were adopted to select DEGs: adjusted p -values <0.05 and $\log_2|\text{fold change}|$ values >1 . The “ggplot2” package was utilized to create volcano plots while heatmaps of this study were constructed using the “pheatmap” package.

RNA-seq data and clinical information of patients diagnosed with STS were retrieved from TCGA (<https://portal.gdc.cancer.gov/>), GSE21050 [17], and GSE30929 [18]. And we subsequently downloaded data of genomic mutation in STS including both and copy number variation from TCGA database. Then these data regarding genomic mutation were analyzed using the R package ‘Rcirco’ to plot the copy number variation landscape that will enable us to better and more directly observe changes of ZNFs on human chromosomes.

Construction and validation of prognostic models

After excluding patients without sufficient follow-up time, we then randomly divided patients into training cohort or testing cohort. We adopted the survival R package to perform univariate Cox regression analysis to determine factors that were significantly associated with prognosis.

Then factors confirmed by the univariate Cox regression analysis to be significantly associated with prognosis were subsequently incorporated into the LASSO-penalized Cox regression analysis to establish prognostic models. The LASSO algorithm was adopted to select variables and for shrinkage using the “glmnet” R package. Additionally risk score of each individual patient in this study was calculated according to the following formula: Risk score = coef gene1 \times Exp gene1 + coef gene2 \times Exp gene2 + coef genei \times Exp genei. Then each patient was assessed based on his or her median risk score calculated according to the aforementioned formula and divided into the high-risk group or low-risk group. Overall survival (OS) of patients in the high-risk group was compared with that of patients in the low-risk group by performing the log-rank test. And predictive capability of the prognostic model was evaluated by performing receiver operating characteristics (ROC) analysis using the “survivalROC” package. Testing cohort obtained from TCGA database and external cohort retrieved from GEO database (GSE21050) were used as the validation group. In a similar way, we established a predictive model for progression-free survival (PFS) as well. In this case, testing cohort obtained from TCGA database and external cohort retrieved from GEO database (GSE30929) were used as the validation group. Then prognostic significance of risk score and the aforementioned clinical variables was evaluated among patients in the TCGA training cohort and testing cohort

by performing univariate and multivariate Cox regression analyses. Ultimately, we established nomograms using the “rms” package to help us more efficiently predict OS and PFS of patients with STS.

GO enrichment and gene set variation analysis (GSVA)

We conducted GO enrichment analysis utilizing the “clusterProfiler” package of R software. Through a hypergeometric distribution using $P < 0.05$ as the significance threshold, we then performed functional enrichment analyses, results of which were made visualized using the “ggplot2” packages of R software.

Variations of key gene sets were estimated using the GSVA package of R software as an unsupervised, non-parametric method [19]. A gene expression matrix was used as the input for GSVA algorithm. GSVA scores were calculated in a non-parametrical way using the Kolmogorov-Smirnov (KS)-like random walk statistic and a negative value for a particular sample and gene set.

Consensus clustering

Consensus clustering was used to classify CAD cases into different subgroups [20]. Consensus clustering was accomplished using the K-means algorithm with the Spearman distance. Maximum number of clusters was set at 10. Final cluster number was identified by cluster consensus score and the consensus matrix.

Immune scores and tumor infiltrating immune cell (TIIC) analysis

The ESTIMATE algorithm was adopted to calculate immune and stromal scores of the samples from TCGA datasets. Based on the median immune score, patients were assigned into the high-score group or low-score group. OS of high-score group was compared with that of low-score group via log-rank test. TIICs in TCGA datasets were assessed using the CIBERSORT analytical tool (<https://cibersort.stanford.edu/>) [21]. 22 immune cells, abundance ratio matrix was identified at p values less than 0.05.

Cell culture and transfection

A673 and SW982 cell lines were purchased from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China). A673 and SW982 cells were cultured in DMEM (Biological Industries, Shanghai, China) or L-15 (Sigma) containing 10% fetal bovine serum (Biological Industries). ZNF141 siRNA and

negative control siRNA oligonucleotides were designed and synthesized by Biosyntech (Suzhou, China). The sequences of si-1 and si-2 are shown in Supplementary Table 2. The siRNA transfections were performed using Lipofectamine RNAiMAX (Invitrogen, Carlsbad, CA, USA). Then we transfected ZNF141-overexpression-promoting plasmid (GV-ZNF141) and negative control plasmid (GV-Vector) that were designed and synthesized by GeneChem (Shanghai, China) into SW982 cells with Lipofectamine 3000 (Invitrogen, Carlsbad, CA, USA).

Real-time PCR and Western blotting

Following the manufacturer’s instructions, total RNA was retrieved with AG RNAex Pro Reagent (Accurate Biology, Changsha, China) and was reverse-transcribed into cDNA using the Evo M-MLV RT Premix Kit (Accurate Biology). Real-time PCR (RT-PCR) assays were performed using the SYBR Green Premix Pro Taq HS qPCR Kit (Accurate Biology) according to the manufacturer’s protocols. The primer sequences are listed in Supplementary Table 2.

Total protein of STS cells were extracted with RIPA lysis buffer (Beyotime Biotechnology, Shanghai, China), which were then separated with the method of SDS-PAGE. Then the proteins separated by SDS-PAGE were transferred onto PVDF membranes (Merck Millipore, Billerica, MA, USA). After being loaded with transferred proteins, these PVDF membranes were soaked in 5% BSA solution for at least one hour at room temperature, which was immediately followed by incubation with different primary antibodies at 4° C overnight. On the second day, the PVDF membranes were incubated with secondary antibodies for one hour at room temperature after washing three times. After washing with TBST solution, the PVDF membranes were visualized using a BeyoECL Plus Kit (Beyotime Biotechnology). Anti-GAPDH (10494-1-AP) were purchased from Proteintech (Wuhan, China) while anti-ZNF141 antibody (TA339080) was produced by OriGene Technologies (Wuxi, China).

Cell proliferation and plate clone formation

Cell proliferation was detected using the Cell Counting Kit-8 (CCK8, Beyotime Biotechnology) according to the manufacturer’s instructions. The cells were cultured in a 96-well plate (2000 cells/well) at 37° C for 0, 24, 48, and 72 h. For the plate clone formation assay, 1000 cells were seeded into 6-well plates and cultivated for two weeks. The number of colonies with >100 cells was counted under a light microscope, and the colonies were visualized.

Statistical analysis

R software was used for most bioinformatics and statistical analyses, including RNA-seq data normalization and transformation, DEG analysis, survival analyses, ROC analysis, CIBERSORT, ESTIMATE, GSVA, and Spearman rank correlation analysis. For the *in vitro* experiments, all quantitative data are presented as mean \pm standard deviation of three independent experiments. Differences between the three groups were analyzed with one-way ANOVA using GraphPad Prism 8.0 (GraphPad, La Jolla, CA, USA). Statistical significance was set at $p < 0.05$.

Data availability statement

All datasets presented in this study are included in the article/Supplementary Material.

RESULTS

Differentially expressed ZNFs in STS

The study design is illustrated in Figure 1. Differentially expressed ZNFs between normal tissues and tumor tissues were identified from GSE2719 using the “limma” package in R software. The heatmap demonstrating expression of ZNFs was presented in Figure 2A. A total of differentially expressed 110 ZNFs were identified with 82 ones up-regulated and 28 down-regulated (Figure 2B and Supplementary Table 3).

Construction and validation of prognostic models

Patients without sufficient follow-up time length were excluded from this study and remaining patients were

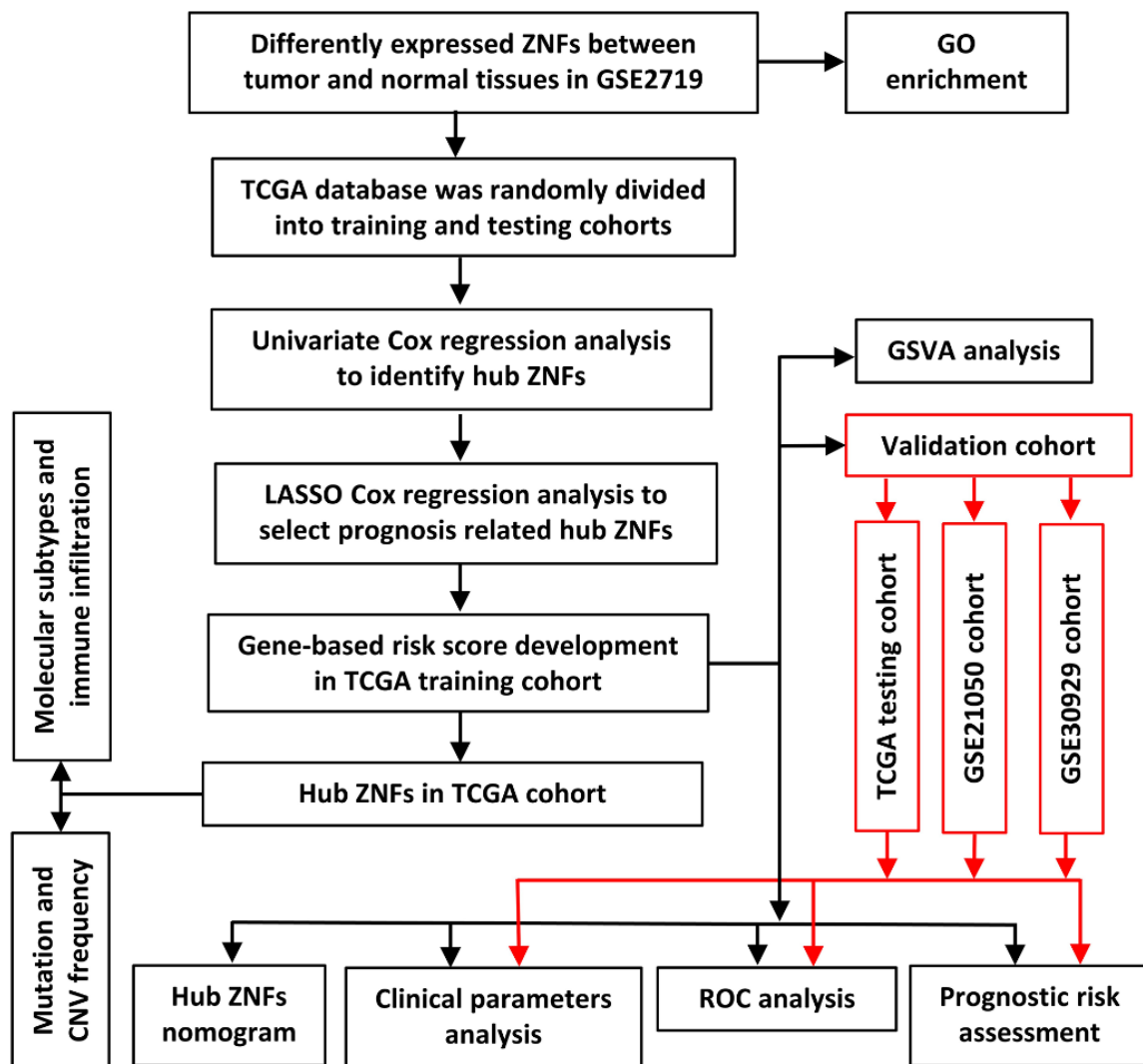
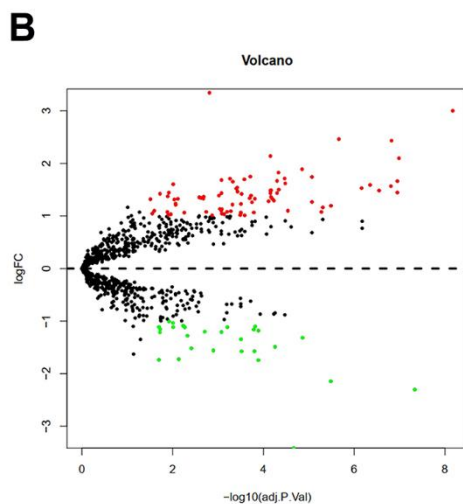
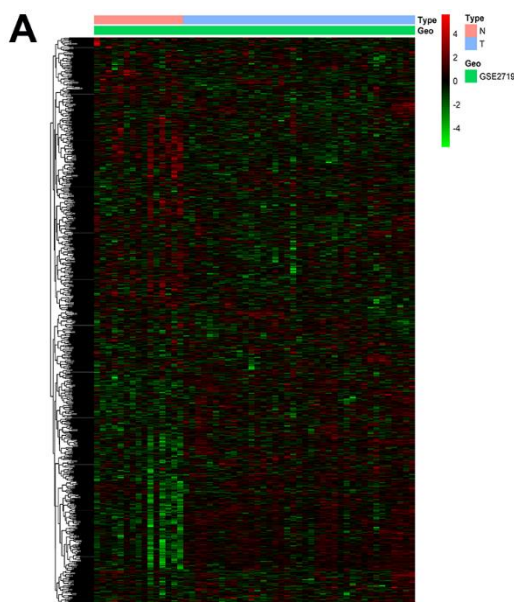


Figure 1. Framework for analyzing ZNFs in STS.

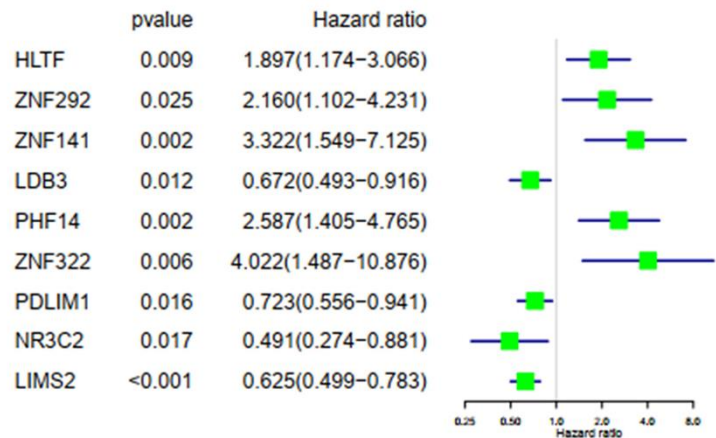
randomly divided into the training cohort or the testing cohort. Then univariate Cox regression analysis was accomplished to determine variables significantly associated with OS of patients in the TCGA training cohort. A total of nine differentially expressed ZNFs were demonstrated to be significantly associated with OS (HLTF, ZNF292, ZNF141, LDB3, PHF14, ZNF322, PDLIM1, NR3C2, and LIMS2) (Figure 2C). LASSO regression analysis was accomplished using the “glmnet” package of R software to select these identified ZNFs. Then the predictive model was established using five-fold cross-validation and we demonstrated the confidence interval under each lambda in Figure 3A. Trajectory of the coefficient for each gene with a value of $-\ln(\lambda)$ is shown in

Figure 3B. There was a total of nine genes identified as the model gene signature. The formula was as follows:

$$\begin{aligned} \text{Risk score} = & (0.225 \times \text{Exp HLTF}) + (0.495 \\ & \times \text{Exp ZNF292}) + (0.971 \\ & \times \text{Exp ZNF141}) + (-0.136 \\ & \times \text{Exp LDB3}) + (0.717 \\ & \times \text{Exp PHF14}) + (0.889 \\ & \times \text{Exp ZNF322}) + (-0.125 \\ & \times \text{Exp PDLIM1}) + (-0.809 \\ & \times \text{Exp NR3C2}) + (-0.098 \\ & \times \text{Exp LIMS2}). \end{aligned}$$



C Univariate COX regression analysis of OS in TCGA training cohort



D

Univariate COX regression analysis of PFS in TCGA training cohort

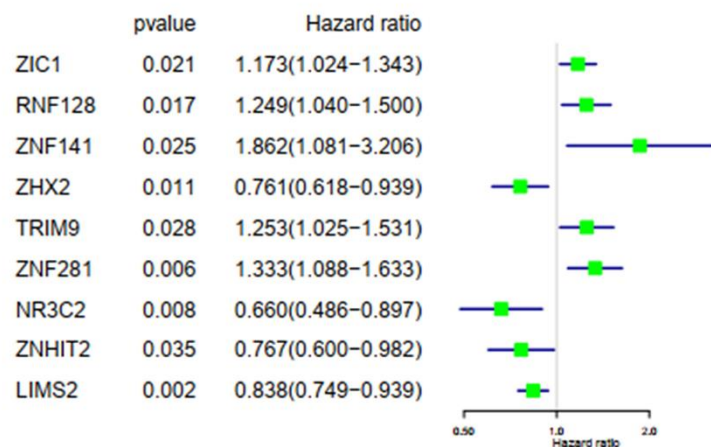


Figure 2. Differentially expressed ZNFs and univariate Cox regression analysis. Heatmap (A) and volcano plot (B) of 1,555 ZNFs in normal and STS tissues from GSE2719. Univariate Cox regression analysis of OS (C) and PFS (D) in TCGA training cohort.

Risk score of each patient was calculated and each patient from TCGA training cohort was divided into low-risk group or high-risk group. Then we performed survival analyses, results of which revealed that in comparison with patients in the low-risk group, those belonging to the high-risk group had significantly worse OS ($p < 0.001$, Figure 3C). Additionally, time-dependent ROC analysis was performed, results of which revealed that areas under ROC curve (AUC) at one, three, and five years were 0.807, 0.792, and 0.799, respectively (Figure 3C). We then conducted model verification using data of the TCGA testing cohort (Figure 3D) and the GSE21050 cohort (Figure 3E) to further assess the validity and stability of the prognostic models, results of which revealed that compared with those in the low-risk group, patients of the high-risk group had significantly poorer prognosis. The AUC of the model in TCGA testing cohort at one, three, and five years was 0.789, 0.699, and 0.636, respectively (Figure 3D), and those of the GSE21050 cohort were 0.613, 0.607, and 0.577, respectively (Figure 3E).

In a similar way, we then established and validated a model predicting PFS of patients with STS. Firstly, datasets of STS from TCGA were randomly divided into the training cohort or the testing cohort. Univariate Cox regression analysis was accomplished to evaluate the impacts of expression levels of ZNFs on PFS of patients from the training cohort. A total of nine ZNFs including ZIC1, RNF128, ZNF141, ZHX2, TRIM9, ZNF281, NR3C2, ZNHIT2, and LIMS2 were demonstrated to be significantly associated with PFS (Figure 2D). Then a subsequent LASSO regression analysis was performed to screen ZNFs (Figure 4A, 4B), results of which revealed that seven genes were included in the predictive model for PFS. The following was the formula:

$$\begin{aligned} \text{Risk score} = & (0.077 \times \text{Exp ZIC1}) + (0.195 \\ & \times \text{Exp ZNF141}) + (-0.273 \\ & \times \text{Exp ZHX2}) + (0.041 \\ & \times \text{Exp ZNF281}) + (-0.420 \\ & \times \text{Exp ZNHIT2}) + (-0.398 \\ & \times \text{Exp NR3C2}) + (-0.041 \\ & \times \text{Exp LIMS2}). \end{aligned}$$

For patients in the training cohort, patients with high risk had significantly shorter PFS than those with low risk ($p < 0.001$, Figure 4C). AUC of the PFS model at one, three, and five years was 0.711, 0.725, and 0.716, respectively (Figure 4C). Then patients from TCGA testing cohort (Figure 4D) and the GSE30929 cohort (Figure 4E) were utilized to further validate this predictive model for PFS. Like in the training cohort,

patients with high risk were demonstrated to have worse PFS than those with low risk. The AUC of the model in TCGA testing cohort was 0.681, 0.635, and 0.624 at one, three, and five years, (Figure 4D), and those in the GSE30929 cohort were 0.694, 0.679, and 0.668, respectively (Figure 4E).

Moreover, prognostic significance of risk score and the aforementioned clinical variables was evaluated among patients from TCGA training cohort and testing cohort by accomplishing Cox regression analysis, results of which revealed that for patients from both cohorts, risk score was an independent predictive factor for both OS ($p < 0.01$, Supplementary Figure 1A, 1B) and PFS ($p < 0.05$, Supplementary Figure 2A, 2B). Thus, through the aforementioned statistical analyses, we identified prognosis-associated ZNFs and established models that could reliably and efficiently predict long-term outcomes of patients diagnosed with STS.

Building predictive nomograms

A clinically useful nomogram that will help physicians more accurately predict OS (Figure 5A) or PFS (Figure 5B) of patients with STS was generated. According to the results of multivariate analysis among patients from the validation cohort, each variable was vested with a corresponding point based on the scale from the nomogram. A horizontal line was drawn to determine each variable's point. The total score of each patient was obtained by adding all the points together and based on the total score, we could more accurately estimate the one-, three-, and five-year survival rates of patients with STS.

GO enrichment analysis and GSVA analysis

In order to evaluate the possible functions of differentially expressed ZNFs in STS, we then performed GO enrichment analysis using the "clusterProfiler" package of R software. As shown in Figure 6A, we revealed that ZNFs were mainly located in PML bodies, stress fibers, contractile actin filament bundles, filamentous actin, and actin filament bundles. As for the specific molecular functions of these ZNFs, we found that ZNFs were mainly involved in DNA-binding transcriptional repressor and activator activity, ubiquitin-like protein transferase activity, ubiquitin-protein transferase activity and steroid hormone receptor activity.

In order to investigate the possible mechanisms through which ZNFs-based risk score model predicted prognosis of patients with STS, we then performed GSVA analysis. In the model predicting OS, the primary differences in the KEGG pathway between high-risk

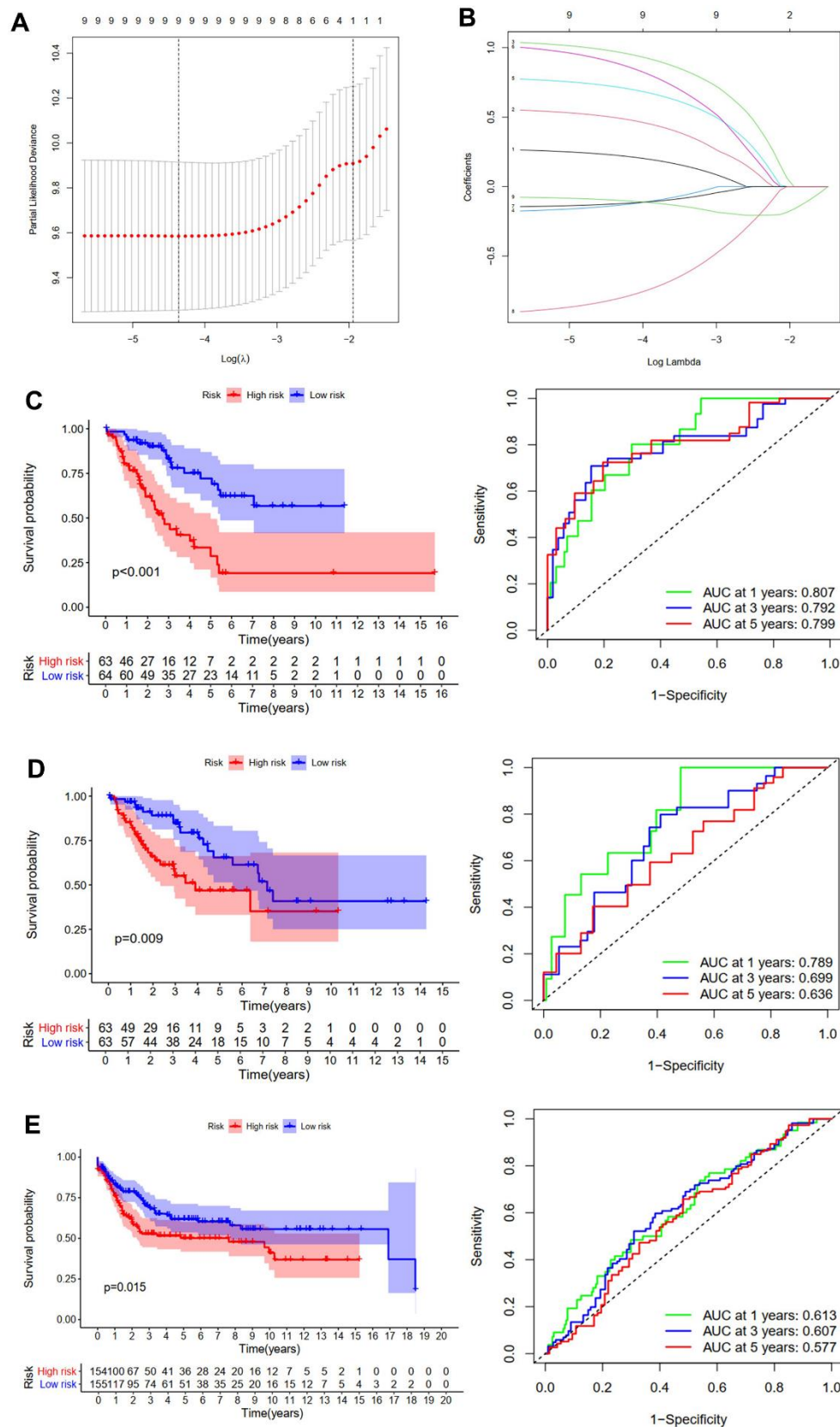


Figure 3. Development and validation of the OS prediction model for STS. (A) The plot of partial likelihood deviance. (B) The changing trajectory of each ZNF variable. Survival curve and ROC curve for low- and high-risk subgroups in TCGA training cohort (C), TCGA testing cohort (D) and GSE21050 cohort (E).

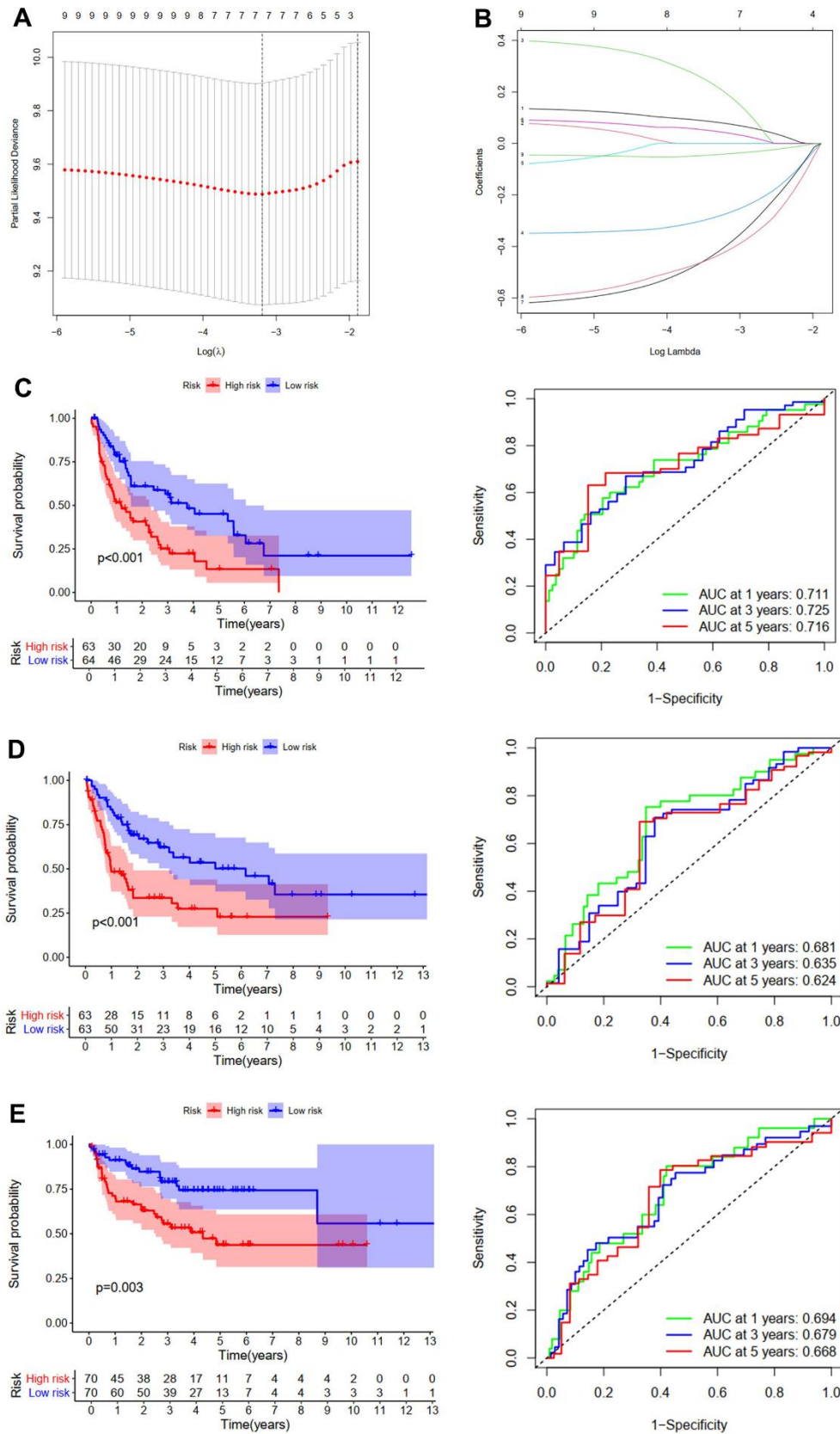


Figure 4. Development and validation of the PFS prediction model for STS. (A) The plot of partial likelihood deviance. (B) The changing trajectory of each ZNF variable. Survival curve and ROC curve for low- and high-risk subgroups in TCGA training cohort (C), TCGA testing cohort (D) and GSE30929 cohort (E).

group and low-risk group were as follows: vascular smooth muscle contraction, phosphatidylinositol signaling system, inositol phosphate metabolism, adipocytokine signaling pathway and inositol phosphate metabolism (Figure 6B), which were significantly more activated among patients in the low-risk group and this might lead to better OS. In the model predicting PFS, the primary differences in the KEGG pathway between high-risk group and low-risk group were as follows: histidine, tyrosine, tryptophan, arachidonic acid, fatty acid, and drug metabolism cytochrome P450, and metabolism of xenobiotics by cytochrome P450 (Figure 6C), which were significantly more activated among patients in the low-risk group and this might lead to better PFS.

Molecular subtypes and immune infiltration of STS based on ZNFs

Thirteen genes (*HLTF*, *LDB3*, *LIMS2*, *NR3C2*, *PDLIM1*, *PHF14*, *ZHX2*, *ZIC1*, *ZNF141*, *ZNF281*, *ZNF292*, *ZNF322*, and *ZNHIT2*) were identified to be significantly associated with prognosis. Then we evaluated the prevalence of somatic mutations in 1,555 ZNFs in STS. Of the 237 samples, a total of 51 (21.52%) ones were demonstrated to be with genetic alterations, primarily missense mutations, nonsense mutations, and multiple hits in ZNFs. PCLO was the frequent mutation (Figure 7A). Then through further analysis of the 13 prognosis-related ZNFs, we demonstrated that copy number variation (CNV) mutations were prevalent. Specifically, *PHF14*, *ZHX2*, and *ZNF141* showed widespread amplifications, while *LDB3*, *PDLIM1*, *ZNF292*, *ZNF322*, and *ZNHIT2* showed prevalent

deletions (Figure 7B). The locations of CNV alterations in the 13 prognosis-related ZNFs on the chromosomes are shown in Figure 7C.

Consensus ClusterPlus package was adopted to divide STS samples from TCGA into k ($k=2-9$) subtypes based on 13 ZNFs significantly associated with prognosis. Best division of the CDF curve based on consensus scores could be achieved when k was 4. Accordingly, we identified four distinct subtypes (Figure 8A), including 34, 67, 87, and 65 cases in cluster A, B, C, and D, respectively (Figure 8B). It was demonstrated through survival analysis that cluster B had prominent survival advantages over other clusters and cluster D was found to have the worst OS ($p<0.001$, Figure 8B) and PFS ($p=0.001$, Figure 8C). The heatmap of the 13 prognosis-related ZNFs for four different subtypes was presented in Figure 8D.

Subsequently, we further evaluated the immunological characteristics of the four subtypes. The immune and stromal scores of samples in the TCGA cohort were quantified using the ESTIMATE algorithm, results of which demonstrated that in comparison with patients with high scores, OS ($p<0.01$, Supplementary Figure 3A) and PFS ($p<0.05$, Supplementary Figure 3B) of those with low scores were significantly worse. Four subtypes were revealed to have significantly different immune scores. Immune scores of subtypes B and C were significantly higher than those of subtypes A and D (Figure 8E). Results of analysis of the immunological characteristics implied that unlike subtypes B and C, immune responses within subtypes A and D STS were significantly suppressed. Furthermore, we assessed the

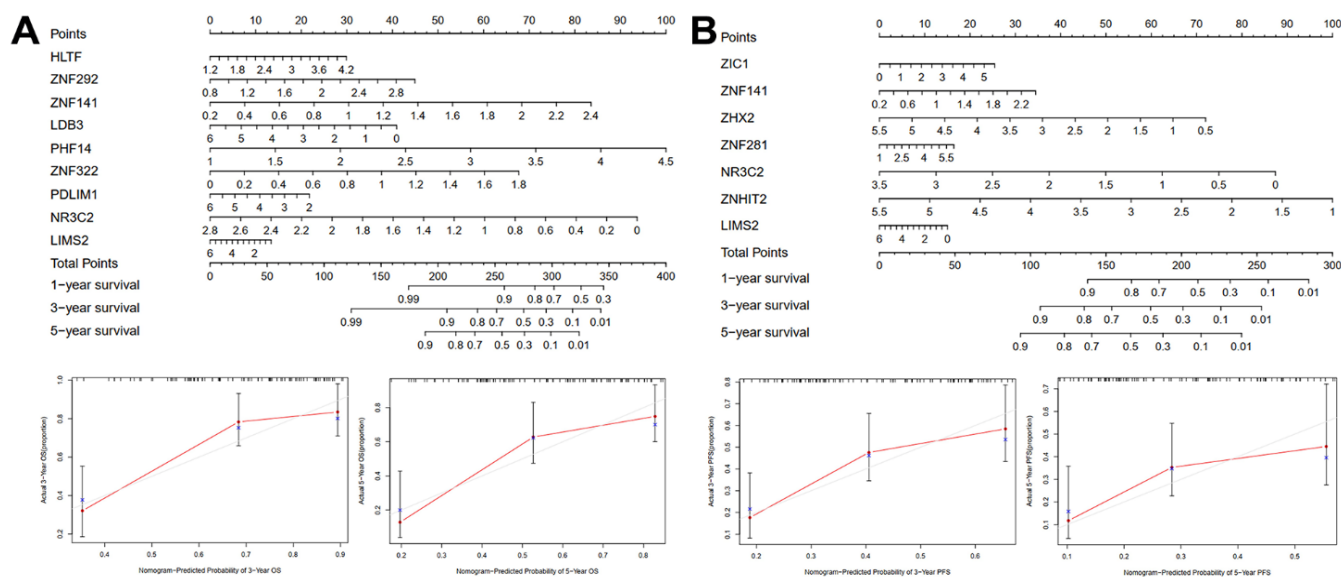


Figure 5. Nomogram and calibration plots for predicting one-, three-, and five-year OS (A) or PFS (B) of TCGA training cohort.

expression of several immune checkpoints in four different subtypes, results of which demonstrated that compared with those in subtypes A and D, expression levels of most immune checkpoints were significantly higher in subtypes B and C (Figure 8F), which was consistent with results of immune score analysis. Then the immune infiltration profile of the TCGA cohort was

characterized with CIBERSORT. For STS samples from TCGA, the major constituents of TIICs were CD8 T cells, resting CD4 memory T cells, resting mast cells, M2 and M0 macrophages (Figure 8G). Additionally, it was also observed that infiltration proportions of five immune cells in four subtypes were significantly different (Figure 8G). Significantly much more resting

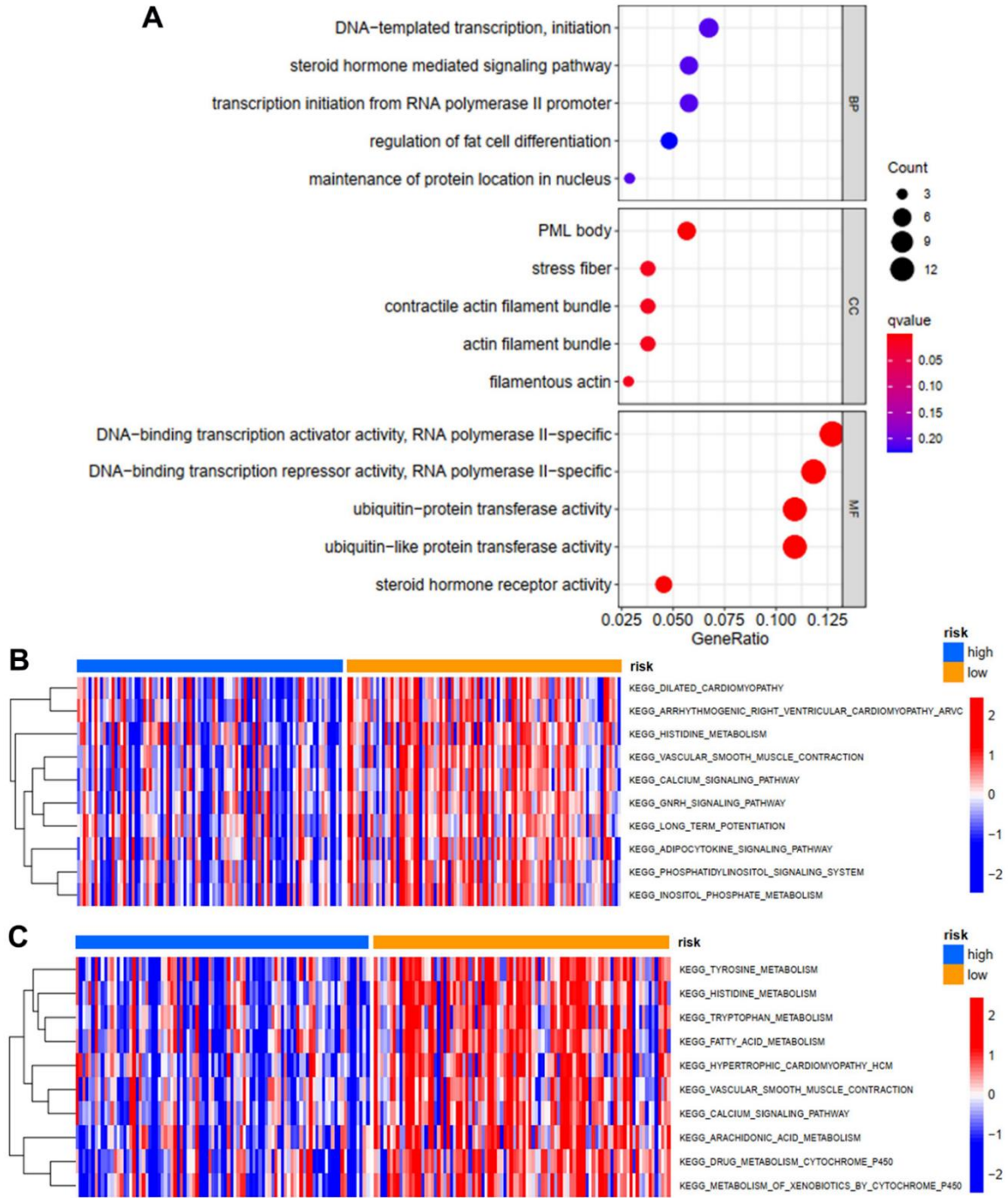


Figure 6. GO enrichment and GSEA analysis. (A) GO enrichment analysis of differentially expressed ZNFs. GSEA analysis for OS (B) and PFS (C) in low- and high-risk subgroups of TCGA training cohort.

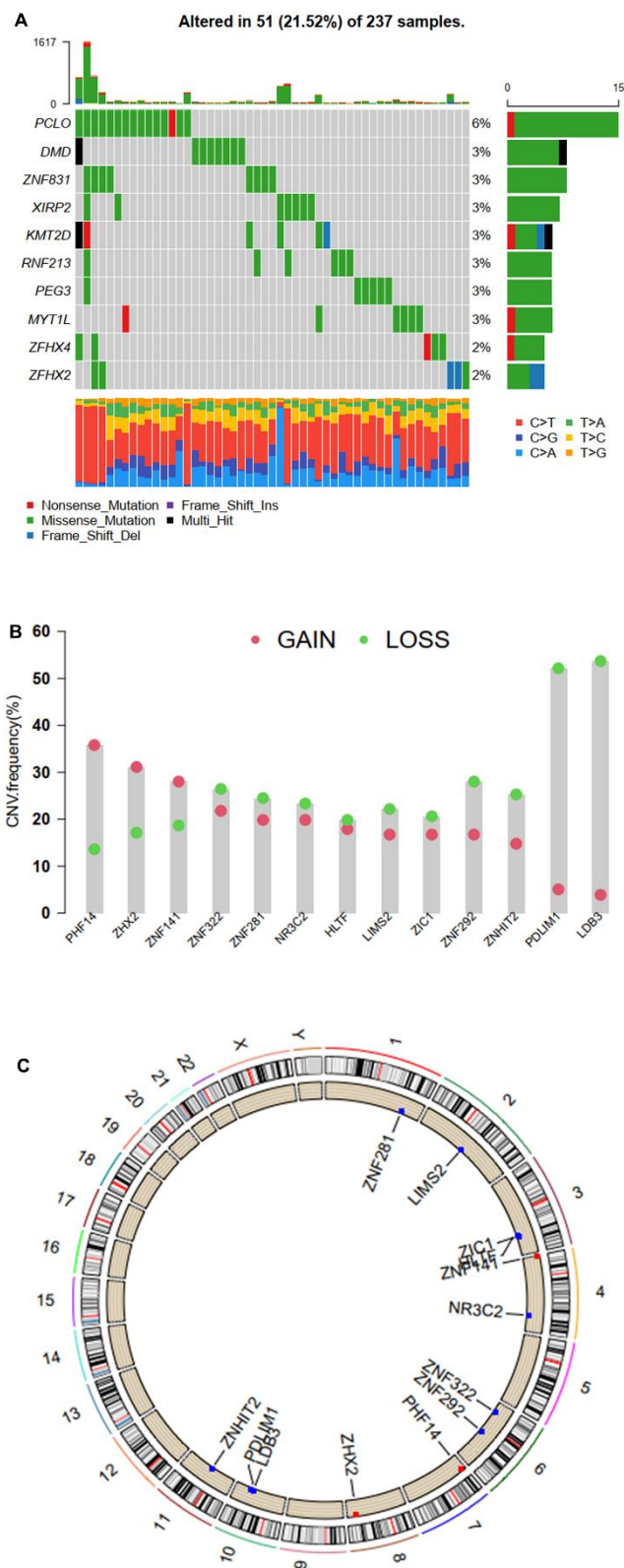


Figure 7. Mutation and CNV frequency of ZNFs. (A) The mutation frequency of ZNFs in 237 patients with STS from TCGA cohort. (B) The CNV frequency of ZNFs in TCGA cohort. (C) The location of CNV alterations of ZNFs on chromosomes using TCGA cohort.

CD4 memory T cells and mast cells were observed in subtype A STS than in other three subtypes. Moreover, it was demonstrated that CD8 T cells within subtype B STS were significantly more prominent while significantly more abundant infiltration of M2 within subtypes C and D STS and M0 within subtype D STS were detected. Therefore, given the fact that the dominant immune cells within subtype B were CD8+ T cells (the main killers of tumor cells, Philip and Schietinger, 2021) while the dominant immune cells within subtype C were M2 macrophages (the main tumor promoters, Najafi et al., 2019), we could explain why prognosis of subtype B STS was significantly better than that of subtype C STS. Similarly, the most prominent immune cells within subtype A STS were resting CD4 memory T cells and mast cells and the most prominent immune cells within subtype D STS were M0 and M2 macrophages, which might explain why patients with subtype A STS had better prognosis than those with subtype D STS (Supplementary Figure 3C). Therefore, considering all the results, we could draw the conclusion that proportion of immune cells within STS would significantly affect patients' prognosis.

ZNF141 promoted the proliferation and viability of STS cells

ZNF141 has been identified as an OS- and RFS-related ZNF, but its role in STS is still unclear. Therefore, we verified the effect of ZNF141 on STS cells *in vitro*. ZNF141 mRNA expression was upregulated in tumor tissues compared to that in normal tissues (Figure 9A). The results from TCGA dataset show that patients with high ZNF141 mRNA levels had poorer OS (Figure 9B) and RFS (Figure 9C) than those with low ZNF141 mRNA levels. Subsequently, siRNAs (si-nc, si-1, and si-2) were transfected to A673 cells. RT-PCR and western blotting assays showed that ZNF141 expression was significantly downregulated in A673 cells transfected with si-1 or si-2 compared with those transfected with si-nc (Figure 9D). The results of the CCK8 showed that the proliferation of A673 cells with knockdown ZNF141 (si-1 or si-2) was weaker than that of the control group (si-nc) (Figure 9E). Plate clone formation assays showed that the colony numbers of A673 cells with ZNF141 knockdown were significantly reduced (Figure 9F). Next, ZNF141 overexpression (GV- ZNF141) or control plasmid (GV-Vector) were transfected to SW982 cells. The successful overexpression of ZNF141 in SW982 cells was verified by both RT-PCR and western blotting assays (Figure 9G). CCK8 assays showed that the proliferation of SW982 cells with ZNF141 overexpression (OE) was stronger than that of the control group (NC) (Figure 9H). Plate clone formation assays showed that the colony

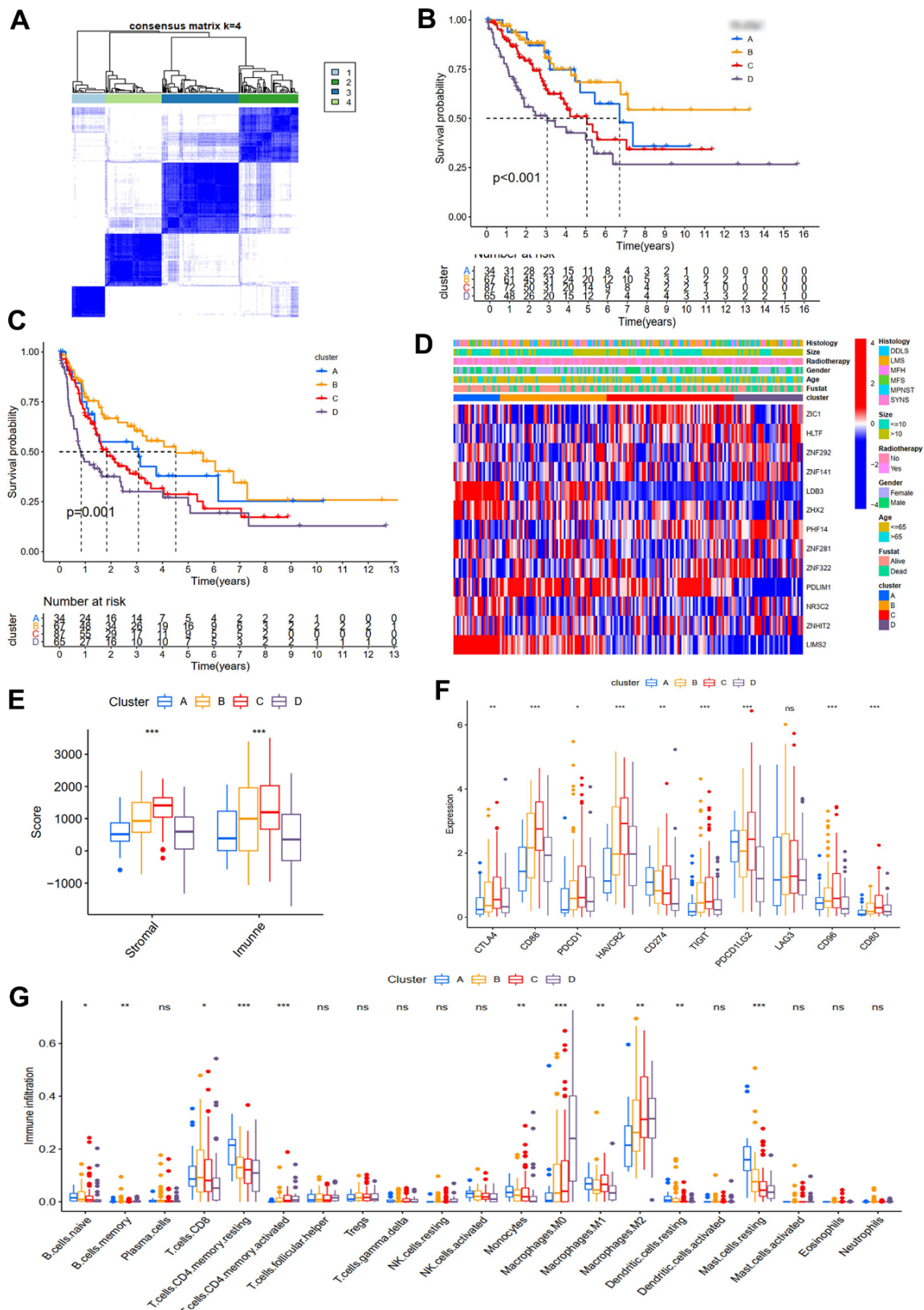


Figure 8. Molecular subtypes and immune infiltration. (A) Consensus matrices of TCGA cohort for k=4. OS (B) and PFS (C) for the four clusters in TCGA cohort. (D) Expression heatmap of ZNF prognosis-related genes in four subtypes. (E) Immune and stromal scores of four subtypes from the ESTIMATE algorithm. (F) The differences of immune checkpoint genes in four subtypes. (G) TIICs of four subtypes.

numbers of SW982 cells with ZNF141 overexpression were significantly increased (Figure 9I). These results suggested that ZNF141 promoted the proliferation and viability of STS cells, and we could conclude that ZNF141 played critical roles in promoting progression of STS.

DISCUSSION

ZNFs were the most abundant transcriptional regulatory factors within mammal cells. It has been reported by many previous studies that ZNFs play vital roles in occurrence and progression of malignant

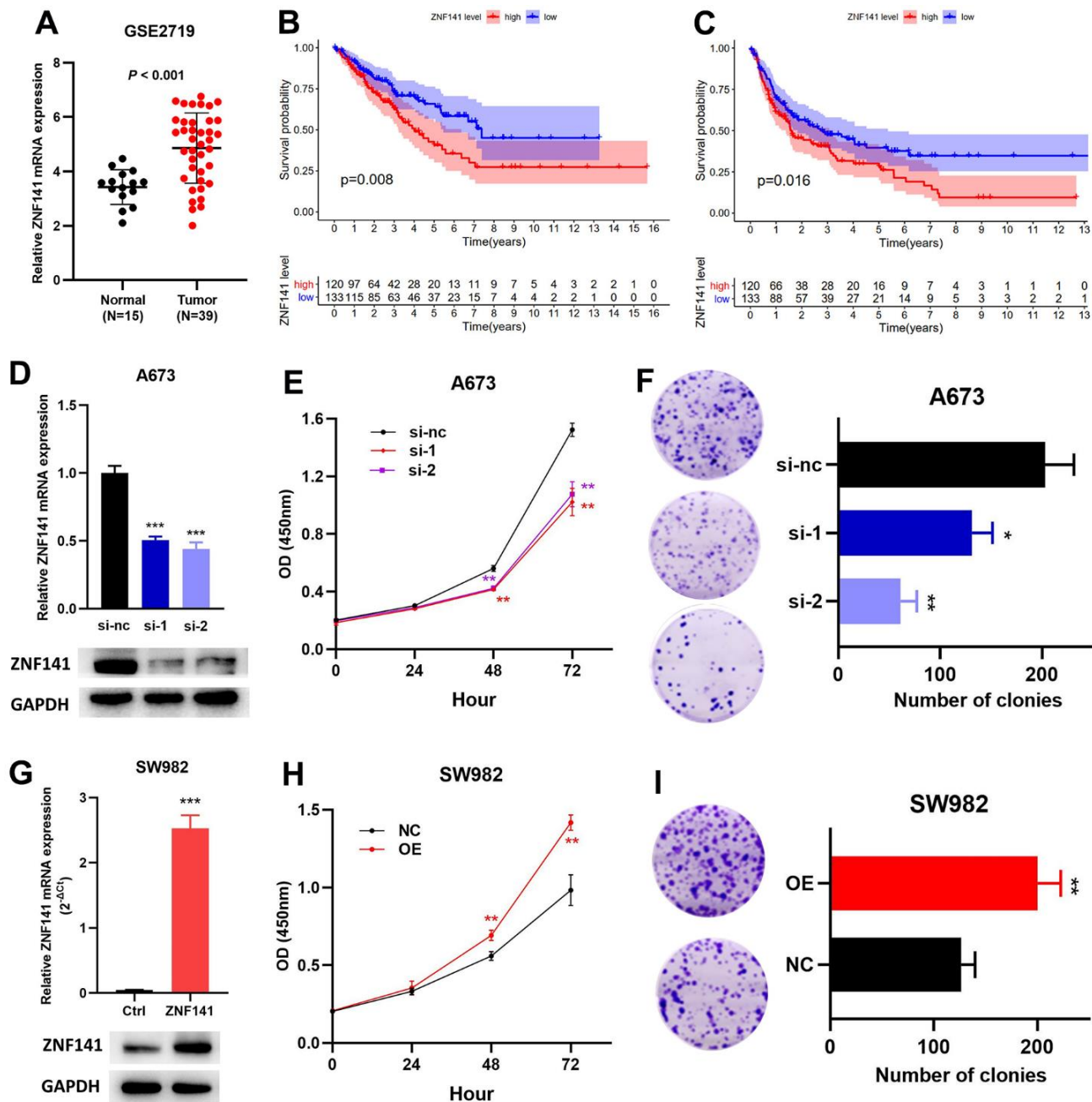


Figure 9. ZNF141 promotes the proliferation and viability of STS cells. (A) ZNF141 expression in normal tissues and STS tissues (GSE2719). The OS (B) and RFS (C) analysis of ZNF141 expression in TCGA cohort. (D) ZNF141 knockdown A673 cells were constructed and confirmed by RT-PCR and Western blotting. (E) The proliferation ability of A673 cells with ZNF141 (si-nc, si-1 or si-2) via CCK8 assays. (F) Plate clone formation assays of A673 cells with ZNF141 (si-nc, si-1 or si-2). (G) ZNF141 overexpression SW982 cells were constructed and confirmed by RT-PCR and Western blotting. (H) The proliferation ability of SW982 cells with ZNF141 overexpression or control plasmid via CCK8 assays. (I) Plate clone formation assays SW982 cells with ZNF141 overexpression or control plasmid.

tumors [9, 10, 14, 15]. However, roles of ZNFs in STS have not been fully investigated. In this study, we identified 110 ZNFs whose expression in STS was differentially between STS samples and normal tissues. We performed univariate Cox regression analysis and revealed that some ZNFs were significantly associated with OS and PFS of patients with STS. Then we established and validated models for OS and PFS using LASSO-penalized Cox regression analysis and these models could efficiently predict prognosis of patients with STS. Additionally, we used consensus clustering to classify subtypes of STS. As far as we know, this is by far the first comprehensive study investigating roles played by ZNFs in STS and our study would build a framework for future studies.

Firstly, we identified 110 differentially expressed ZNFs among the 1555 total ZNFs. We then evaluated prognostic significance of these differentially expressed ZNFs. Subsequently, we accomplished LASSO-penalized Cox survival analysis and established models predicting OS of the TCGA training cohort. Nine genes (*HLTF*, *ZNF292*, *ZNF141*, *LDB3*, *PHF14*, *ZNF322*, *PDLIM1*, *NR3C2*, and *LIMS2*) were selected to establish the OS prediction model, and seven genes (*ZIC1*, *ZNF141*, *ZHX2*, *ZNF281*, *ZNHIT2*, *NR3C2*, and *LIMS2*) were utilized to build the PFS prediction model. It was then demonstrated through ROC analysis that the prediction models for OS and PFS were reliable and efficient predicting prognosis of patients with STS, which was then further validated by the testing cohort from TCGA and GEO cohorts (GSE21050 for the OS model and GSE30929 for the PFS model). These findings suggested the clinical usefulness and efficiency of models predicting OS and PFS. Ultimately, we established OS and PFS nomograms to enable physicians to better predict survival of patients with STS at one, three, and five years.

Then potential functions of ZNFs in STS and the possible pathways through which ZNFs exert these functions were investigated. It was revealed through GO enrichment analysis that ZNFs played vital roles in DNA-binding transcription activator or repressor activity, ubiquitin-protein transferase activity, and steroid hormone receptor activity. Then we performed GSVA analysis and demonstrated that high-risk groups and low-risk groups were different in terms of the following pathways: amino acid and fatty acid metabolism, adipocytokine signaling pathway, and drug metabolism cytochrome P450. Previous studies reported that amino acid metabolism was an important factor promoting the formation immunosuppressive tumor microenvironment and resistance of cancer cells to drugs [22, 23]. Additionally, it has also been reported that growth and survival of tumor cells are significantly

affected by fatty acid metabolism and enzymes regulating synthesis of fatty acid are potential targets treating malignant tumors [24]. Similarly, it has also been reported that cytochrome P450 is important in growth of tumors and could regulate resistance of tumor cells to drugs [25]. However, the roles of these pathways in STS should be further evaluated to improve survival of patients with STS.

Next, based on 13 prognosis-related ZNFs, four subtypes with significantly distinct prognosis and remarkably different immune infiltration characteristics were identified. Of these four subtypes, patients with subtype B had a significantly prominent survival advantages while OS and PFS of patients with subtype D were the worst. Then immune cell infiltration characteristics of the four subtypes were analyzed, results of which revealed that subtypes A and D were with low immune scores while subtypes B and C were with high immune scores. Then through CIBERSORT, we revealed distinct characteristics of TIICs among the four subtypes. Subtype A is dominated by resting CD4 memory T cells and resting mast cells, subtype B by CD8 T cells, subtype C by M2 macrophages, and subtype D by M0 and M2 macrophages. These TIICs have different effects on tumor prognosis. According to some previously published studies, resting mast cells and CD4 memory T cells are predictors for better prognosis [26], which was in line with our finding that resting mast cells were positively correlated with prognosis. Moreover, CD8 T cells are killers that specially kill tumor cells [27], while M2 cells are tumor promoters [28], and M0 macrophages are strongly associated with poor outcomes [26]. These TIICs also play different roles in STS, which further influence the prognosis of the four subtypes. Therefore, making more precise treatment plans according to the characteristics of the four subtypes will be beneficial for improving the prognosis of patients with STS.

Moreover, ZNF141 was related to both OS and RFS, but the role of ZNF141 in STS remains unclear. The role of ZNF141 in STS cells was preliminarily explored. Our results suggest that ZNF141 promotes the proliferation and viability of STS cells. Recent studies suggest that ZNF141 is a potential marker of extranodal NK/T-cell lymphoma [29]. Therefore, we could conclude that ZNF141 played critical roles in promoting progression of STS and may serve as a potential prognostic biomarker. Obviously, further research needs to be explored.

In conclusion, we performed a comprehensive study investigating prognostic significance and possible functions of differentially expressed ZNFs in STS. Predictive models that could efficiently and reliably

predict OS and PFS were also developed and validated. Additionally, four subtypes of STS with significantly distinct prognosis and remarkably different immune infiltration characteristics were identified. Taking all these results together, we identified some potential targets for developing novel treatment strategies that may improve prognosis of patients with STS.

AUTHOR CONTRIBUTIONS

ZM and YJ conceptualized and supervised the study. LJQ and ZQ performed bioinformatic analysis and statistical analysis. ZM performed *in vitro* experiments. ZCS and ZHM performed data curation and interpretation. LJQ wrote the original manuscript.

ACKNOWLEDGMENTS

We appreciate the opening of TCGA and GEO databases to researchers.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING

No funding was provided to this study.

Editorial note

[&]This corresponding author has a verified history of publications using a personal email address for correspondence.

REFERENCES

1. Massarweh NN, Dickson PV, Anaya DA. Soft tissue sarcomas: staging principles and prognostic nomograms. *J Surg Oncol.* 2015; 111:532–9. <https://doi.org/10.1002/jso.23851> PMID:[25482536](https://pubmed.ncbi.nlm.nih.gov/25482536/)
2. Blay JY. Treatment of advanced soft tissue sarcoma by histological subtype: wish, prediction or reality? *Future Oncol.* 2019; 15:5–10. <https://doi.org/10.2217/fon-2019-0488> PMID:[31500446](https://pubmed.ncbi.nlm.nih.gov/31500446/)
3. Savina M, Le Cesne A, Blay JY, Ray-Coquard I, Mir O, Toulmonde M, Cousin S, Terrier P, Ranchere-Vince D, Meeus P, Stoeckle E, Honoré C, Sargos P, et al. Patterns of care and outcomes of patients with METAstatic soft tissue SARComa in a real-life setting: the METASARC observational study. *BMC Med.* 2017; 15:78. <https://doi.org/10.1186/s12916-017-0831-7> PMID:[28391775](https://pubmed.ncbi.nlm.nih.gov/28391775/)

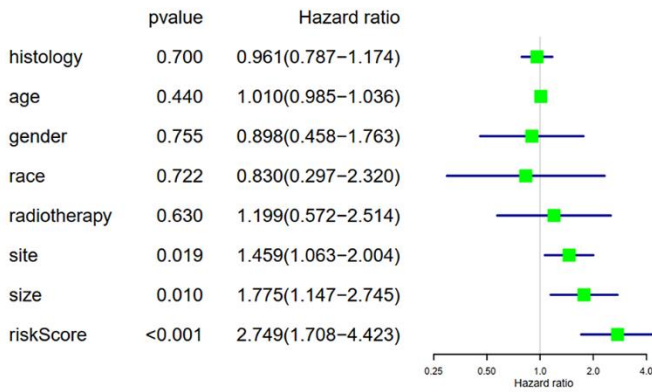
4. Ray-Coquard I, Serre D, Reichardt P, Martín-Broto J, Bauer S. Options for treating different soft tissue sarcoma subtypes. *Future Oncol.* 2018; 14:25–49. <https://doi.org/10.2217/fon-2018-0076> PMID:[29768052](https://pubmed.ncbi.nlm.nih.gov/29768052/)
5. Tanaka K, Ozaki T. Adjuvant and neoadjuvant chemotherapy for soft tissue sarcomas: JCOG Bone and Soft Tissue Tumor Study Group. *Jpn J Clin Oncol.* 2021; 51:180–4. <https://doi.org/10.1093/jjco/hyaa231> PMID:[33313851](https://pubmed.ncbi.nlm.nih.gov/33313851/)
6. Yang W, Lei C, Song S, Jing W, Jin C, Gong S, Tian H, Guo T. Immune checkpoint blockade in the treatment of malignant tumor: current status and future strategies. *Cancer Cell Int.* 2021; 21:589. <https://doi.org/10.1186/s12935-021-02299-8> PMID:[34727927](https://pubmed.ncbi.nlm.nih.gov/34727927/)
7. Ngwa CJ, Farrukh A, Pradel G. Zinc finger proteins of *Plasmodium falciparum*. *Cell Microbiol.* 2021; 23:e13387. <https://doi.org/10.1111/cmi.13387> PMID:[34418264](https://pubmed.ncbi.nlm.nih.gov/34418264/)
8. Laity JH, Lee BM, Wright PE. Zinc finger proteins: new insights into structural and functional diversity. *Curr Opin Struct Biol.* 2001; 11:39–46. [https://doi.org/10.1016/s0959-440x\(00\)00167-6](https://doi.org/10.1016/s0959-440x(00)00167-6) PMID:[11179890](https://pubmed.ncbi.nlm.nih.gov/11179890/)
9. Jen J, Wang YC. Zinc finger proteins in cancer progression. *J Biomed Sci.* 2016; 23:53. <https://doi.org/10.1186/s12929-016-0269-9> PMID:[27411336](https://pubmed.ncbi.nlm.nih.gov/27411336/)
10. Weng H, Wang X, Li M, Wu X, Wang Z, Wu W, Zhang Z, Zhang Y, Zhao S, Liu S, Mu J, Cao Y, Shu Y, et al. Zinc finger X-chromosomal protein (ZFX) is a significant prognostic indicator and promotes cellular malignant potential in gallbladder cancer. *Cancer Biol Ther.* 2015; 16:1462–70. <https://doi.org/10.1080/15384047.2015.1070994> PMID:[26230915](https://pubmed.ncbi.nlm.nih.gov/26230915/)
11. Horiguchi K, Sakamoto K, Koinuma D, Semba K, Inoue A, Inoue S, Fujii H, Yamaguchi A, Miyazawa K, Miyazono K, Saitoh M. TGF- β drives epithelial-mesenchymal transition through δ EF1-mediated downregulation of ESRP. *Oncogene.* 2012; 31: 3190–201. <https://doi.org/10.1038/onc.2011.493> PMID:[22037216](https://pubmed.ncbi.nlm.nih.gov/22037216/)
12. Liu X, Ge X, Zhang Z, Zhang X, Chang J, Wu Z, Tang W, Gan L, Sun M, Li J. MicroRNA-940 promotes tumor cell invasion and metastasis by downregulating ZNF24 in gastric cancer. *Oncotarget.* 2015; 6:25418–28. <https://doi.org/10.18632/oncotarget.4456> PMID:[26317898](https://pubmed.ncbi.nlm.nih.gov/26317898/)
13. Cheng Y, Liang P, Geng H, Wang Z, Li L, Cheng SH, Ying J, Su X, Ng KM, Ng MH, Mok TS, Chan AT, Tao Q. A

- novel 19q13 nucleolar zinc finger protein suppresses tumor cell growth through inhibiting ribosome biogenesis and inducing apoptosis but is frequently silenced in multiple carcinomas. *Mol Cancer Res*. 2012; 10:925–36.
<https://doi.org/10.1158/1541-7786.MCR-11-0594>
PMID:22679109
14. Yu J, Liang QY, Wang J, Cheng Y, Wang S, Poon TC, Go MY, Tao Q, Chang Z, Sung JJ. Zinc-finger protein 331, a novel putative tumor suppressor, suppresses growth and invasiveness of gastric cancer. *Oncogene*. 2013; 32:307–17.
<https://doi.org/10.1038/onc.2012.54>
PMID:22370639
15. Hu R, Wang E, Peng G, Dai H, Lin SY. Zinc finger protein 668 interacts with Tip60 to promote H2AX acetylation after DNA damage. *Cell Cycle*. 2013; 12:2033–41.
<https://doi.org/10.4161/cc.25064> PMID:2377805
16. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015; 43:e47.
<https://doi.org/10.1093/nar/gkv007> PMID:25605792
17. Chibon F, Lagarde P, Salas S, Pérot G, Brouste V, Tirode F, Lucchesi C, de Reynies A, Kauffmann A, Bui B, Terrier P, Bonvalot S, Le Cesne A, et al. Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. *Nat Med*. 2010; 16:781–7.
<https://doi.org/10.1038/nm.2174> PMID:20581836
18. Gobble RM, Qin LX, Brill ER, Angeles CV, Ugras S, O'Connor RB, Moraco NH, Decarolis PL, Antonescu C, Singer S. Expression profiling of liposarcoma yields a multigene predictor of patient outcome and identifies genes that contribute to liposarcomagenesis. *Cancer Res*. 2011; 71:2697–705.
<https://doi.org/10.1158/0008-5472.CAN-10-3588>
PMID:21335544
19. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013; 14:7.
<https://doi.org/10.1186/1471-2105-14-7>
PMID:23323831
20. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010; 26:1572–3.
<https://doi.org/10.1093/bioinformatics/btq170>
PMID:20427518
21. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015; 12:453–7.
<https://doi.org/10.1038/nmeth.3337> PMID:25822800
22. Tabe Y, Lorenzi PL, Konopleva M. Amino acid metabolism in hematologic malignancies and the era of targeted therapy. *Blood*. 2019; 134:1014–23.
<https://doi.org/10.1182/blood.2019001034>
PMID:31416801
23. Wang W, Zou W. Amino Acids and Their Transporters in T Cell Immunity and Cancer Therapy. *Mol Cell*. 2020; 80:384–95.
<https://doi.org/10.1016/j.molcel.2020.09.006>
PMID:32997964
24. Fhu CW, Ali A. Fatty Acid Synthase: An Emerging Target in Cancer. *Molecules*. 2020; 25:3935.
<https://doi.org/10.3390/molecules25173935>
PMID:32872164
25. Stipp MC, Acco A. Involvement of cytochrome P450 enzymes in inflammation and cancer: a review. *Cancer Chemother Pharmacol*. 2021; 87:295–309.
<https://doi.org/10.1007/s00280-020-04181-2>
PMID:33112969
26. Li X, Li J, Wu P, Zhou L, Lu B, Ying K, Chen E, Lu Y, Liu P. Smoker and non-smoker lung adenocarcinoma is characterized by distinct tumor immune microenvironments. *Oncoimmunology*. 2018; 7:e1494677.
<https://doi.org/10.1080/2162402X.2018.1494677>
PMID:30288364
27. Philip M, Schietinger A. CD8⁺ T cell differentiation and dysfunction in cancer. *Nat Rev Immunol*. 2022; 22:209–23.
<https://doi.org/10.1038/s41577-021-00574-3>
PMID:34253904
28. Najafi M, Hashemi Goradel N, Farhood B, Salehi E, Nashtaei MS, Khanlarkhani N, Khezri Z, Majidpoor J, Abouzaripour M, Habibi M, Kashani IR, Mortezaee K. Macrophage polarity in cancer: A review. *J Cell Biochem*. 2019; 120:2756–65.
<https://doi.org/10.1002/jcb.27646> PMID:30270458
29. Wang Y, Tan H, Yu T, Ma X, Chen X, Jing F, Zou L, Shi H. The identification of gene signatures in patients with extranodal NK/T-cell lymphoma from a pair of twins. *BMC Cancer*. 2021; 21:1303.
<https://doi.org/10.1186/s12885-021-09023-9>
PMID:34872521

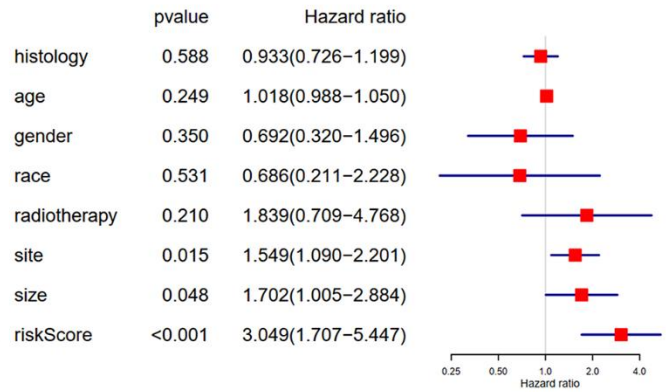
SUPPLEMENTARY MATERIALS

Supplementary Figures

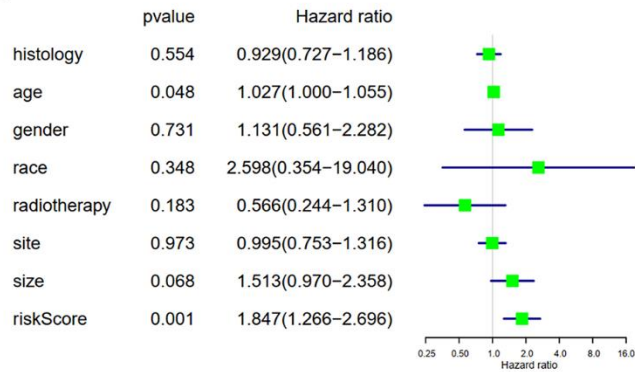
A Univariate COX regression analysis of OS in TCGA training cohort



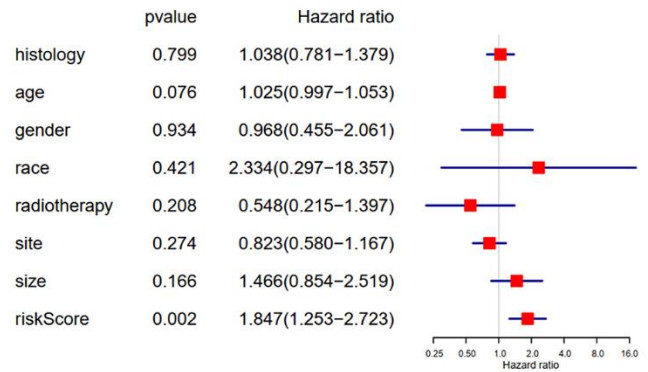
Multivariate COX regression analysis of OS in TCGA training cohort



B Univariate COX regression analysis of OS in TCGA testing cohort

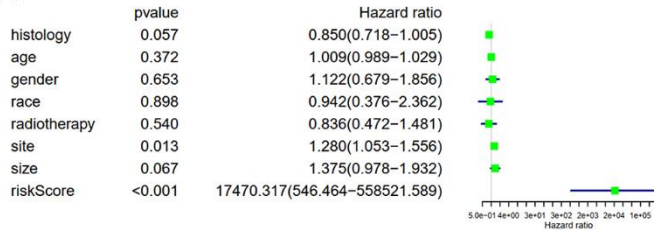


Multivariate COX regression analysis of OS in TCGA testing cohort

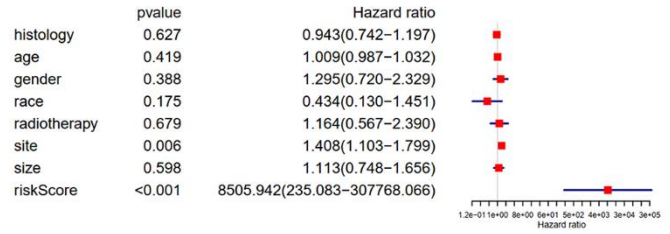


Supplementary Figure 1. Univariate and multivariate Cox regression analysis of OS in TCGA training (A) and testing cohorts (B).

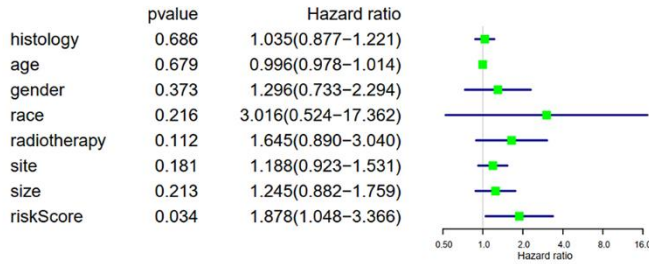
A Univariate COX regression analysis of PFS in TCGA training cohort



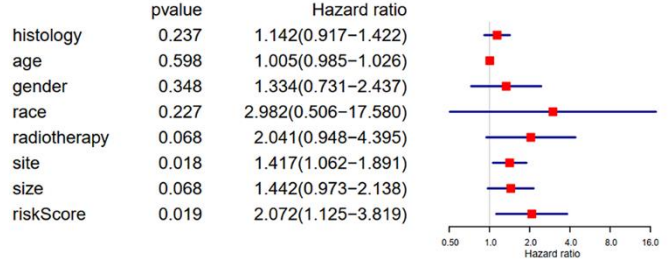
Multivariate COX regression analysis of PFS in TCGA training cohort



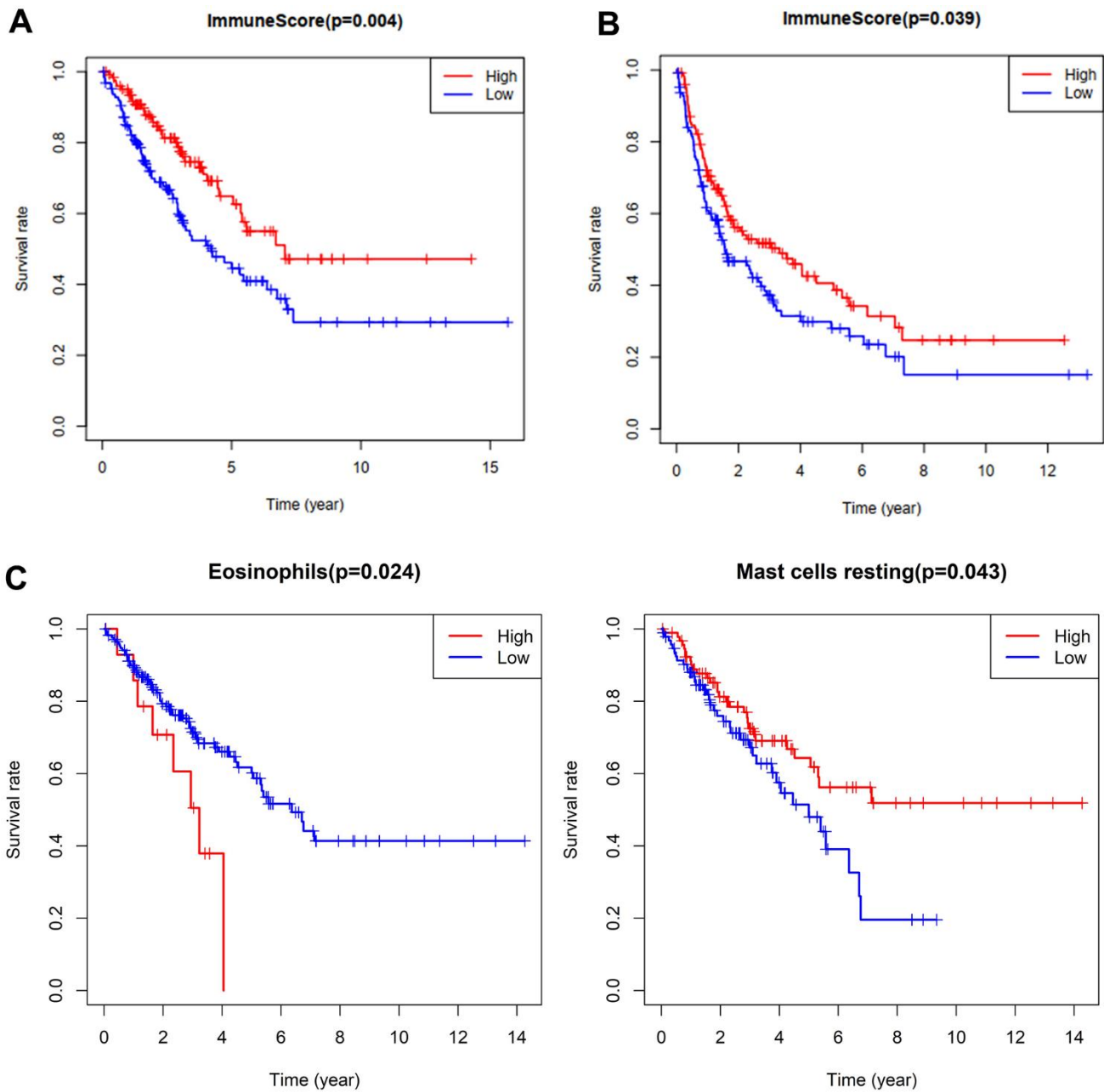
B Univariate COX regression analysis of PFS in TCGA testing cohort



Multivariate COX regression analysis of PFS in TCGA testing cohort



Supplementary Figure 2. Univariate and multivariate Cox regression analysis of PFS in TCGA training (A) and testing cohorts (B).



Supplementary Figure 3. Survival curve of immune score and tumor-infiltrating immune cells. OS (A) and PFS (B) curve of patients with high- or low- immune scores in the TCGA cohort. (C) OS curve of eosinophils and mast cells resting in the TCGA cohort.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1, 3.

Supplementary Table 1. Symbols and categories of the 1,555 ZNFs.

Supplementary Table 2. The sequences of primer and siRNA oligonucleotides.

ZNF141	F	5'- GACGGTCCACAGATCGGAG-3'
	R	5'- TGACTCAGGAGCGAAAATTGTT -3'
GAPDH	F	5'- ACAACTTTGGTATCGTGAAG -3'
	R	5'- GCCATCACGCCACAGTTTC -3'
si-1	SS	5'-GUGUCAAGUUGUUAGUAA-3'
	AS	5'-UUACUAACAACUUUGACAC-3'
si-2	SS	5'-GAGAUGUGAUGUUGGAGAA-3'
	AS	5'-UUCUCCAACAUCACAUCUC-3'

Supplementary Table 3. Differentially expressed ZNFs in GSE2719.