



Construction of Two Alternative Polyadenylation Signatures to Predict the Prognosis of Sarcoma Patients

Chuan Hu^{1†}, Chuan Liu^{2†}, Jianyi Li^{1†}, Tengbo Yu³, Jun Dong⁴, Bo Chen⁵, Yukun Du¹, Xiaojie Tang¹ and Yongming Xi^{1*}

¹ Department of Orthopaedic Surgery, The Affiliated Hospital of Qingdao University, Qingdao, China, ² Graduate School, China Medical University, Shenyang, China, ³ Department of Sports Medicine, The Affiliated Hospital of Qingdao University, Qingdao, China, ⁴ Department of Orthopedics, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, China, ⁵ The First Clinical College, Wenzhou Medical University, Wenzhou, China

OPEN ACCESS

Edited by:

Ritu Kulshreshtha,
Indian Institute of Technology Delhi,
India

Reviewed by:

Hyun Jung Park,
University of Pittsburgh, United States
Liu Le Ping,
Central South University, China

*Correspondence:

Yongming Xi
xym700118@163.com

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Epigenomics and Epigenetics,
a section of the journal
Frontiers in Cell and Developmental
Biology

Received: 16 August 2020

Accepted: 12 April 2021

Published: 14 June 2021

Citation:

Hu C, Liu C, Li J, Yu T, Dong J,
Chen B, Du Y, Tang X and Xi Y (2021)
Construction of Two Alternative
Polyadenylation Signatures to Predict
the Prognosis of Sarcoma Patients.
Front. Cell Dev. Biol. 9:595331.
doi: 10.3389/fcell.2021.595331

Background: Increasing evidence indicates that alternative polyadenylation (APA) is associated with the prognosis of cancers.

Methods: We obtained gene expression and APA profiles of 259 sarcoma patients from the TCGA datportal and TC3A database, respectively. The prognostic signatures, clinical nomograms, and regulatory networks were studied by integrated bioinformatics analyses. Then, the immune cell infiltration profile was obtained from the ImmCellAI. The association between APA-based signature and immune cells was studied.

Results: A total of 61 and 38 APA events were identified as overall survival (OS)- and progress free-survival (PFS)-related biomarkers, respectively. Two signatures were generated. The area under the curves (AUC) values of OS signature were 0.900, 0.928, and 0.963 over 2-, 4-, and 6-years, respectively. And the AUC values of PFS signature at 2-, 4-, and 6-years were 0.826, 0.840, and 0.847, respectively. Overall and subgroup analyses indicated that high-risk patients had a worse prognosis than low-risk patients (all p -values < 0.05). In addition, immunomics analyses indicated that there are different patterns of immune cell infiltration between low- and high-risk patients. Furthermore, two clinical-APA nomograms were established and the C-indexes were 0.813 and 0.809 for OS nomogram and PFS nomogram, respectively. Finally, two APA regulatory networks were constructed. FIP1L1-VPS26B was identified as a key regulating relationship and validated in the pan-cancer analyses.

Conclusion: In this study, we identified prognostic predictors based on APA events with high accuracy for risk stratification in sarcoma patients and uncovered interesting regulatory networks in sarcoma that could be underlying mechanisms. This study not only provides novel potential prognostic biomarkers but promote precision medicine and provide potential novel research interests for immunotherapy.

Keywords: alternative polyadenylation, sarcoma, overall survival, progress free-survival, nomogram

INTRODUCTION

Sarcomas are a heterogeneous group of mesenchymal malignancies that can develop at any age, comprising approximately 1% of all adult malignancies and 15% of pediatric malignancies (Von Mehren et al., 2018). Although the incidence of sarcoma is relatively rare, more than 10,000 patients are diagnosed with soft tissue sarcomas in the United States and 40,000 in China each year (Yang et al., 2019). Surgery, radiotherapy, and chemotherapy are three mainstream treatments for sarcoma patients that have shown progressive effects (Nussbaum et al., 2016; Albertsmeier et al., 2018; Thanindrataran et al., 2019). Unfortunately, the prognosis of sarcoma patients is still unsatisfactory due to local recurrence and distant metastases (Cipriano et al., 2020). Therefore, it is urgent to develop a reliable prognostic predictor for guiding clinical practice. Based on either clinical data, gene expression profile, or tumor-infiltrating immune cell, several prognostic models have been developed for sarcoma patients before (Callegaro et al., 2019; Huang et al., 2019; Gu et al., 2020). Nevertheless, there have not been any reliable models due to the complexity and heterogeneity of sarcoma.

Alternative polyadenylation (APA) is an important post-transcriptional regulation mechanism, which occurs in >70% of human genes (Mayr and Bartel, 2009; Hoque et al., 2013). It was found that APA plays an essential role in protein diversification, mRNA stability, mRNA nuclear export and repression of gene expression by producing mRNAs with different 3' untranslated regions (3' UTRs) and/or encoding variable protein isoforms (Edwards-Gilbert et al., 1997; Tian and Manley, 2017). Therefore, from the perspective of epigenetics, once APA is dysregulated, it will cause diverse pathological processes, such as cancer, viral infection, amyotrophic lateral sclerosis, and so on (Chen et al., 2017).

Nowadays, the deregulation of APA has caused widespread interest in cancer research, because APA generates mRNA 3' UTR isoforms with potentially different stabilities, subcellular localizations, translation efficiencies, and functions. In recent years, with the rapid development of high-throughput sequencing technology, genome-wide profiling for APA events has become a reality. In 2018, Li et al., (Xiang et al., 2018) have completed an analysis of pan-cancer analysis that helps us to understand the regulatory mechanisms and functional consequences of APA alterations in tumorigenesis. In total, 17 tumor types were studied and a series of important roles of APA in the tumor were discovered, such as gene expression regulation and cellular pathway remodeling (Xiang et al., 2018). Recently, Venkat et al. (2020) firstly performed a cancer-specific analysis and found that APA was an independent prognostic biomarker for pancreatic ductal adenocarcinoma patients. Generally, as with other well-researched post-transcriptional regulation mechanisms like alternative splicing and mRNA m6A methylation, APA also plays a vital role in the genesis, progress, and prognosis of cancers. Despite the effect of APA was preliminarily confirmed in this research, there were few attempts to study the role of APA for sarcoma, and no APA-based signature was constructed for such patients.

In the present study, a comprehensive analysis of APA events was performed based on a large cohort from the TCGA-SARC dataset. The prognostic value of APA events for sarcoma patients was uncovered, and two APA-based signatures were constructed. We further explored the potential relationship between APA signatures and clinicopathological data and developed two clinical-APA nomograms. Finally, we established the regulatory network between APA events and APA factors to elucidate the underlying mechanisms.

MATERIALS AND METHODS

Data Acquisition

The gene expression files and clinical data of TCGA-SARC were obtained from the UCSC Xena browser¹. Additionally, the Percentage of Distal polyA site Usage Index (PDUI) value was used to quantify each APA event, and the data of PDUI of TCGA-SARC was downloaded from the TC3A database (Feng et al., 2018). According to the definition in published research, the PDUI score quantifies the relative poly (A) site usage for that gene in the sample by computing the abundances of 3'-UTR long and short forms (Venkat et al., 2020). Genes favoring distal PAS usage (long 3' UTRs) have PDUI scores near 1, whereas genes favoring proximal PAS usage (short 3' UTRs) have PDUI scores near 0 (Venkat et al., 2020). To construct as credible a set of APA events as possible, only APA events that met the following criteria were included in this study: (Von Mehren et al., 2018) percentage of samples with PDUI value $\geq 75\%$; (Yang et al., 2019) average PDUI value ≥ 0.05 (Li et al., 2019; Zhang et al., 2020b). All data were downloaded from the public databases hence it was not required to obtain additional ethical approval for our study.

Survival Analysis and Enrichment Analysis

To comprehensively understand the role of APA event in sarcoma, two primary endpoints were studied, including overall survival (OS) and progress-free survival (PFS). According to the PDUI value, all patients were stratified into low- and high-PDUI groups by the median cut. Then, the univariate Cox analysis was performed to identify prognostic APA events for sarcoma patients, including OS- and PFS-related APA events. To obtain the robust prognostic APA events, only APA events with a p -value < 0.01 in the univariate Cox analysis were selected for further analyses (Zhu et al., 2018). To further understand the mechanisms involved in APA events affecting the prognosis of patients, the parent genes of identified APA events were then incorporated into Gene Ontology (GO) enrichment analysis. The enrichment analyses were performed in the Metascape².

Establishment and Evaluation of APA-Based Signatures

To avoid overfitting among prognostic APA events, a machine learning algorithm called Least absolute shrinkage and

¹<https://xenabrowser.net/>

²<http://metascape.org>

selection operator (LASSO) was used to filter significant APA events. Finally, the multivariate Cox analysis was performed to identify independent prognostic APA events. A robust prognostic signature is valuable for prognostic prediction, clinical management, and accurate clinical trial. Previous research indicated that APA events can serve as effective prognostic biomarkers for survival prediction and that the power of APA events exceeded clinical covariates (Venkat et al., 2020; Zhang et al., 2020b). Therefore, based on corresponding independent prognostic APA events, two prognostic signatures were established, including OS and PFS signatures. To show the discrimination of signatures, a time-dependent receiver operating characteristic (ROC) curve with area under the curve (AUC) was generated (Heagerty et al., 2000). In addition, according to the median of risk score, all patients were stratified into low- and high-risk groups, and the survival curve with a log-rank test was used to show the distinct prognosis between low- and high-risk groups.

To confirm the stability of APA-based signatures in different subgroups, subgroup analyses were performed. Kaplan-Meier survival curves of low- and high-risk patients in several clinical subgroups, including age, sex, histological type, metastatic status, tumor site, surgical margin status, and multifocal status, were constructed. A log-rank test was used to compare the

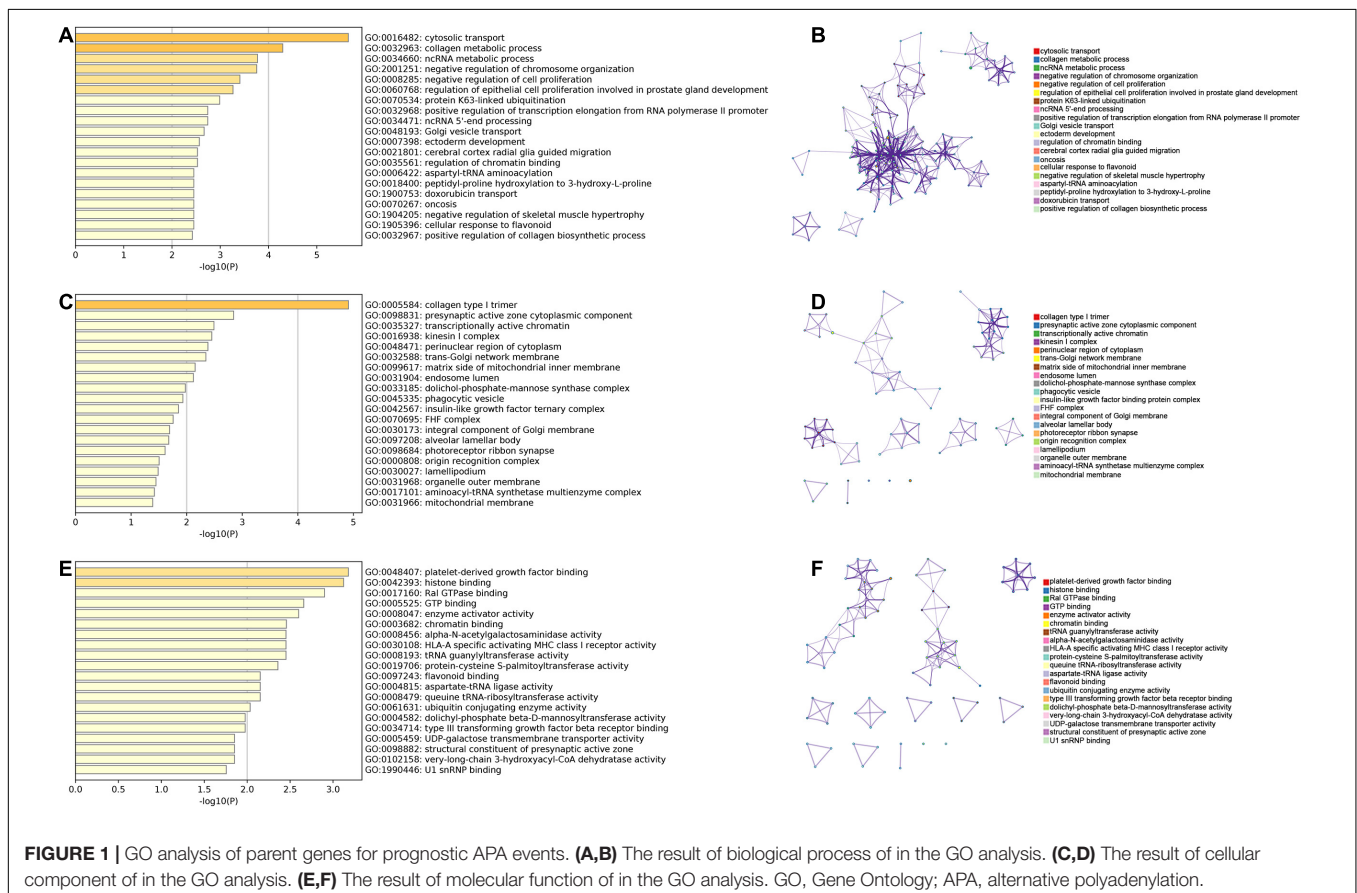
prognostic difference between patients in low- and high-risk groups.

GSVA Analysis and Immune Cell Infiltration

Gene set variation analysis (GSVA) is a non-parametric and unsupervised method that is commonly used to estimate the variation of the pathway and biological process activity in expression cohort samples (Hänzelmann et al., 2013). The gene set of “c2.cp.kegg.v7.1.symbols” was downloaded from MSigDB database for running GSVA analysis. Adjusted *P* with value less than 0.05 was considered as statistically significance. Previous studies reported that post-transcriptional regulation mechanism plays an important role in the formation of the tumor microenvironment (Li et al., 2019; Yi et al., 2020; Zhang et al., 2020a). Therefore, we further elucidated the association between immune cell infiltration and APA signatures. The immune cell infiltration profile was obtained from the ImmCellAI (Miao et al., 2020). The difference between low- and high-risk groups were evaluated by Wilcoxon tests.

Development of Clinical-APA Nomogram for Sarcoma Patients

Nomogram is one of the effective tools for clinical practice, which was widely used as a prognostic device for cancer



patients (Iasonos et al., 2008). Therefore, to confirm that APA-based signatures are an independent prognostic predictor for sarcoma and to develop two clinical-APA nomograms, the clinical data, including age, sex, race, tumor site, histological type, distant metastatic status, postoperative radiotherapy, pharmacotherapy, multifocal status, and surgical margin resection status were extracted for further analyses. The univariate Cox analysis was used to identify prognostic variables, and variables with a $p < 0.05$ were incorporated into the multivariate Cox analysis. Then, two nomograms were developed by incorporating several independent predictors. A concordance index (C-index) was used to show the discrimination of nomograms, and the calibration curve was selected to show the calibration of nomograms (Iasonos et al., 2008).

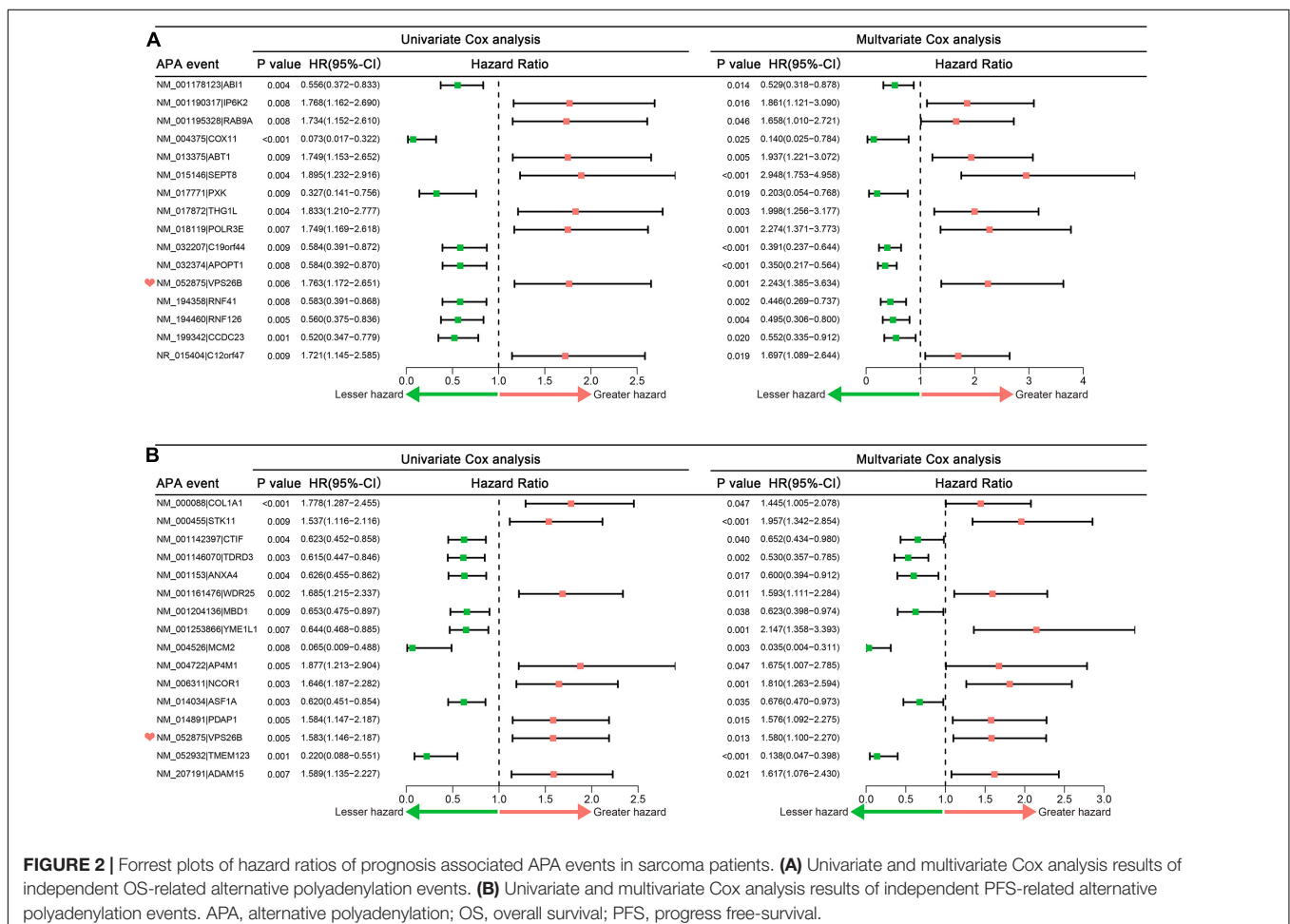
Correlation Network Between APA and APA Factors

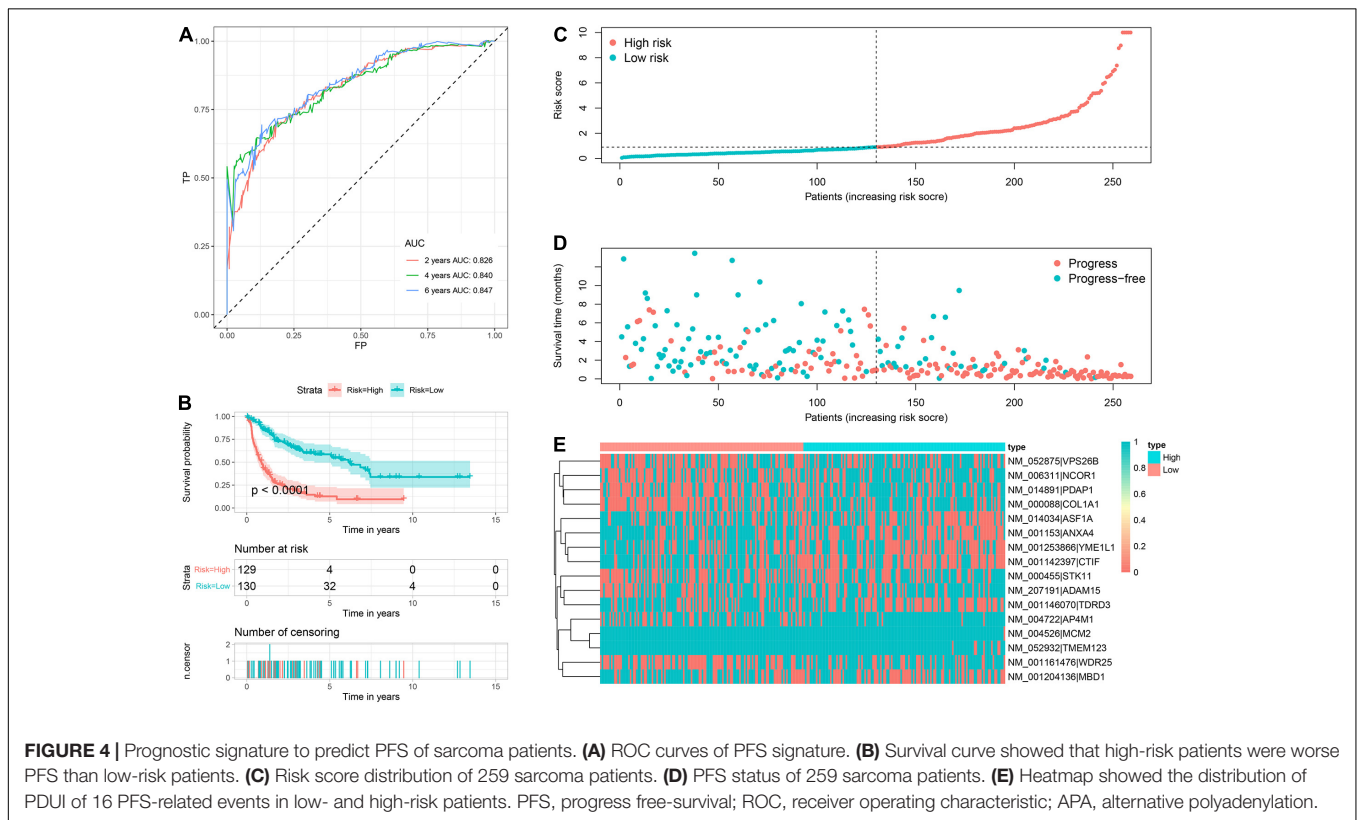
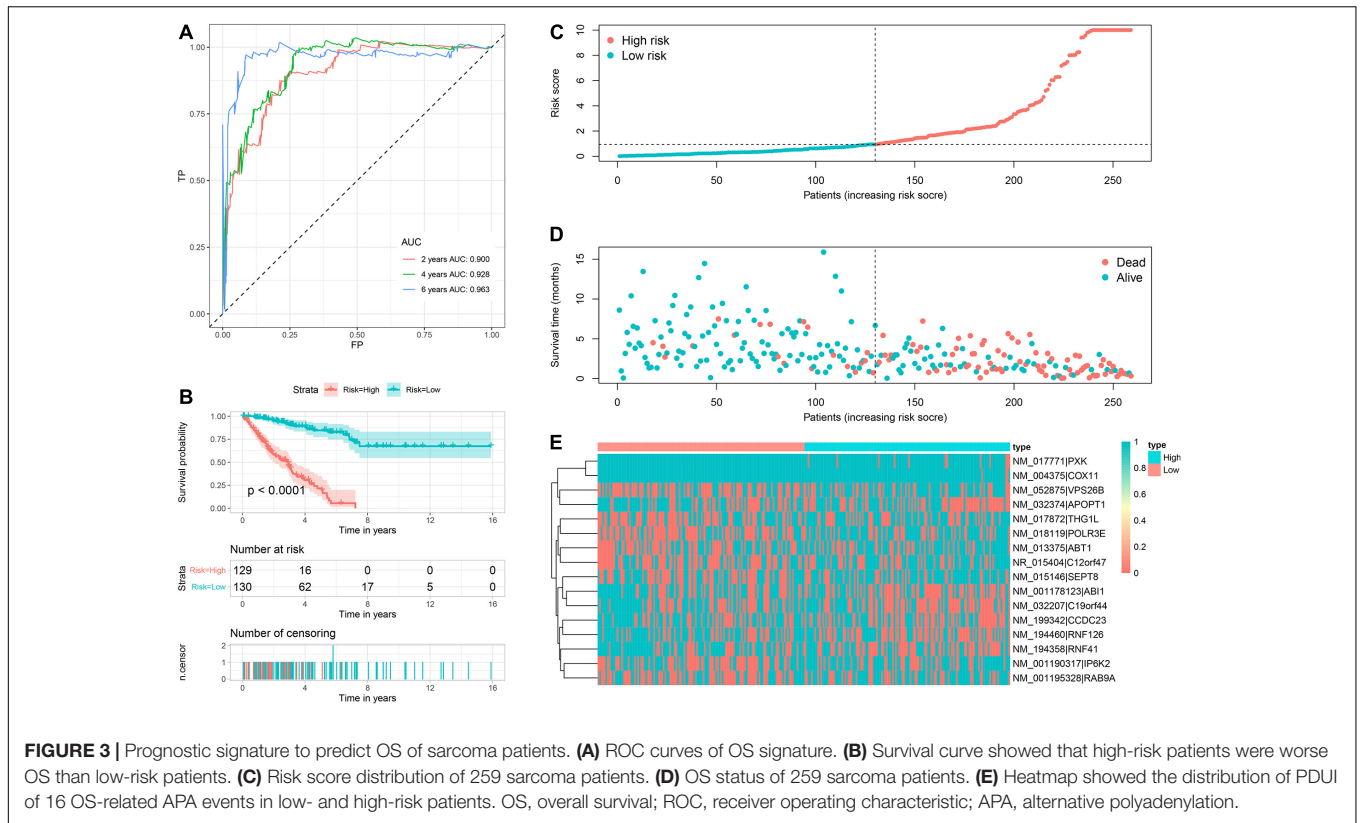
The 3' end-processing machinery is composed of multiple protein factors and several APA factors were confirmed as vital regulatory factors for APA. Based on a published articles, 28 core APA factors were included (Tang et al., 2016; Qiu et al., 2017; Xiang et al., 2018; Chatrikhi et al., 2019; Jia et al., 2019;

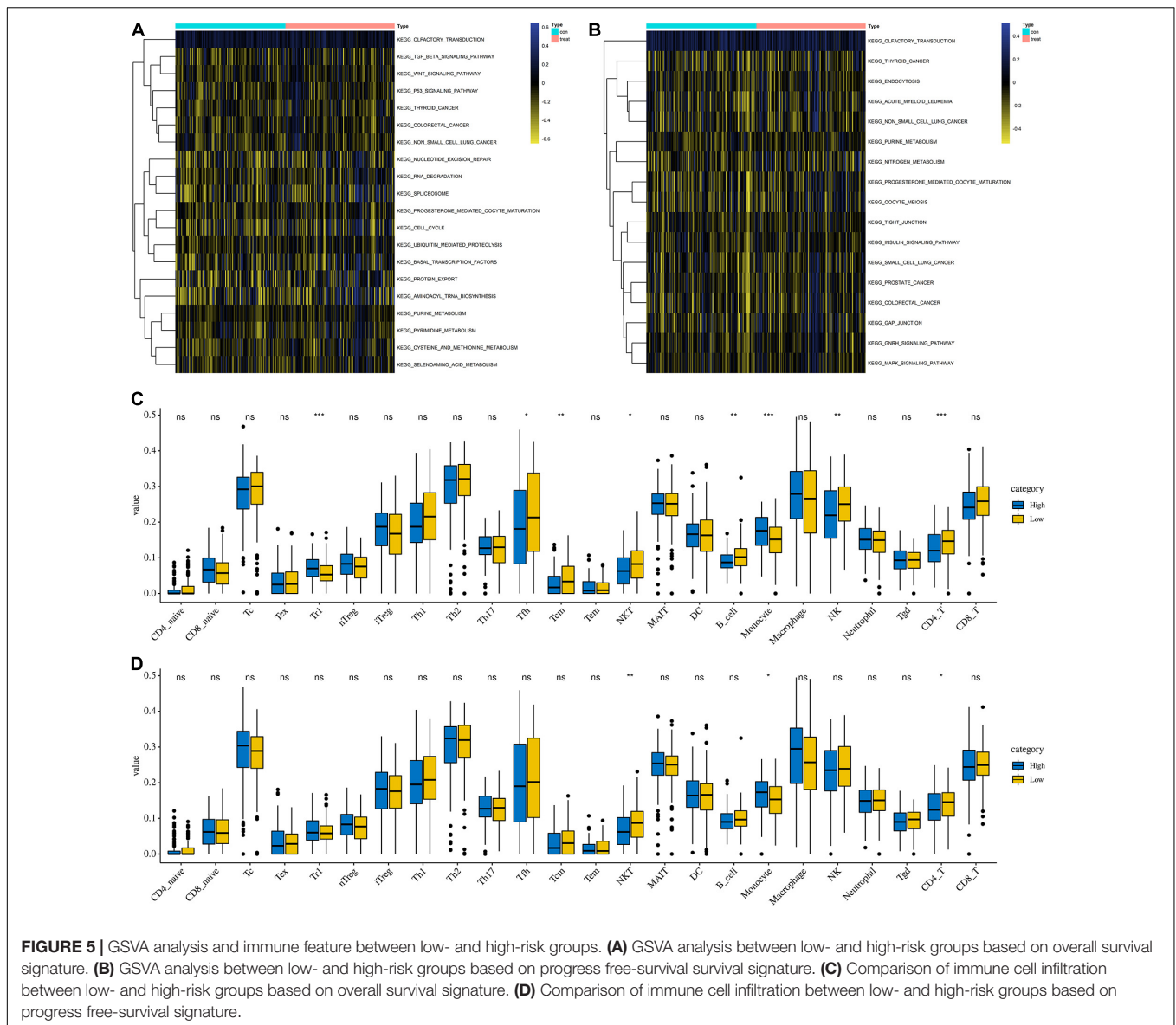
Zhang et al., 2019b). The expression of 28 APA factors was downloaded from the UCSC Xena browser (see text footnote 1). The prognostic role of all APA factors was investigated, and OS- and PFS-related APA factors were used for further correlation analyses. Pearson correlation analysis was used to determine the correlation coefficient between the PDUI of APA events and the expression of APA factors and to identify the potential regulatory networks between them ($|r| > 0.2$ and $p < 0.05$). The regulatory networks was visualized by Cytoscape 3.7.2 (Shannon et al., 2003).

Statistical Analyses

In the present study, all statistical analyses were performed with R software (version 3.6.1). Except for the special instructions, a p -value < 0.05 (two sides) was considered as statistically significant. Univariate, LASSO, and multivariate Cox analyses were used to select independent prognostic APA events by “survival” and “glmnet” packages. The “survivalROC” package was used to develop time-dependent ROC curves, and corresponding AUC values were generated simultaneously. The survival curve was generated by a “survminer” package. In addition, Pearson correlation analysis was used to identify the







potential regulatory network between APA events and APA factors ($|r| > 0.2$ and $p < 0.05$).

RESULTS

Overview of APA Events Profiling in Sarcoma

According to the aforementioned criterion, a total of 259 primary sarcoma patients were included in our study. The mean age was 60.71 ± 14.59 , and the average follow up time was 3.26 years (range: 0.04–15.56 years). In total, 98 patients died during the follow-up duration and 153 patients progressed. For APA events, 8864 APA events were detected for the TCGA-SARC cohort in TC3A database. A total of 2179 APA events were excluded because more than 25% of patients lacked PDUI of these APA

events or the average of PDUI value < 0.05 . Finally, 6685 APA events were used for further analyses.

Identification of Prognosis-Related APA Events and Enrichment Analysis

According to the median of PDUI, all patients were stratified as low- and high-groups for each APA event. The univariate Cox analysis indicated that 61 and 38 APA events were OS- and PFS-related biomarkers, respectively ($p < 0.01$) (**Supplementary Tables 1, 2**). The enrichment of the GO analysis is illustrated in **Figure 1**, which showed that specific GO categories were significantly related to sarcomas, like cytosolic transport, collagen metabolic process and negative regulation of chromosome organization.

Furthermore, LASSO analysis excluded 22 OS-related APA events and 7 PFS-related APA events

(Supplementary Figures 1A–D). Finally, 16 and 16 APA events were confirmed as independent OS- and PFS-related biomarkers, respectively (Figures 2A,B). Interestingly, we found that the APA event of VPS26B was the overlapping independent prognostic APA event between OS and PFS.

Construction of APA-Based Signatures

Based on corresponding independent prognostic APA events, two prognostic signatures were constructed to predict the OS and PFS, respectively. The AUC values of OS signature for 2-, 4-, and 6-years were 0.900, 0.928, and 0.963, respectively (Figure 3A). All AUC values were up to 0.900, indicating the great discrimination of this signature. In addition, the AUC values of PFS signature for 2-, 4-, and 6-years were 0.826, 0.840, and 0.847, respectively, which also suggested favorable discrimination (Figure 4A). According to the median of the risk score, all patients were stratified into low- and high-risk groups. Log-rank tests showed that the patients in the high-risk group has worse OS and PFS than the low-risk patients (Figures 3B, 4B). The distribution of risk score (Figures 3C, 4C), prognostic status (Figures 3D, 4D), and PDUI of each sample (Figures 3E, 4E) were visualized to facilitate the understanding of prognostic signatures. Furthermore, we compared the distribution

of clinical covariates across high- and low- risk groups (Supplementary Figure 2). For the OS signature, the distribution of multifocal indicator, margin status, and metastasis were significantly different between two risk groups (Supplementary Figure 2A). Additionally, for the PFS signature, the distributions of age, margin status, and metastasis were significantly different between two risk groups (Supplementary Figure 2B).

The Differences of GSVA Analysis and Immune Cells Between High- and Low-Risk Groups

To explore the functions of differentially expressed parental genes in APA, GSVA analysis was conducted to predict the possible functions of APA events ($p < 0.05$). Two heatmaps were used to show differences between high- and low-risk groups (Figures 5A,B). The wnt signaling pathway, p53 signaling pathway, RNA degradation, and nucleotide excision repair were enriched in the high-risk group, which played a vital role in tumorigenesis. Meanwhile, we found that other malignant tumors were also enriched in the high-risk group, including colorectal cancer, prostate cancer, and small cell lung cancer, which suggested that the APA signature was

TABLE 1 | Overall survival analysis of APA signature and clinical data for sarcoma.

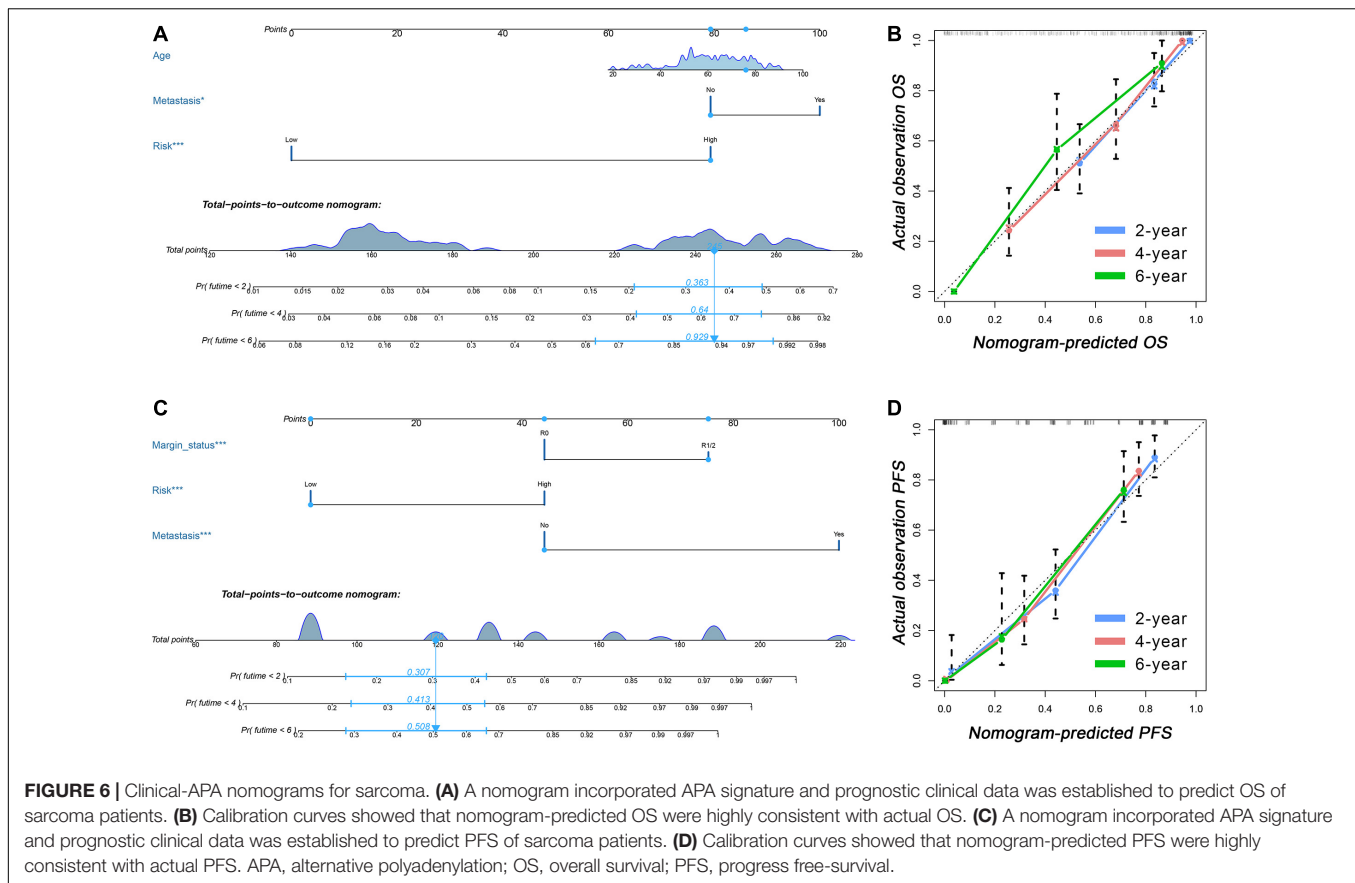
	Univariate Cox analysis			Multivariate Cox analysis		
	HR	95% CI	P-value	HR	95% CI	P-value
Risk (Low)	0.093	0.054–0.159	0.000	0.091	0.038–0.217	0.000
Age	1.020	1.005–1.036	0.010	1.025	1.002–1.049	0.032
Sex (Male)	0.855	0.572–1.277	0.443			
Race						
Asian						
Black	1.085	0.132–8.924	0.939			
White	0.791	0.108–5.773	0.817			
Histological						
DLP						
LMS	0.842	0.512–1.386	0.500			
MFS	0.703	0.328–1.507	0.364			
Other	0.739	0.319–1.710	0.479			
UPS	0.901	0.481–1.691	0.746			
Tumor site (Other)	1.229	0.792–1.906	0.357			
Metastasis (Yes)	3.014	1.834–4.954	0.000	2.410	1.335–4.349	0.004
Margin status (R1/2)	2.554	1.668–3.910	0.000	1.416	0.786–2.550	0.247
Multifocal indicator (Yes)	2.404	1.502–3.847	0.000	1.085	0.517–2.278	0.830
Radiotherapy (Yes)	0.988	0.619–1.579	0.961			
Pharmacotherapy (Yes)	1.382	0.815–2.341	0.230			

APA, alternative polyadenylation; HR, hazard ratio; CI, confidence interval; DLP, dedifferentiated liposarcoma; LMS, leiomyosarcoma; MFS, myxofibrosarcoma; UPS, undifferentiated pleomorphic sarcoma.

TABLE 2 | Progress free-survival analysis of APA signature and clinical data for sarcoma.

	Univariate Cox analysis			Multivariate Cox analysis		
	HR	95% CI	P-value	HR	95% CI	P-value
Risk (Low)	0.251	0.178–0.356	0.000	0.221	0.138–0.354	0.000
Age	1.006	0.995–1.018	0.267			
Sex (Male)	1.096	0.797–1.506	0.573			
Race						
Asian						
Black	3.146	0.402–24.609	0.275			
White	2.954	0.412–21.170	0.281			
Histological						
DLP						
LMS	0.926	0.619–1.386	0.708			
MFS	0.730	0.395–1.349	0.315			
Other	0.807	0.421–1.547	0.519			
UPS	0.870	0.528–1.433	0.584			
Tumor site (Other)	0.990	0.706–1.388	0.952			
Metastasis (Yes)	6.001	3.948–9.122	0.000	6.412	4.012–10.247	0.000
Margin status (R1/2)	2.202	1.565–3.099	0.000	2.589	1.634–4.100	0.000
Multifocal indicator (Yes)	1.972	1.307–2.975	0.001	1.567	0.850–2.887	0.150
Radiotherapy (Yes)	1.108	0.769–1.597	0.581			
Pharmacotherapy (Yes)	1.437	0.929–2.222	0.103			

APA, alternative polyadenylation; HR, hazard ratio; CI, confidence interval; DLP, dedifferentiated liposarcoma; LMS, leiomyosarcoma; MFS, myxofibrosarcoma; UPS, undifferentiated pleomorphic sarcoma.



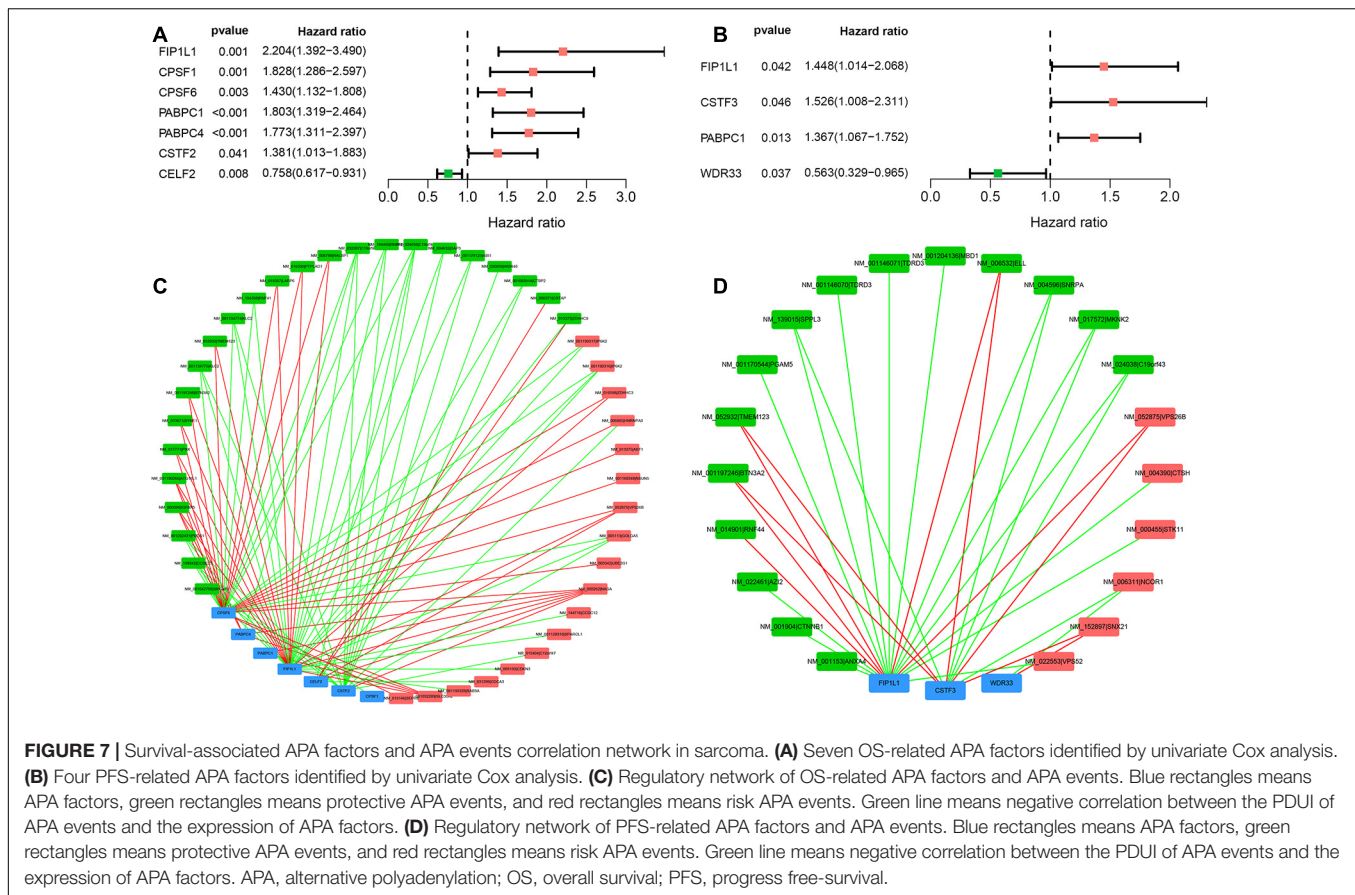
positively related to the process of these kinds of tumors. Additionally, the difference in immune cell infiltration were also observed. Among 24 types of immune cells, the NKT cell, monocyte, and CD4 T cell were significantly different between low- and high-risk groups no matter if they were the in OS signature and PFS signature (Figures 5C,D). Moreover, the filtration of Tr1 cells, Tfh cells, Tcm cells, B cells, and NK cells was distinct between low OS group and high OS group (Figure 5C).

Subgroup Analyses of Signatures in Different Subgroups

To study the ability of prognostic signatures in different clinical subgroups, the Kaplan-Meier survival curves in several subgroups were constructed (Supplementary Figures 3, 4). For the OS signature, in all 16 subgroups, survival analyses showed that patients in the low-risk group have a favorable OS than high-risk patients (Supplementary Figures 3A–H). Log-rank tests in all subgroups were statistically significant (all p -values < 0.05). Similarly, for the PFS signature, subgroup analyses indicated that low-risk patients had a better prognosis than high-risk patients (all p -values < 0.05) (Supplementary Figures 4A–H). Generally, subgroup analyses showed the stability and robustness of APA-based signatures, which further verified that APA events can serve as satisfactory prognostic predictors.

Development of Two Clinical-APA Nomograms

The results of the univariate Cox analysis indicated that APA signature, age, metastasis status, margin status, and multifocal indicator were OS-related variables ($p < 0.05$) (Table 1), while APA signature, metastasis status, margin status, and multifocal indicator were PFS-related variables (Table 2). We incorporated the above significant variables into the multivariate Cox analyses. The results indicated that both APA-based OS signature and APA-based PFS signature were robust prognostic predictors and were independent of clinical data (Tables 1, 2). Meanwhile, age and metastasis status were confirmed as independent OS-related variables (Table 1), while metastasis and margin status were independent PFS-related variables (Table 2). Two nomograms were developed to predict the OS and PFS, respectively (Figures 6A,C). In the nomogram, values for individual patients are located along the variable axes, and a line is drawn upward to the point axis to determine the number of points assigned for each variable. There is a total point line at the bottom of the nomogram, and each variable score is summed to give the total points. Then, a vertical line is drawn from the total point scale to the bottom three lines to obtain the prognosis at 2-, 4-, and 6-years. For example, in the nomogram of PFS, a patient with low risk, no distant metastasis, and a margin status of R1/2 would have a total score of 120, and his 2-year, 4-year, and 6-year PFS rates would be 30.7, 41.3, and 50.8%, respectively. The



C-index for OS nomogram was 0.813 and 0.809 for the PFS nomogram, which means that both nomograms have favorable discrimination. The calibration curves for 2-, 4-, and 6-years indicated that the nomogram-predicted outcomes were highly consistent with actual observation outcomes, no matter whether in OS or PFS nomogram (Figures 6B,D).

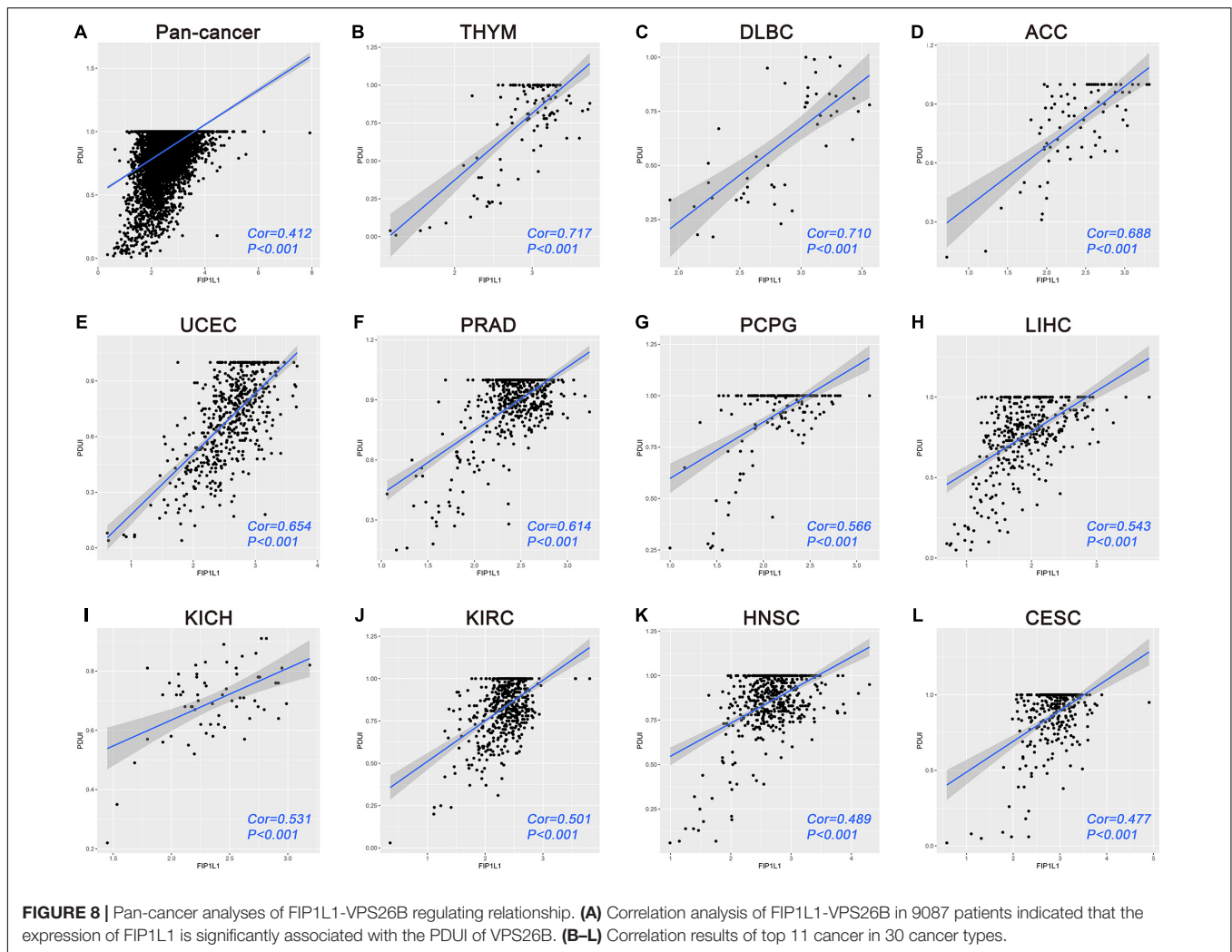
A Network of Prognostic APA Events and APA Factors

According to the univariate Cox analysis, 61 and 38 APA events were OS- and PFS-related biomarkers, respectively. In addition, the expression of 28 core APA factors was extracted. In total, seven APA factors were identified as OS-related APA factors, and four APA factors were identified as PFS-related APA factors (Figures 7A,B). The correlation between APA events and APA factors was studied. Finally, two APA-APA factors regulatory networks were established (Figures 7C,D). The regulatory between APA events and APA factors were initially elucidated. The first network contained seven APA factors and 42 APA events to show the regulatory mechanism of OS-related APA events (Figure 7C). Another network including three APA factors and 21 APA events was established to show the regulatory mechanism of PFS-related APA events (Figure 7D). Throughout the two networks, we can find that one APA factor can regulatory more than one APA event. Additionally, one APA

event can be regulated by more than one APA factor even two opposite regulations.

Pan-Cancer Analyses of FIP1L1-VPS26B Regulating Relationship

In two regulatory networks, one interesting regulating relationship presented itself for our attention. According to the survival analyses, APA events of VPS26B were identified as overlapping independent prognostic APA events between OS and PFS. In addition, FIP1L1 was shown to be significantly associated with both OS and PFS. More importantly, in the regulatory network, the expression of FIP1L1 was significantly related to the PDUI of VPS26B. Hence, we speculated this regulating relationship may have an important role in malignancy to comprehensively understand this regulating relationship in malignancy. We performed a pan-cancer analysis based on the UCSC Xena browser and TC3A database. A total of 9087 patients were included, including 30 cancer types. The Pearson correlation analysis indicated that the expression of FIP1L1 is significantly associated with the PDUI of VPS26B (Figure 8A). Furthermore, we analyzed the correlation of FIP1L1-VPS26B in specific cancers. Among 29 types (sarcoma was excluded), only four tumor cohorts showed that there was no significant correlation between FIP1L1 and VPS26B. The correlation coefficient is the highest in the TCGA-THYM



cohort (cor = 0.717), followed by TCGA-DLBC (cor = 0.710) (Figures 8B–L).

DISCUSSION

Alternative polyadenylation is a RNA-processing mechanism that generates distinct 3' termini on mRNAs and other RNA polymerase II transcripts (Tian and Manley, 2017). It is widespread across all eukaryotic species and is recognized as a major mechanism of gene regulation (Tian and Manley, 2017). The role of APA in human cancers is only beginning to be appreciated, which shows the potentially robust predictive value for tumor patients. More importantly, compared with genomic data only, APA signature may have a better predictive ability in patients' prognosis (Xia et al., 2014; Feng et al., 2018).

In the present study, we mainly focused on the profiling of the prognostic value of APA events to explore the utilization of APA signatures in predicting the outcome of sarcoma patients. In total, 259 primary sarcoma patients from TC3A database and TCGA dataportal were included. A total of 61 and 38 APA

events were determined as OS- and PFS-related biomarkers, respectively. Two 16 APA-based signatures were built and showed favorable prognostic predictors, with all AUC values above 0.900 for OS signature and 0.820 for the PFS signature. Combined with independent prognostic clinical variables, two APA-clinical nomograms were developed and showed satisfactory discrimination and calibration. Finally, combined survival and correlation network analyses between APA events and APA factors, our research uncovered the underlying mechanism of APA events involved in patient prognosis. To our knowledge, the present study is the first study to establish APA signatures for predicting the survival of sarcoma patients.

Identifying effective biomarker and constructing an ideal prognostic signature or nomogram has long been the focus of oncologists, which can individually predict the specific outcomes to guide the management of tumor patients. At present, a great number of biomarkers have been reported, and several prognostic models for sarcoma were constructed (Benassi et al., 2015; Callegaro et al., 2017, 2019; Yang et al., 2017; Huang et al., 2019; Raut et al., 2019; Zhang et al., 2019a; Gu et al., 2020; Hu et al., 2020; Zhu et al., 2020). For example, MS

Benassi et al. (2015) reported that the expression of IGFBP7, considered a tumor stroma marker in mesenchymal-derived cells, was highly prognostic in poor metastasis-free survival for soft tissue sarcoma. Additionally, clinicopathological variables, ncRNA data, or immune cells were also confirmed as predictors and used to develop prognostic models (Callegaro et al., 2017, 2019; Yang et al., 2017; Huang et al., 2019; Raut et al., 2019; Zhang et al., 2019a; Gu et al., 2020; Hu et al., 2020; Zhu et al., 2020). Nevertheless, clinical practice is not optimistic. Current research mainly focused on the prognostic role of clinical data or gene level, overlooking the transcriptional level. As one of the important post-transcriptional regulatory mechanisms, APA events have shown a potentially robust predictive value for tumor patients (Venkat et al., 2020; Zhang et al., 2020b). Our study confirmed that APA events are valuable predictors for sarcoma patients, no matter whether in general patients or several subgroups. In the mechanism, APA play a vital role in gene regulation and diverse cellular processes, including mRNA metabolism, protein diversity, and protein localization (Tian and Manley, 2017).

In total, 31 APA events, including 15 APA events for OS signature only, 15 APA events for PFS signature only, and 1 APA event for both OS and PFS signatures, were used to construct two signatures. Among the corresponding genes of these APA events, most of which were associated with tumorigenesis and progression. For example, by activating the EMT and non-canonical WNT signaling, ABI1 can drive the tumorigenesis of prostate cancer (Nath et al., 2019). Additionally, dysregulation of ABI1 was confirmed associated with the prognosis of gastric cancer, epithelial ovarian cancer, and breast cancer (Cui et al., 2010; Wang et al., 2011; Zhang et al., 2015). Another widely studied gene is MCM2, which is a vital initiation factor for DNA replication in humans. It also presents in the nucleus and is overexpressed in proliferating cells. The prognostic value of MCM2 was confirmed in pancreatic cancer, lung cancer, multiple myeloma, and oral cancer (Ramnath et al., 2001; Torres-Rendon et al., 2009; Deng et al., 2020; Quan et al., 2020). Significantly, the APA event of VPS26B was determined as an overlapping independent prognostic biomarker for sarcoma patients. Nevertheless, few studies reported the role of VPS26B or related regulation in cancers. Future studies may focus on its prognostic value in other tumors and their regulatory role in tumors.

In our research, APA and APA factor regulatory networks were also constructed. Various of regulating relationships between them were identified. Interestingly, one core regulating relationship was confirmed, and the same trend was also confirmed in the pan-cancer analysis. As a novel regulator for APA events, FIP1L1 can regulate the 3'UTR lengthening of leukemia-associated genes, including NRAS, BAALC, and MAPKAPK3 (Davis et al., 2018). Interestingly, both overexpression and knockdown of FIP1L1 are harmful to leukemia cells, demonstrating that mild alteration of gene expression may dramatically impact on cell fitness (Davis et al., 2018). In addition to FIP1L1, eight APA factors were included incorporated into regulatory networks. The mechanism of part of factors has been preliminarily elucidated. As a vital

cleavage/polyadenylation factor, CSTF2 can shorten the length of 3'UTR RAC1 in human urothelial carcinoma of the bladder by mediating slow transcriptional elongation at RAC1 (Chen et al., 2018). Moreover, F6 was considered as a vulnerability target for breast cancer patients (Binothman et al., 2017). Despite this research, the regulatory mechanism between APA and APA factors in sarcoma remains unclear. More than 100 potential regulating relationships were detected in our research, which pointed out the directions for future research.

Although the strict bioinformatic and statistical methods were used and the prognostic value of APA events in sarcoma patients have been discovered in our research, there are some limitations. First, due to the relative rarity of sarcoma, only bioinformatic analyses were used in this study, and no further experimental analysis based on clinical samples was performed to validate our results. Second, external validation is vital for clinical application of prognostic signatures or nomograms. Unfortunately, no available independent cohort can be obtained from other databases. Thirdly, no available normal sample can be incorporated into our study. Therefore, the diagnostic value and the potential carcinogenic effect of APA events cannot be studied in this research. Finally, the data used in the study were obtained from public datasets from which the available clinical data is limited and incomplete. Therefore, some prognostic variables, such as tumor grade and size, were not available and were not analyzed in the present study.

CONCLUSION

Our data revealed the prognostic value of survival-related APA events for sarcoma. Some key APA factors might play essential roles in tumor initiation and progression by regulating the corresponding APA events. Our findings might offer a new prospect for effective therapies targeted at APA events for sarcoma.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://xenabrowser.net/> and tc3a.org.

AUTHOR CONTRIBUTIONS

CH, YX, and CL performed the data analysis and wrote the manuscript. JL, CL, JD, and XT contributed to the data analysis and manuscript revision. CH, TY, BC, and YD contributed to literature search and data extraction. CH and YX conceived and designed the study. All authors have read and approved the final version of the manuscript.

FUNDING

This study was supported by the Taishan Scholar Project of Shandong Province, China (No. ts20190985).

ACKNOWLEDGMENTS

We are sincerely acknowledge the contributions from the TCGA project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2021.595331/full#supplementary-material>

Supplementary Figure 1 | LASSO analysis of OS- and PFS-related APA events. **(A,C)** Dotted vertical lines were drawn at the optimal values by using the minimum criteria. **(B,D)** LASSO coefficient profiles of the candidate OS **(B)** and PFS **(D)**-related alternative polyadenylation events. A coefficient profile plot was produced against the log λ sequence. LASSO, least absolute shrinkage and

selection operator; APA, alternative polyadenylation; OS, overall survival; PFS, progress free-survival.

Supplementary Figure 2 | The distribution of clinical covariates across low- and high-risk groups. **(A)** Overall survival signature; **(B)** Progress free-survival signature. DLP, dedifferentiated liposarcoma; LMS, leiomyosarcoma; MFS, myxofibrosarcoma; UPS, undifferentiated pleomorphic sarcoma.

Supplementary Figure 3 | Subgroup analyses of OS signature. Survival curves showed that high-risk patients were significant worse OS than low-risk patients in subgroups of age **(A)**, sex **(B)**, cancer histological type **(C,D)**, metastatic status **(E)**, tumor site **(F)**, margin status **(G)**, and multifocal indicators **(H)**. OS, overall survival.

Supplementary Figure 4 | Subgroup analyses of PFS signature. Survival curves showed that high-risk patients were significant worse PFS than low-risk patients in subgroups of age **(A)**, sex **(B)**, cancer histological type **(C,D)**, metastatic status **(E)**, tumor site **(F)**, margin status **(G)**, and multifocal indicators **(H)**. PFS, progress-free survival.

REFERENCES

- Albertsmeier, M., Rauch, A., Roeder, F., Hasenhütl, S., Pratschke, S., Kirschneck, M., et al. (2018). External beam radiation therapy for resectable soft tissue sarcoma: a systematic review and meta-analysis. *Ann. Surg. Oncol.* 25, 754–767. doi: 10.1245/s10434-017-6081-2
- Benassi, M. S., Pazzaglia, L., Novello, C., Quattrini, I., Pollino, S., Magagnoli, G., et al. (2015). Tissue and serum IGFBP7 protein as biomarker in high-grade soft tissue sarcoma. *Am. J. Cancer Res.* 5, 3446–3454.
- Binothman, N., Hachim, I. Y., Lebrun, J.-J., and Ali, S. (2017). CPSF6 is a clinically relevant breast cancer vulnerability target: role of CPSF6 in breast cancer. *EBioMedicine* 21, 65–78. doi: 10.1016/j.ebiom.2017.06.023
- Callegaro, D., Miceli, R., Bonvalot, S., Ferguson, P. C., Strauss, D. C., van Praag, V. V. M., et al. (2019). Development and external validation of a dynamic prognostic nomogram for primary extremity soft tissue sarcoma survivors. *EClinicalMedicine* 17:100215. doi: 10.1016/j.eclinm.2019.11.008
- Callegaro, D., Miceli, R., Mariani, L., Raut, C. P., and Gronchi, A. (2017). Soft tissue sarcoma nomograms and their incorporation into practice. *Cancer* 123, 2802–2820. doi: 10.1002/cncr.30721
- Chatrikhi, R., Mallory, M. J., Gazzara, M. R., Agosto, L. M., Zhu, W. S., Litterman, A. J., et al. (2019). RNA binding protein CELF2 regulates signal-induced alternative polyadenylation by competing with enhancers of the polyadenylation machinery. *Cell Rep.* 28, 2795–2806. doi: 10.1016/j.celrep.2019.08.022
- Chen, W., Jia, Q., Song, Y., Fu, H., Wei, G., and Ni, T. (2017). Alternative polyadenylation: methods, findings, and impacts. *Genom. Proteom. Bioinform.* 15, 287–300. doi: 10.1016/j.gpb.2017.06.001
- Chen, X., Zhang, J.-X., Luo, J.-H., Wu, S., Yuan, G.-J., Ma, N.-F., et al. (2018). CSTF2-induced shortening of the RAC1 3'UTR promotes the pathogenesis of urothelial carcinoma of the bladder. *Cancer Res.* 78:5848. doi: 10.1158/0008-5472.CAN-18-0822
- Cipriano, C. A., Jang, E., and Tyler, W. (2020). Sarcoma surveillance: a review of current evidence and guidelines. *J. Am. Acad. Orthop. Surg.* 28, 145–156. doi: 10.5435/jaaos-d-19-00002
- Cui, M., Yu, W., Dong, J., Chen, J., Zhang, X., and Liu, Y. (2010). Downregulation of ABI1 expression affects the progression and prognosis of human gastric carcinoma. *Med. Oncol.* 27, 632–639. doi: 10.1007/s12032-009-9260-6
- Davis, A. G., Shima, T., Wang, R., Zheng, D., Tian, B., and Zhang, D.-E. (2018). FIP1L1 regulates alternative polyadenylation of leukemia-associated genes in acute myeloid leukemia. *Blood* 132(Suppl. 1):3882. doi: 10.1182/blood-2018-99-119582
- Deng, Y., Ma, H., Hao, J., Xie, Q., and Zhao, R. (2020). MCM2 and NUSAP1 Are potential biomarkers for the diagnosis and prognosis of pancreatic cancer. *BioMed Res. Int.* 2020:8604340.
- Edwards-Gilbert, G., Veraldi, K. L., and Milcarek, C. (1997). Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res.* 25, 2547–2561. doi: 10.1093/nar/25.13.2547
- Feng, X., Li, L., Wagner, E. J., and Li, W. (2018). TC3A: the cancer 3' UTR atlas. *Nucleic Acids Res.* 46, D1027–D1030. doi: 10.1093/nar/gkx892
- Gu, H.-Y., Zhang, C., Guo, J., Yang, M., Zhong, H.-C., Jin, W., et al. (2020). Risk score based on expression of five novel genes predicts survival in soft tissue sarcoma. *Aging* 12:3807. doi: 10.18632/aging.102847
- Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 14:7. doi: 10.1186/1471-2105-14-7
- Heagerty, P. J., Lumley, T., and Pepe, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 56, 337–344. doi: 10.1111/j.0006-341x.2000.00337.x
- Hoque, M., Ji, Z., Zheng, D., Luo, W., Li, W., You, B., et al. (2013). Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat. Methods* 10, 133–139. doi: 10.1038/nmeth.2288
- Hu, Q., Zhou, S., Hu, X., Zhang, H., Huang, S., and Wang, Y. (2020). Systematic screening identifies a 2-gene signature as a high-potential prognostic marker of undifferentiated pleomorphic sarcoma/myxofibrosarcoma. *J. Cell. Mol. Med.* 24, 1010–1021. doi: 10.1111/jcmm.14814
- Huang, R., Meng, T., Chen, R., Yan, P., Zhang, J., Hu, P., et al. (2019). The construction and analysis of tumor-infiltrating immune cell and ceRNA networks in recurrent soft tissue sarcoma. *Aging* 11, 10116–10143. doi: 10.18632/aging.102424
- Iasonos, A., Schrag, D., Raj, G. V., and Panageas, K. S. (2008). How to build and interpret a nomogram for cancer prognosis. *J. Clin. Oncol.* 26, 1364–1370. doi: 10.1200/jco.2007.12.9791
- Jia, Q., Nie, H., Yu, P., Xie, B., Wang, C., Yang, F., et al. (2019). HN1HNRNPA1-mediated 3' UTR length changes of contributes to cancer- and senescence-associated phenotypes. *Aging* 11, 4407–4437. doi: 10.18632/aging.102060
- Li, Z.-X., Zheng, Z.-Q., Wei, Z.-H., Zhang, L.-L., Li, F., Lin, L., et al. (2019). Comprehensive characterization of the alternative splicing landscape in head and neck squamous cell carcinoma reveals novel events associated with tumorigenesis and the immune microenvironment. *Theranostics* 9:7648. doi: 10.7150/thno.36585
- Mayr, C., and Bartel, D. P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138, 673–684. doi: 10.1016/j.cell.2009.06.016
- Miao, Y. R., Zhang, Q., Lei, Q., Luo, M., Xie, G. Y., Wang, H., et al. (2020). ImmuCellAI: a unique method for comprehensive T-Cell subsets abundance prediction and its application in cancer immunotherapy. *Adv. Sci.* 7:1902880. doi: 10.1002/advs.201902880
- Nath, D., Li, X., Mondragon, C., Post, D., and Kotula, L. (2019). Abi1 loss drives prostate tumorigenesis through activation of EMT and non-canonical WNT signaling. *Cell Commun. Signal.* 17:120.
- Nussbaum, D. P., Rushing, C. N., Lane, W. O., Cardona, D. M., Kirsch, D. G., Peterson, B. L., et al. (2016). Preoperative or postoperative radiotherapy versus

- surgery alone for retroperitoneal sarcoma: a case-control, propensity score-matched analysis of a nationwide clinical oncology database. *Lancet Oncol.* 17, 966–975. doi: 10.1016/s1470-2045(16)30050-x
- Qiu, F., Fu, Y., Lu, C., Feng, Y., Wang, Q., Huo, Z., et al. (2017). Small nuclear ribonucleoprotein polypeptide a-mediated alternative polyadenylation of STAT5B during Th1 cell differentiation. *J. Immunol.* 199, 3106–3115. doi: 10.4049/jimmunol.1601872
- Quan, L., Qian, T., Cui, L., Liu, Y., and Si, C. (2020). Prognostic role of minichromosome maintenance family in multiple myeloma. *Cancer Gene Ther.* 27, 819–829. doi: 10.1038/s41417-020-0162-2
- Ramnath, N., Hernandez, F. J., Tan, D. F., Huberman, J. A., and Bepko, G. (2001). MCM2 Is an independent predictor of survival in patients with non-small-cell lung cancer. *J. Clin. Oncol.* 19, 4259–4266. doi: 10.1200/jco.2001.19.22.4259
- Raut, C. P., Callegaro, D., Miceli, R., Barretta, F., Rutkowski, P., Blay, J.-Y., et al. (2019). Predicting survival in patients undergoing resection for locally recurrent retroperitoneal sarcoma: a study and novel nomogram from TARPSWG. *Clin. Cancer Res.* 25, 2664–2671. doi: 10.1158/1078-0432.ccr-18-2700
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Tang, S., Patel, A., and Krause, P. R. (2016). Herpes simplex virus ICP27 regulates alternative pre-mRNA polyadenylation and splicing in a sequence-dependent manner. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12256–12261. doi: 10.1073/pnas.1609695113
- Thanindratarn, P., Dean, D. C., Nelson, S. D., Hornicek, F. J., and Duan, Z. (2019). Advances in immune checkpoint inhibitors for bone sarcoma therapy. *J. Bone Oncol.* 15:100221. doi: 10.1016/j.jbo.2019.100221
- Tian, B., and Manley, J. L. (2017). Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.* 18:18. doi: 10.1038/nrm.2016.116
- Torres-Rendon, A., Roy, S., Craig, G. T., and Speight, P. M. (2009). Expression of Mcm2, geminin and Ki67 in normal oral mucosa, oral epithelial dysplasias and their corresponding squamous-cell carcinomas. *Br. J. Cancer* 100, 1128–1134. doi: 10.1038/sj.bjc.6604967
- Venkat, S., Tisdale, A. A., Schwarz, J. R., Alahmari, A. A., Maurer, H. C., Olive, K. P., et al. (2020). Alternative polyadenylation drives oncogenic gene expression in pancreatic ductal adenocarcinoma. *Genome Res.* 30, 347–360. doi: 10.1101/gr.257550.119
- Von Mehren, M., Randall, R. L., Benjamin, R. S., Boles, S., Bui, M. M., Ganjoo, K. N., et al. (2018). Soft tissue sarcoma, version 2.2018, NCCN clinical practice guidelines in oncology. *J. Natl. Compr. Canc. Netw.* 16, 536–563.
- Wang, C., Tran-Thanh, D., Moreno, J. C., Cawthorn, T. R., Jacks, L. M., Wang, D. Y., et al. (2011). Expression of Abl interactor 1 and its prognostic significance in breast cancer: a tissue-array-based investigation. *Breast Cancer Res. Treat.* 129, 373–386. doi: 10.1007/s10549-010-1241-0
- Xia, Z., Donehower, L. A., Cooper, T. A., Neilson, J. R., Wheeler, D. A., Wagner, E. J., et al. (2014). Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat. Commun.* 5:5274.
- Xiang, Y., Ye, Y., Lou, Y., Yang, Y., Cai, C., Zhang, Z., et al. (2018). Comprehensive characterization of alternative polyadenylation in human cancer. *J. Natl. Cancer Inst.* 110, 379–389. doi: 10.1093/jnci/djx223
- Yang, L., Forker, L., Irlam, J. J., Pillay, N., Choudhury, A., and West, C. M. L. (2017). Validation of a hypoxia related gene signature in multiple soft tissue sarcoma cohorts. *Oncotarget* 9, 3946–3955. doi: 10.18632/oncotarget.23280
- Yang, Z., Zheng, R., Zhang, S., Zeng, H., Li, H., and Chen, W. (2019). Incidence, distribution of histological subtypes and primary sites of soft tissue sarcoma in China. *Cancer Biol. Med.* 16, 565–574. doi: 10.20892/j.issn.2095-3941.2019.0041
- Yi, L., Wu, G., Guo, L., Zou, X., and Huang, P. (2020). Comprehensive analysis of the PD-L1 and immune infiltrates of m(6)A RNA methylation regulators in head and neck squamous cell carcinoma. *Mol. Ther. Nucleic Acids* 21, 299–314. doi: 10.1016/j.omtn.2020.06.001
- Zhang, B., Wu, Q., Li, B., Wang, D., Wang, L., and Zhou, Y. L. (2020a). m⁽⁶⁾A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. *Mol. Cancer* 19:53. doi: 10.1186/s12943-020-01170-0
- Zhang, J., Pan, Z., Yang, J., Yan, X., and Lyu, J. (2019a). A nomogram for determining the disease-specific survival in ewing sarcoma: a population study. *BMC Cancer* 19:667. doi: 10.1186/s12885-019-5893-9
- Zhang, J., Tang, L., Chen, Y., Duan, Z., Xiao, L., Li, W., et al. (2015). upregulation of abelson interactor protein 1 predicts tumor progression and poor outcome in epithelial ovarian cancer. *Hum. Pathol.* 46, 1331–1340. doi: 10.1016/j.humpath.2015.05.015
- Zhang, Y., Wang, Y., Li, C., and Jiang, T. (2020b). Systemic analysis of the prognosis-associated alternative polyadenylation events in breast cancer. *Front. Genet.* 11:1352. doi: 10.3389/fgene.2020.590770
- Zhang, Z., So, K., Peterson, R., Bauer, M., Ng, H., Zhang, Y., et al. (2019b). Elav-mediated exon skipping and alternative polyadenylation of the Dscam1 gene are required for axon outgrowth. *Cell Rep.* 27, 3808–3817. doi: 10.1016/j.celrep.2019.05.083
- Zhu, X.-F., Zhu, B.-S., Wu, F.-M., and Hu, H.-B. (2018). DNA methylation biomarkers for the occurrence of lung adenocarcinoma from TCGA data mining. *J. Cell. Physiol.* 233, 6777–6784. doi: 10.1002/jcp.26531
- Zhu, Z., Jin, Z., Zhang, H., Zhang, M., and Sun, D. (2020). Integrative clustering reveals a novel subtype of soft tissue sarcoma with poor prognosis. *Front. Genet.* 11:69. doi: 10.3389/fgene.2020.00069

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Hu, Liu, Li, Yu, Dong, Chen, Du, Tang and Xi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.