Genome-wide profiling of prognosis-related alternative splicing signatures in sarcoma

Weifeng Hong^{1#}, Weicong Zhang^{2#}, Renguo Guan^{3#}, Yuying Liang^{1#}, Shixiong Hu^{3#}, Yayun Ji¹, Mouyuan Liu¹, Hai Lu², Min Yu³, Liheng Ma¹

¹Department of Medical Imaging, The First Affiliated Hospital of Guangdong Pharmaceutical University, Guangzhou 510080, China; ²Department of Orthopaedics, The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai 519000, China; ³Department of General Surgery, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510000, China

Contributions: (I) Conception and design: W Hong; (II) Administrative support: W Zhang, R Guan; (III) Provision of study materials or patients: Y Liang, S Hu; (IV) Collection and assembly of data: Y Ji, M Liu; (V) Data analysis and interpretation: H Lu, M Yu, L Ma; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

"These authors contributed equally to this work.

Correspondence to: Liheng Ma. Department of Medical Imaging, The First Affiliated Hospital of Guangdong Pharmaceutical University, Guangzhou 510080, China. Email: maliheng@gdpu.edu.cn; Min Yu. Department of General Surgery, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510000, China. Email: yuminzhongda@163.com; Hai Lu. Department of Orthopaedics, The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai 519000, China. Email: luhai04@qq.com.

Background: Sarcomas (SARCs) are rare malignant tumors with poor prognosis. Increasing evidence has suggested that aberrant alternative splicing (AS) is strongly associated with tumor initiation and progression. We considered whether survival-related AS events might serve as prognosis predictors and underlying targeted molecules in SARC treatment.

Methods: RNA-Seq data of the SARC cohort were downloaded from The Cancer Genome Atlas (TCGA) database. Survival-related AS events were selected by univariate and multivariate Cox regression analyses. Metascape was used for constructing a gene interaction network and performing functional enrichment analysis. Then, prognosis predictors were established based on statistically significant survival-related AS events and evaluated by receiver operator characteristic (ROC) curve analysis. Finally, the potential regulatory network was analyzed via Pearson's correlation between survival-related AS events and splicing factors (SFs).

Results: A total of 3,610 AS events and 2,291 genes were found to be prognosis-related in 261 SARC samples. The focal adhesion pathway was identified as the most critical molecular mechanism corresponding to poor prognosis. Notably, several prognosis predictors based on survival-related AS events showed excellent performance in prognosis prediction. The area under the curve of the ROC of the risk score was 0.85 in the integrated predictor. The splicing network proved complicated regulation between prognosis-related SFs and AS events. Also, driver gene mutations were significantly associated with AS in SARC patients.

Conclusions: Survival-related AS events may become ideal indictors for the prognosis prediction of SARCs. Corresponding splicing regulatory mechanisms are worth further exploration.

Keywords: Alternative splicing (AS); sarcomas; prognosis; splicing factor (SF); driver gene

Submitted Aug 31, 2019. Accepted for publication Sep 06, 2019. doi: 10.21037/atm.2019.09.65 **View this article at:** http://dx.doi.org/10.21037/atm.2019.09.65

Page 2 of 17

Introduction

Alternative splicing (AS) is a ubiquitous regulatory mechanism that enables eukaryotic cells to generate vast and diverse mature mRNAs from a limited number of genes. AS remarkably increases both transcriptome and proteome complexity via different precursor mRNA splicing processes (1,2). Throughout the whole developmental stage, AS plays a crucial role in the regulation of cell differentiation, cell-specific functions, and cell senescence to further control growth and metabolism (3-7). Given this, multiple pathological processes, including cancers, are also produced because of AS deregulation. Aberrant AS may initiate loss-of-function in tumor suppressors, the activation of oncogenes or cancer pathways to affect tumorigenesis, development, and prognosis through abnormal proliferation, altered epigenetics, and resistance to apoptosis and chemotherapy (8-12). Therefore, AS events have been proposed as ideal biomarkers for cancer diagnosis and prognosis and even potential targets for new anti-carcinogen discoveries (8,10,12,13).

Sarcomas (SARCs), which arise from mesenchymal tissues, are malignant tumors comprising less than 10% of all cancers (14). Statistically, approximately 10,700 diagnoses and 3,800 deaths per year in the US are relevant to SARC, while in Europe, SARC represents only 1-2% of all cancers in adults with an overall annual incidence of 5.6 cases per 100,000 adults (15,16). Compared to the low incidence in adults, SARCs account for a higher percentage of overall cancer morbidity and mortality in children and young adults (ages 20-39). SARCs are a group of heterogeneous mesenchymal malignancies with over eighty histological subtypes, which are broadly categorized as soft-tissue SARCs and bone SARCs (17). Clinically, surgical resection, radiotherapy, and chemotherapy adopted singly or together be effective in treating specific localized SARCs. However, for many SARC patients who have chemotherapy resistance or metastatic foci, new therapeutic regimens are needed.

Along with the discovery of an increasing number of pivotal genes and biological mechanisms underlying carcinogenesis and disease progression, correlative pharmacologic and genetic therapies based on these are becoming progressively essential in routine treatment. In turn, molecular characteristics gradually guide the nosological classification of SARCs. Hence, it is imperative to explore molecular mechanisms more deeply in the prognosis of SARCs. Several studies have been undertaken to expand the genomic diversity of oncogenic drivers and identify potential therapeutic targets in SARC through genome-scale analyses of mRNAs, microRNAs, and protein, and alterations in DNA sequences, methylation, and copy numbers (18,19). AS is worth further exploring as a new field, as it is an indispensable part in the molecular mechanisms of SARCs.

In recent years, many studies of AS have mainly focused on distinguishing "SARC-specific" AS events by molecular experiments (20-23). However, comprehensive analyses of survival-related AS events are still scarce in SARC. Powered by high-throughput RNA-seq, the investigation of cancerrelated AS events and counterpart molecular mechanisms in a relatively larger population has become obtainable with the rapid accumulation of transcriptome data and clinical data (24,25). Through the analysis of genome-wide profiling, prognostic AS signatures have been reported in many cancers, such as non-small cell lung cancer, colorectal cancer, prostate adenocarcinoma, esophageal carcinoma, and gastrointestinal pan-adenocarcinomas (26-31).

The Cancer Genome Atlas (TCGA) project provides a rich source for the research of AS in cancer with mass data of transcript isoforms, corresponding genes, and the survival status of tumor patients (24). We systematically analyzed genome-wide AS events in the SARC cohort from TCGA using a series of bioinformatics methods to obtain a certain amount of survival-related AS events and revealed the latent splicing networks. More importantly, high-performance prognostic predictors were constructed to evaluate the potential of AS events in the prognosis prediction of SARC. These findings, for the first time, supply systematic novel insights into the latent functions of SARC-specific AS events.

Methods

AS event curation from TCGA RNA-seq data

RNA-seq data from the TCGA SARC cohorts were available from the TCGA data portal (https://tcga-data.nci. nih.gov/tcga/).

TCGASpliceSeq (http://bioinformatics.mdanderson. org/TCGASpliceSeq) is a web-based resource that includes the mRNA AS patterns of tumors from TCGA (32). We used this tool to analyze the different AS patterns of TCGA SARC samples. The SpliceSeq program aligns to the gene's protein-coding transcripts in the ensemble gene database and conveniently generates a unified splice graph. Then, the splice graph is aligned to the TCGA sample reads, and summary statistics are collected for each transcript isoform.



Figure 1 Illustrations of the seven subtypes of alternative splicing. (A) An example of a full-length pre-mRNA with six exons. Mutually exclusive exons (MEd), retained intron (RI), exon skip (ES), alternate terminator (AT), alternate promoter (AP), alternate acceptor site (AA), and alternate donor site (AD). (B) The number of preliminary alternative splicing events and involved genes obtained from the TCGA sarcoma cohort. The red bar indicates the number of alternative splicing events. The green bar indicates the number of involved genes.

SpliceSeq can calculate the percent-spliced-in (PSI) value for each AS event, which supplies a clear view of splice junctions and the proportion of exons. The PSI value is a standard, intuitive ratio for quantifying seven types of AS events (33): exon skip (ES), retained intron (RI), mutually exclusive exons (MEs), alternate promoter (AP), alternate terminator (AT), alternate donor site (AD), and alternate acceptor site (AA). Only AS events with a PSI value >75% and a standard deviation >0.1 were included in the present analysis. A schematic diagram illustrating the seven types of AS events are shown in *Figure 1A*.

Survival analysis

A total of 261 SARC patients with detailed overall survival (OS) information were included in this study. The summary characteristics of these patients are depicted in *Table S1*. Univariate Cox regression analysis was performed to identify survival-related AS events in SARC with P values less than 0.05. Then, multivariate Cox regression analysis was further employed to determine whether splicing events were independent prognostic factors. Finally, the risk score for the seven types of AS events was calculated for survival prediction. Furthermore, predictive models of Kaplan-Meier (K-M) survival analysis and receiver operating characteristic (ROC) curves were constructed to evaluate the clinical value of the risk scores. The area under the ROC curve (AUC) was used to assess the predictive

accuracy of the prognostic models.

UpSet plot and gene interaction network

The intersections between the different types of AS events were also investigated by the UpSet package of R software. UpSet plots are a more intuitive and superior alternative to Venn diagrams for visualizing intersecting sets (34). After selection by univariate Cox regression analysis, the candidate genes of prognosis-related AS events with P value <0.005 were submitted to Metascape (http://metascape.org) for protein-protein interaction (PPI) analysis (35,36). The molecular complex detection (MCODE) algorithm was then applied to identify neighborhoods where proteins were densely connected in the PPI network. Subsequently, Gene Ontology (GO) enrichment analysis was performed for each MCODE network to assign biological interpretations to the network components via Metascape. We further set the P value to less than 0.05 in the univariate Cox regression analysis to obtain genes of survival-related AS events for the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and GO analyses performed by Metascape. The ggplot2 package of R software was used to visualize the KEGG pathway enrichment analysis.

Splicing correlation network construction

The association between survival-related AS events and

Page 4 of 17

 Table 1 Baseline Characteristics according to TCGA Clinical data

 about SARC (n=261)

Characteristics	No. of patient (%)					
Age (y)						
<60	116 (44.4)					
≥60	145 (55.6)					
Gender						
Male	119 (45.6)					
Female	142 (54.4)					
Event						
Death	99 (37.9)					
Alive	162 (62.1)					
Overall survival						
≥365	216 (82.8)					
<365	45 (17.2)					
Race						
White	226 (86.6)					
Asian	6 (2.3)					
Black or African American	18 (6.9)					
Not reported	11 (4.2)					
Histological type						
Dedifferentiated liposarcoma	59 (22.6)					
Desmoid tumor	2 (0.8)					
Leiomyosarcoma	105 (40.2)					
Malignant peripheral nerve sheath tumors	9 (3.4)					
Myxofibrosarcoma	25 (9.6)					
Synovial sarcoma	10 (3.8)					
Undifferentiated pleomorphic sarcoma	51 (19.5)					
Tumor site						
Chest	13 (5.0)					
Head and neck	5 (1.9)					
Lower abdominal/pelvic	16 (6.1)					
Lower extremity	73 (28.0)					
Not available	1 (0.4)					
Ovary	1 (0.4)					
Retroperitoneum/upper abdominal	99 (37.9)					
Superficial trunk	12 (4.6)					
Upper extremity	12 (4.6)					
Uterus	29 (11.1)					

splicing factors (SFs) was further investigated. A list of known SFs was extracted from the SpliceAid 2 (www.introni. it/spliceaid.html) database (37). The level 3 expression profiles of SFs were downloaded from the TCGA database and converted into transcripts per million (TPM), which is considered a more reasonable data format for RNA-seq (38). After that, the log-rank test was conducted to select survival-related SFs. Pearson's correlation test was further applied to assess the association between survival-related AS events and SFs.

DriverDBv2 (a database for human cancer driver gene research; http://driverdb.tms.cmu.edu.tw/driverdbv2/) was used to discover driver genes and their mutational types (39). Pearson's correlation test was performed to investigate the correlation between the expression of driver genes and AS events. Also, *t*-test was conducted to find the statistical significance between gene mutation and AS events. All statistical analyses were performed using R/Bioconductor (version 3.4.3) and reported P values of <0.05 were considered statistically significant (P values were two-sided).

Results

Integrated mRNA splicing event profiles in the SARC cohort of TCGA

All splicing events were calculated by SpliceSeq. A total of 9,673 genes with 40,184 AS events were detected in 261 sarcomatous samples, including 15,311 ESs in 6,038 genes, 8,287 ATs in 3,616 genes, 7,837 APs in 3,156 genes, 3,197 AAs in 2,295 genes, 2,816 ADs in 1,987 genes, 2,572 RIs in 1,741 genes and 164 MEs in 163 genes (*Figure 1B*). Overall, the results showed that one gene might have an average of 4.1 AS events. Among these splicing subtypes, ES was the main subtype of AS events, while MEs were rare in sarcomas. Also, only a small fraction of all AS events (1,438 out of 40,184) were novel splices. The summary clinical characteristics of the SARC patients are shown in *Table 1*.

Survival-associated AS events in the SARC cohort

Univariate Cox analyses of OS were applied to find survivalassociated AS events in the SARC cohort. The results showed that a total of 3,610 AS events from 2,291 genes were significantly associated with patient OS (P<0.05), including 203 RIs from 185 genes, 252 AAs from 240 genes, 252 Ads from 229 genes, 758 APs from 444 genes, 776 ATs from 423 genes, 1,359 ESs from 1,083 genes, and 10 MEs

from 10 genes (*Figure 2A*). As depicted in the UpSet plot, most genes had only one prognosis-related AS subtype, while some genes had two or more AS subtypes associated with survival. Among these survival-related genes, 293 genes had more than one type of AS event, but none of them had the seven AS subtypes simultaneously (*Figure 2B*).

Gene network construction and functional enrichment analysis

To reveal the interaction effects among the genes of prognosis-related AS events, a PPI network was constructed by Metascape (Figure 2C). After the molecular complex detection analysis, thirteen MCODE components identifying neighborhoods where proteins were densely connected were extracted from the PPI network (Figure 2D). Also, the GO enrichment analysis of each MCODE network is depicted in Table 2, in which the top three GO results are displayed. Among these thirteen MCODEs, the most significant part was MCODE2, which included twelve genes: RNPS1, POLR2E, SNRNP27, SRSF2, SNRPN, NCBP2, HNRNPF, HNRNPD, HNRNPH1, CPSF3, PPIE, and SRSF7. As shown in Table 2, these genes involved in MCODE2 were enriched in reactions with bulged adenosine as nucleophile, mRNA splicing via the spliceosome, and RNA splicing via transesterification reactions.

To analyze the biological classification of the survivalrelated AS genes, KEGG pathway and GO enrichment analyses were performed. The KEGG pathway analysis revealed that "focal adhesion" was the most significantly enriched by these genes, followed by "endocytosis" (*Figure 3A*). GO analysis revealed that "cell adhesion molecule binding," "adherens junction" and "lipid modification" were the most significantly enriched (*Figure 3B*). As showed in the hierarchical cluster tree, the highest Kappa-statistical similarities existed in the "adherens junction" among gene memberships (*Figure 3C*), while the "cell adhesion molecule binding" was the most statistically significant (*Figure 3D*).

Construction of AS-based prognostic predictors

To clinically evaluate the prognostic value of AS in SARC, we constructed prognostic multivariate Cox regression models based on the top twenty significantly prognosisrelated events in each AS subtype (*Figure 4*). The hazard ratio of each AS subtype was processed logarithmically to base 10. Owing to the quantitative differences in survivalassociated AS events, we chose different p values to construct best risk scores with regard to each AS subtype, which included all splicing types (P<0.0001), AA (P<0.001), AP (P<0.001), ES (P<0.0001), RI (P<0.005), AD (P<0.005), and AT (P<0.001). However, the risk score of the ME type could not be constructed because AS events of the ME type were insufficient. The details of the AS events used for constructing the seven risk scores are shown in Table S1. SARC patients were classified into low-risk and high-risk groups according to the median value of the risk score as the cut-off (Figure 5). The Kaplan-Meier analysis proved that the survival time was diverse between the two groups in the Cox regression model of each AS subtype, excluding the ME type (P value <0.0001). Then, we integrated the six types of AS events and developed a comprehensive prognostic predictor, which also showed moderate performance in predicting prognosis (P value <0.0001) (Figure 6A). Notably, the ROC curve analysis confirmed that the integrated prognostic predictor displayed more of an ideal performance than the other models based on a single specific type of AS event. The AUC of the ROC curve for the integrated prognostic predictor was 0.85, followed by the AP and ES models with AUCs of 0.82 and 0.80, respectively (Figure 6B).

Survival-associated SFs and driver genes

A total of 71 SFs from the SpliceAid 2 database were selected for survival analysis, and 26 survival-related SFs were identified (P<0.001). Furthermore, Spearman's test was performed to explore the correlations between the PSI values of top significant AS events and prognosis-related SFs. Only relationships with the significant association (P<0.001) were included in the correlative network (Figure 7). As depicted in the network, most SFs were positively correlated with risk survival prognostic AS events while negatively correlated with protective survival prognostic AS events (*Figure* 7G). For example, the high expression of the SFs SYNCRIP and HNRNPC were significantly associated with poor survival prognosis (*Figure 7A*, D). In concordance, as shown in dot plots (Figure 7B,E), risk survival prognostic AS events such as the AT of PTK2B and AT of HMGA2 were positively correlated with the SFs SYNCRIP and HNRNPC, respectively. In contrast, protective survival prognostic AS events such as the AT of PPIE and AP of OXR1 were negatively correlated with the SFs SYNCRIP and HNRNPC (Figure 7C,F) respectively, which showed

Page 6 of 17



Figure 2 Statistics for survival-related alternative splicing events and the construction of a PPI network. (A) The number of survival-related splicing events. The green bar indicates the number of survival-related splicing events. The green bar indicates the number of survival-related genes. (B) The UpSet plot of survival-related alternative splicing events. One gene can have up to four subtypes of survival-related splicing events. (C) PPI network of survival-related genes via Metascape analysis (http://metascape.org). A unique color is assigned to each MCODE network. The larger circles are more significant in the network. (D) Thirteen MCODE networks were extracted from the PPI network. MCODE, molecular complex detection; PPI network, protein-protein interaction network.

MCODE	GO	Description	Log10 (P value)
MCODE_1	GO:0044391	Ribosomal subunit	-8.6
MCODE_1	GO:0000313	Organellar ribosome	-8.5
MCODE_1	GO:0005761	Mitochondrial ribosome	-8.5
MCODE_2	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	-21.7
MCODE_2	GO:0000398	mRNA splicing, via spliceosome	-21.7
MCODE_2	GO:0000375	RNA splicing, via transesterification reactions	-21.7
MCODE_3	GO:0000980	RNA polymerase II distal enhancer sequence-specific DNA binding	-5.2
MCODE_3	GO:2001233	Regulation of apoptotic signaling pathway	-5.1
MCODE_3	GO:0001158	Enhancer sequence-specific DNA binding	-5
MCODE_5	GO:0045296	Cadherin binding	-6.8
MCODE_5	GO:0050839	Cell adhesion molecule binding	-6.1
MCODE_5	GO:0016055	Wnt signaling pathway	-6
MCODE_6	GO:0001325	Formation of extrachromosomal circular DNA	-9.1
MCODE_6	GO:0090737	Telomere maintenance via telomere trimming	-9.1
MCODE_6	GO:0090656	t-circle formation	-9.1
MCODE_7	hsa03010	Ribosome	-6
MCODE_9	GO:0000407	Phagophore assembly site	-11.6
MCODE_9	GO:0006501	C-terminal protein lipidation	-10.2
MCODE_9	GO:0005776	Autophagosome	-9.7
MCODE_10	GO:0006281	DNA repair	-4.3
MCODE_11	GO:0006406	mRNA export from nucleus	-7
MCODE_11	GO:0071427	mRNA-containing ribonucleoprotein complex export from nucleus	-7
MCODE_11	GO:0071426	Ribonucleoprotein complex export from nucleus	-6.8
MCODE_12	GO:0010565	Regulation of cellular ketone metabolic process	-6.4
MCODE_12	GO:0043401	Steroid hormone mediated signaling pathway	-6.4
MCODE_12	GO:0009755	Hormone-mediated signaling pathway	-6
MCODE_13	GO:0008380	RNA splicing	-5.1

poor survival prognosis under high expression as well.

A list of driver genes was generated by at least four bioinformatics tools using the DriverDBv2, which is a database for cancer driver genes and mutations. The results showed that a total of 14 driver genes were identified (*Figure S1A*). In the mutation profile of the driver genes, mutations in TP53, RB1, and ATRX occurred in most of the SARC cohort from TCGA. Regarding mutation classification, truncating and missense mutations were the two main types for the driver genes TP53, RB1, ATRX, and MUC2 (*Figure S1B,C*). The correlations between the expression of the driver genes and PSI values of the top 30 survival-related AS events were evaluated. The results revealed that expression of PLCG1, PAK2, and RB1 was closely related to most of the top 30 survival-associated AS events with a high correlation coefficient (*Figure S2A*). Also, we conducted *t*-test to confirm the statistical significance of AS events and mutation profiles and found that the mutation



Page 9 of 17



Figure 4 Forest plots of the HRs of the top twenty survival-associated alternative splicing events for AA, AD, AP, AT, ES, and RI subtypes and the whole sarcoma cohort. Only the top ten survival-associated splicing events of the ME type were included owing to insufficient splicing events. The log of HR and 95% CI is taken as the base on the e value (ln). The splicing events are either positively survival-associated (HR <0) or negatively survival-associated (HR >0). The circles represent the HR in the plots, and the horizontal bars represent the 95% CIs. HR, hazard ratio; CIs, confidence intervals; AA, alternate acceptor site; AD, alternate donor site; AP, alternate promoter; AT, alternate terminator; ES, exon skip; RI, retained intron; ME, mutually exclusive exon.



Hong et al. Prognosis-related alternative splicing signatures in SARC

© Annals of Translational Medicine. All rights reserved.

Page 10 of 17

sarcoma patients, and the bottom displays the heatmap of the PSI values for several splicing events. Colors from blue to red represent increasing PSI values from low to high.

PSI, percent-spliced-in. ME, mutually exclusive exon.

Annals of Translational Medicine, Vol 7, No 20 October 2019



Figure 6 Kaplan-Meier and ROC curve analyses of the prognostic predictors. (A) Kaplan-Meier survival analysis of the prognostic predictors based on risk scores. The full lines represent the low-risk groups, while the dotted lines represent the high-risk groups. (B) ROC curve analysis for all prognostic predictors constructed with different splicing subtypes. ROC curves, receiver operating characteristic curves.

status of TP53 and RB1 significantly correlated with most of the top 30 survival-associated AS events (*Figure S2B*).

Discussion

Considerable investigations have confirmed that by altering the composition of the proteome, the perturbation in mRNA AS patterns trigger the initiation and progression of diverse cancers (11,12). In the present study, we screened reasonable quantities of prognosis-related AS events and correlative SFs in SARC patients. Corresponding SF-AS event networks were also set up to elucidate potential regulatory mechanisms for the abnormalities in SARC. The genes of survival-related AS events were found to be enriched in several metabolic pathways and interacted closely with each other, especially in dense regions of the PPI network. We also implemented correlation analyses between variant driver genes and survival-associated AS events with the intent to guide further mutual mechanistic studies. Remarkably, all prognostic predictors we constructed based on the most significant prognosisrelated AS events showed superior performance, especially the integrated prognostic predictor. Our comprehensive investigation first focused on the AS event characteristics at the genome-wide level to evaluate the potential of survivalrelated AS events as prognostic, predictive factors, and therapeutic target molecules in SARC.

A single type of pre-mRNA can be spliced in different

ways to engender similar but not identical mature mRNA patterns. Nevertheless, the proteins translated by these homologous transcript isoforms may exert analogous or completely contrary functions in normal physiological conditions. For example, polypyrimidine tract binding protein 1 (PTBP1) is expressed at high levels during the mitotic period of neural stem cells, but upon neuronal differentiation, it becomes repressed and allows the induction of PTBP2 expression (40). Also, it has been demonstrated that the ubiquitous (Mef2Da1) and the muscle-specific (Mef2Da2) splicing isoforms antagonistically regulate muscle cell differentiation (41). Accumulating evidence has proven that aberrant splicing is often interrelated with the activation of oncogenes and the inactivation of tumor-suppressor genes. In one aspect, dysregulated splicing induces the activation of oncogenes by producing novel splice variants with tumorigenic properties. MDM2-ALT1, an isoform of the oncogene murine double minute 2 (MDM2), is typically induced in response to genotoxic stress. It is pervasive in various aggressive metastatic cancers, including pediatric rhabdomyosarcoma and lymphoma (42,43). In another aspect, the imbalanced components of splice variants from the same gene can affect the related downstream regulatory network, which also contributes to tumor development. For instance, EWS-FLI1, a fusion of Ewing's sarcoma (EWS) and FLI1 genes, stimulates transcription of the proto-oncogene CCND1 encoding cyclin D1a and a less abundant but more



 \triangleleft positive correlation between the expression of SYNCRIP and the PSI value of PTK2B-83154-AT was associated with poor prognosis. (C) A negative correlation between the expression of SYNCRIP and the PSI value of PPIE-1902-AT was associated with good prognosis. (D) High expression of the splicing factor HNRNPC was significantly with poor prognosis. (F) A negative correlation between the expression of HNRNPC and the PSI value of OXR1-84842-AP was associated with good prognosis. (G) Splicing correlation network between survival-related splicing factors and splicing events in sarcomas. Most of the 26 survival-related splicing factors (purple dots) are positively (red associated with poor overall survival in sarcoma. (E) A positive correlation between the expression of HNRNPC and the PSI value of HMGA2-22880-AT was associated Figure 7 Splicing factor regulatory network. (A) High expression of the splicing factor SYNCRIP was significantly associated with poor overall survival in sarcoma. (B) lines) associated with adverse splicing events (red dots), whereas they are negatively (green lines) associated with favorable splicing events (green dots).

Ann Transl Med 2019;7(20):557 | http://dx.doi.org/10.21037/atm.2019.09.65

© Annals of Translational Medicine. All rights reserved.

oncogenic isoform, D1b.

In contrast to the EWS oncogene preferring the D1a isoform, EWS-FLI1 elevates the D1b/D1a transcript ratio, accelerating EWS cell growth (44). Vascular endothelial growth factor (VEGF), well known as a proangiogenic factor, can also confer antiangiogenic effects due to the divergent splice site choice. Its proangiogenic isoform (VEGF165) and antiangiogenic isoform (VEGF165b) are generated from proximal or distal splice site selections in exon 8, respectively. Any disturbance to the coordination of the proand antiangiogenic isoforms of VEGF can interfere with the maintenance of internal homeostasis, even resulting in angiogenic pathologies in different cancers (45). It is difficult to determine whether a gene is a tumor suppressor or oncogene since variant AS events from the same premRNA may have similar or even opposite biological functions because of minute sequence differences. Therefore, multiple functions of a versatile gene should be discriminated depending on the specific predominant AS events.

In the present study, a series of survival-related AS events were identified via high-throughput technology. It was noted that most of these survival-associated AS events were adverse prognostic signatures in SARC. We used the most significant prognosis-related genes to construct a PPI network, from which thirteen MCODEs were highlighted, indicating potential molecular complexes. Thereinto, MCODE2 was the most statistically significant with GO enrichment in reactions with bulged adenosine as a nucleophile, mRNA splicing via the spliceosome and RNA splicing via transesterification reactions. With the assistance of the spliceosome, a multi-subunit RNA-protein complex comprising small nuclear RNAs (snRNAs) and spliceosomal protein, pre-mRNA splicing encompasses two sequential transesterification reactions (branching and exon ligation) to remove introns to produce mature mRNAs with uninterrupted protein-coding sequences (46,47). In the first transesterification step of splicing, the U2 snRNP interacts directly with the pre-mRNA branch point adenosine acting as the nucleophile within the bulged duplex formed between them (48,49). Based on the above physiological process, we have enough evidence to speculate that the survival-related genes of AS enriched in MCODE2 are linked to sarcomas through intricate splicing.

According to the results of the KEGG enrichment analysis, we found that focal adhesion was the most significant interference pathway related to AS. Multifunctional focal adhesion complexes facilitating contact between the extracellular matrix and actin

cytoskeleton play an important role in controlling the cellular morphology and cytoplasmic signaling for survival, proliferation, differentiation, and motility. The uniqueness of focal adhesions rests with the cooperative connection between integrins and growth factor receptors accessing the cytoplasmic signaling network (50). As a key mediator for downstream signaling of integrins and growth factor receptors, focal adhesion kinase (FAK) has been indicated to be upregulated in many cancers (51,52). FAK plays a prominent role in tumor initiation, progression, and metastasis via its regulation of both tumor cells and their microenvironments, including tumor cell migration, invasion, angiogenesis and epithelial to mesenchymal transition (53). Moreover, it has been reported that FAK splice variants were experimentally detected in promoting cancer cell proliferation and migration in leukemia, nonsmall cell lung cancer and breast cancer (54-56).

Due to the high malignancy and mortality of SARCs, it is urgent to explore effective prognostic signatures and therapeutic targets in SARC more deeply. Some prognosisrelated molecules for SARC have been consecutively proposed in recent years, such as mRNAs (57), miRNAs (58) and lncRNAs (59,60). With advances in high-throughput RNA-seq, several researchers have illuminated the prognostic value of AS events in multiple cancers at the genome-wide level (26-31). They suggested that AS events could be preeminent hallmarks for cancer prognosis prediction and further potential therapeutic targets. Hence, we equally used high-throughput RNA-seq to integrate clinical outcome data and AS events, aiming to offer a comprehensive recognition of prognosis-related AS events in SARC. According to the high accuracy of the prediction model we constructed, the signatures of survival associated AS events are ideal for predicting the prognosis of SARC.

As requisite regulatory molecules, SFs assist the spliceosome to successfully recognize and bind to specific pre-mRNA sequences, which subsequently promotes the generation of mature mRNA (61). Hence, it is imperative to uncover the regulatory network between AS events and SFs. In our research, after screening the most significant prognosis-related SFs and AS events, we constructed an SF-AS regulatory network. Interestingly, there was a general trend that most SFs were positively correlated with adverse AS events while negatively correlated with favorable AS events. Also, as observed in the intricate network, limited SFs dominated AS in a synergistic or antagonistic manner to engender an enormous amount of AS events. SFs comprise the serine/arginine-rich proteins (SR proteins)

Page 14 of 17

and the heterogeneous nuclear ribonucleoprotein (hnRNP) families. Physiologically, these two families perform opposite functions in splicing and bind to enhancer sites and silencer sites, respectively. It has been widely reported that dysregulated expression levels of SFs with or without mutations can affect cancer pathogenesis (13). However, the potential molecular interaction and detailed regulatory network of SFs during the AS process remain unclear, and more attention should be paid to future experimental verification.

SARCs are groups of histologically heterogeneous tumors that have historically complicated driver gene discovery. With a better understanding of the molecular mechanism of SARC, an increasing number of essential driver genes have been identified, including *TP53*, *PIK3CA*, *SETD2*, *AKT1*, *FBXW7*, *FGFR1*, *FOXA2* and *PDK1* (62-65). Subsequent signal pathway changes induced by mutations in such genes lead to accelerated cancer development and progression. Also, SF genes are significantly differentially expressed between cancer and corresponding normal samples and have reduced interindividual expression variation in cancer (66).

However, few studies have been conducted on the association between driver genes and AS events. Potential driver genes were screened by a bioinformatic tool in the present study, and we confirmed that mutations in TP53, RB1, and ATRX occurred in most of the SARC cohort from TCGA. Concordantly, these three driver genes were also included in fourteen genes obtained by the exome sequencing analysis of osteosarcoma (67) and are highly recurrently mutated across sarcoma types (18). Through the correlation analyses between the expression of driver genes and top survival-related AS events, RB1 was found to have a high correlation coefficient and a significant correlation simultaneously. Our study findings built a bridge between driver genes and AS events in SARC, and the mutually concrete regulatory mechanisms still need to be further expounded.

However, several limitations should be considered. First, the number of patients included in the SARC cohort was small. Second, the prognostic predictors proposed here have not been reproducibly verified due to the lack of another independent cohort of SARC patients. Third, further experimental validation should be performed in the future.

Conclusions

In summary, for the first time, our comprehensive research

focused on extensive data on AS events in SARCs at the genome-wide level. The interpretation of aberrant AS patterns and corresponding regulatory networks may improve the management of SARCs, particularly broadening the novel field of prognosis prediction and underlying molecular-targeted therapies.

Acknowledgments

Funding: This grant of the study was from the National Natural Science Foundation of China (Grant No. 81572628) and the Guangdong Province Science and Technology plan project (Grant No. 2014A020212304). This funding made a significant contribution to study design, data interpretation, and writing.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work (including full data access, integrity of the data and the accuracy of the data analysis) in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This article does not contain any studies with human participants or animals performed by any of the authors. Therefore, informed consent is not needed.

References

- Yang X, Coulombe-Huntington J, Kang S, et al. Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. Cell 2016;164:805-17.
- Liu Y, Gonzalez-Porta M, Santos S, et al. Impact of Alternative Splicing on the Human Proteome. Cell Rep 2017;20:1229-41.
- Porter RS, Jaamour F, Iwase S. Neuron-specific alternative splicing of transcriptional machineries: Implications for neurodevelopmental disorders. Molecular and Cellular Neuroscience 2018;87:35-45.
- Gallego-Paez LM, Bordone MC, Leote AC, et al. Alternative splicing: the pledge, the turn, and the prestige: The key role of alternative splicing in human biological systems. Hum Genet 2017;136:1015-42.
- 5. Fiszbein A, Kornblihtt AR. Alternative splicing switches: Important players in cell differentiation. Bioessays

2017;39.

- Deschênes M, Chabot B. The emerging role of alternative splicing in senescence and aging. Aging Cell 2017;16:918-33.
- Baralle FE, Giudice J. Alternative splicing as a regulator of development and tissue identity. Nat Rev Mol Cell Biol 2017;18:437-51.
- Song X, Zeng Z, Wei H, et al. Alternative splicing in cancers: From aberrant regulation to new therapeutics. Semin Cell Dev Biol 2018;75:13-22.
- Siegfried Z, Karni R. The role of alternative splicing in cancer drug resistance. Curr Opin Genet Dev 2018;48:16-21.
- Pradella D, Naro C, Sette C, et al. EMT and stemness: flexible processes tuned by alternative splicing in development and cancer progression. Mol Cancer 2017;16:8.
- Climente-González H, Porta-Pardo E, Godzik A, et al. The Functional Impact of Alternative Splicing in Cancer. Cell Rep 2017;20:2215-26.
- Chen J, Weiss WA. Alternative splicing in cancer: implications for biology and therapy. Oncogene 2015;34:1-14.
- 13. Jyotsana N, Heuser M. Exploiting differential RNA splicing patterns: a potential new group of therapeutic targets in cancer. Expert Opin Ther Targets 2018;22:107-21.
- 14. Helman LJ, Meltzer P. Mechanisms of sarcoma development. Nat Rev Cancer 2003;3:685-94.
- 15. Stiller CA, Trama A, Serraino D, et al. Descriptive epidemiology of sarcomas in Europe: report from the RARECARE project. Eur J Cancer 2013;49:684-95.
- Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2009. CA Cancer J Clin 2009;59:225-49.
- 17. Doyle LA. Sarcoma classification: an update based on the 2013 World Health Organization Classification of Tumors of Soft Tissue and Bone. Cancer 2014;120:1763-74.
- Cancer Genome Atlas Research Network. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. Cell 2017;171:950-965.e28.
- 19. Barretina J, Taylor BS, Banerji S, et al. Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. Nat Genet 2010;42:715-21.
- Zhang M, Zhu B, Davie J. Alternative splicing of MEF2C pre-mRNA controls its activity in normal myogenesis and promotes tumorigenicity in rhabdomyosarcoma cells. J Biol Chem 2015;290:310-24.
- 21. Frampton GM, Ali SM, Rosenzweig M, et al. Activation of MET via diverse exon 14 splicing alterations occurs

in multiple tumor types and confers clinical sensitivity to MET inhibitors. Cancer Discov 2015;5:850-9.

- 22. Dewaele M, Tabaglio T, Willekens K, et al. Antisense oligonucleotide-mediated MDM4 exon 6 skipping impairs tumor growth. J Clin Invest 2016;126:68-84.
- Comiskey DF, Jr., Jacob AG, Singh RK, et al. Splicing factor SRSF1 negatively regulates alternative splicing of MDM2 under damage. Nucleic Acids Res 2015;43:4202-18.
- 24. Shen S, Wang Y, Wang C, et al. SURVIV for survival analysis of mRNA isoform variation. Nat Commun 2016;7:11548.
- Griffith M, Griffith OL, Mwenifumbo J, et al. Alternative expression analysis by RNA sequencing. Nat Methods 2010;7:843-7.
- Zong Z, Li H, Yi C, et al. Genome-Wide Profiling of Prognostic Alternative Splicing Signature in Colorectal Cancer. Front Oncol 2018;8:537.
- Xiong Y, Deng Y, Wang K, et al. Profiles of alternative splicing in colorectal cancer and their clinical significance: A study based on large-scale sequencing data. EBioMedicine 2018;36:183-95.
- Mao S, Li Y, Lu Z, et al. Survival-associated alternative splicing signatures in esophageal carcinoma. Carcinogenesis 2019;40:121-30.
- Lin P, He RQ, Ma FC, et al. Systematic Analysis of Survival-Associated Alternative Splicing Signatures in Gastrointestinal Pan-Adenocarcinomas. EBioMedicine 2018;34:46-60.
- Li Y, Sun N, Lu Z, et al. Prognostic alternative mRNA splicing signature in non-small cell lung cancer. Cancer Lett 2017;393:40-51.
- Huang ZG, He RQ, Mo ZN. Prognostic value and potential function of splicing events in prostate adenocarcinoma. Int J Oncol 2018;53:2473-87.
- Ryan M, Wong WC, Brown R, et al. TCGASpliceSeq a compendium of alternative mRNA splicing in cancer. Nucleic Acids Res 2016;44:D1018-22.
- Ryan MC, Cleland J, Kim R, et al. SpliceSeq: a resource for analysis and visualization of RNA-Seq data on alternative splicing and its functional impacts. Bioinformatics 2012;28:2385-7.
- Lex A, Gehlenborg N, Strobelt H, et al. UpSet: Visualization of Intersecting Sets. IEEE Trans Vis Comput Graph 2014;20:1983-92.
- 35. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systemslevel datasets. Nat Commun 2019;10:1523.

Hong et al. Prognosis-related alternative splicing signatures in SARC

Page 16 of 17

- 36. Li T, Wernersson R, Hansen RB, et al. A scored human protein-protein interaction network to catalyze genomic interpretation. Nat Methods 2017;14:61-4.
- Piva F, Giulietti M, Burini AB, et al. SpliceAid 2: a database of human splicing factors expression data and RNA target motifs. Hum Mutat 2012;33:81-5.
- Wagner GP, Kin K, Lynch VJ. A model based criterion for gene expression calls using RNA-seq data. Theory Biosci 2013;132:159-64.
- Chung IF, Chen CY, Su SC, et al. DriverDBv2: a database for human cancer driver gene research. Nucleic Acids Res 2016;44:D975-9.
- Li Q, Zheng S, Han A, et al. The splicing regulator PTBP2 controls a program of embryonic splicing required for neuronal maturation. Elife 2014;3:e01201.
- 41. Sebastian S, Faralli H, Yao Z, et al. Tissue-specific splicing of a ubiquitously expressed transcription factor is essential for muscle differentiation. Genes Dev 2013;27:1247-59.
- 42. Jacob AG, O'Brien D, Singh RK, et al. Stress-Induced Isoforms of MDM2 and MDM4 Correlate with High-Grade Disease and an Altered Splicing Network in Pediatric Rhabdomyosarcoma. Neoplasia 2013;15:1049-63.
- Comiskey DF Jr, Jacob AG, Sanford BL, et al. A novel mouse model of rhabdomyosarcoma underscores the dichotomy of MDM2-ALT1 function in vivo. Oncogene 2018;37:95-106.
- 44. Sanchez G, Bittencourt D, Laud K, et al. Alteration of cyclin D1 transcript elongation by a mutated transcription factor up-regulates the oncogenic D1b splice isoform in cancer. Proc Natl Acad Sci U S A 2008;105:6004-9.
- 45. Nowak DG, Amin EM, Rennel ES, et al. Regulation of vascular endothelial growth factor (VEGF) splicing from pro-angiogenic to anti-angiogenic isoforms: a novel therapeutic strategy for angiogenesis. J Biol Chem 2010;285:5532-40.
- Scheres SH, Nagai K. CryoEM structures of spliceosomal complexes reveal the molecular mechanism of pre-mRNA splicing. Curr Opin Struct Biol 2017;46:130-9.
- 47. Galej WP. Structural studies of the spliceosome: past, present and future perspectives. Biochem Soc Trans 2018;46:1407-22.
- Smith DJ, Konarska MM, Query CC. Insights into branch nucleophile positioning and activation from an orthogonal pre-mRNA splicing system in yeast. Mol Cell 2009;34:333-43.
- Schellenberg MJ, Dul EL, MacMillan AM. Structural model of the p14/SF3b155. branch duplex complex. RNA 2011;17:155-65.

- 50. Eke I, Cordes N. Focal adhesion signaling and therapy resistance in cancer. Semin Cancer Biol 2015;31:65-75.
- Tomar S, Plotnik JP, Haley J, et al. ETS1 induction by the microenvironment promotes ovarian cancer metastasis through focal adhesion kinase. Cancer Lett 2018;414:190-204.
- 52. Luo M, Guan JL. Focal adhesion kinase: a prominent determinant in breast cancer initiation, progression and metastasis. Cancer Lett 2010;289:127-39.
- 53. Zhao J, Guan JL. Signal transduction by focal adhesion kinase in cancer. Cancer Metastasis Rev 2009;28:35-49.
- 54. Zhou B, Wang GZ, Wen ZS, et al. Somatic Mutations and Splicing Variants of Focal Adhesion Kinase in Non-Small Cell Lung Cancer. J Natl Cancer Inst 2018. doi: 10.1093/ jnci/djx157.
- 55. Yao L, Li K, Peng W, et al. An aberrant spliced transcript of focal adhesion kinase is exclusively expressed in human breast cancer. J Transl Med 2014;12:136.
- 56. Despeaux M, Chicanne G, Rouer E, et al. Focal adhesion kinase splice variants maintain primitive acute myeloid leukemia cells through altered Wnt signaling. Stem Cells 2012;30:1597-610.
- 57. Rot S, Taubert H, Bache M, et al. Low HIF-1alpha and low EGFR mRNA Expression Significantly Associate with Poor Survival in Soft Tissue Sarcoma Patients; the Proteins React Differently. Int J Mol Sci 2018. doi: 10.3390/ijms19123842.
- 58. Smolle MA, Leithner A, Posch F, et al. MicroRNAs in Different Histologies of Soft Tissue Sarcoma: A Comprehensive Review. Int J Mol Sci 2017. doi: 10.3390/ ijms18091960.
- Yang Z, Li X, Yang Y, et al. Long noncoding RNAs in the progression, metastasis, and prognosis of osteosarcoma. Cell Death Dis 2016;7:e2389.
- Min L, Garbutt C, Tu C, et al. Potentials of Long Noncoding RNAs (LncRNAs) in Sarcoma: From Biomarkers to Therapeutic Targets. Int J Mol Sci 2017. doi: 10.3390/ijms18040731.
- 61. de Almeida SF, Carmo-Fonseca M. Design principles of interconnections between chromatin and pre-mRNA splicing. Trends Biochem Sci 2012;37:248-53.
- 62. Wada M, Horinaka M, Yasuda S, et al. PDK1 is a potential therapeutic target against angiosarcoma cells. J Dermatol Sci 2015;78:44-50.
- 63. Le Gallo M, Rudd ML, Urick ME, et al. The FOXA2 transcription factor is frequently somatically mutated in uterine carcinosarcomas and carcinomas. Cancer 2018;124:65-73.

Page 17 of 17

- 64. Joseph CG, Hwang H, Jiao Y, et al. Exomic analysis of myxoid liposarcomas, synovial sarcomas, and osteosarcomas. Genes Chromosomes Cancer 2014;53:15-24.
- 65. Chudasama P, Renner M, Straub M, et al. Targeting Fibroblast Growth Factor Receptor 1 for Treatment of Soft-Tissue Sarcoma. Clin Cancer Res 2017;23:962-73.

Cite this article as: Hong W, Zhang W, Guan R, Liang Y, Hu S, Ji Y, Liu M, Lu H, Yu M, Ma L. Genome-wide profiling of prognosis-related alternative splicing signatures in sarcoma. Ann Transl Med 2019;7(20):557. doi: 10.21037/atm.2019.09.65

- 66. Sveen A, Kilpinen S, Ruusulehto A, et al. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. Oncogene 2016;35:2413-27.
- Kovac M, Blattmann C, Ribi S, et al. Exome sequencing of osteosarcoma reveals mutation signatures reminiscent of BRCA deficiency. Nat Commun 2015;6:8940.

Table S1 Details of the seven risk scores of splicing events

Туре	ID	Exp (coef)	Lower	Upper	Se (coef)	Z score	Pr (> z)
All	COMT.61100.AT	5.98E+12	2.45E+07	1.46E+18	6.33E+00	4.65E+00	3.36E-06
,							
	PDHX.14973.ES	5.33E-05	4.68E-07	6.07E-03	2.42E+00	4.07E+00	4.64E-05
	RGS3.87287.AP	1.38E-03	5.70E-05	3.34E-02	1.63E+00	4.05E+00	5.13E-05
	NARF44392 AP	1 90F+04	3 57E+02	1.01E+06	2 03E+00	4 86F+00	1 18E-06
	ISPAN14.12377.ES	1.37E+02	1.48E+01	1.27E+03	1.14E+00	4.33E+00	1.50E-05
	RPRD1B.59373.ES	7.75E+02	3.49E+01	1.72E+04	1.58E+00	4.21E+00	2.60E-05
	SPRYD7.25916.ES	6.82E-03	6.91E-04	6.73E-02	1.17E+00	4.27E+00	1.95E-05
		0.012 00	0.012 01	0.102 02	=		
	ABHD14B.65146.AA	9.04E+01	9.64E+00	8.48E+02	1.14E+00	3.94E+00	8.03E-05
	ADCY1.79596.AT	1.48E+01	4.31E+00	5.09E+01	6.30E-01	4.28E+00	1.90E-05
	PTGER3.3415.AT	2.06F-02	3.08E-03	1.38E-01	9.70F-01	4.00E+00	6.28E-05
		2.002.02					0.202 00
	RBMS3.63800.AT	3.18E+25	5.67E+12	1.78E+38	1.50E+01	3.92E+00	8.83E-05
	CTNNAL1.87156.ES	1.09E-34	1.53E-49	7.84E-20	1.75E+01	4.48E+00	7.44E-06
	ZC3H14.28724.ES	1.72E+01	4.81E+00	6.18E+01	6.52E-01	4.37E+00	1.25E-05
			4.012100	0.102101	0.022 01	1.07 2100	1.202 00
	NUP160.15807.ES	1.87E-05	8.30E-08	4.20E-03	2.76E+00	3.94E+00	8.14E-05
AA	DCAF4.28239.AA	2.81E+02	3.60E+01	2.19E+03	1.05E+00	5.38E+00	7.64E-08
	DGK7,15549,AA	3.96F+10	1.63E+05	9.66F+15	6.33E+00	3.86E+00	1.15E-04
		0.002110		0.002.10	0.002.00	0.002.00	
	DPH5.3896.AA	3.40E+05	3.52E+02	3.28E+08	3.51E+00	3.63E+00	2.81E-04
	FKBP8.48446.AA	6.87E-07	5.41E-10	8.71E-04	3.65E+00	3.89E+00	9.92E-05
	LAIR1.51883.AA	3.14F+01	6.40F+00	1.54E+02	8.11E-01	4.25E+00	2.15E-05
		0.142101	0.402100			4.202100	2.102 00
	LTA4H.23825.AA	3.79E-15	2.54E-21	5.67E-09	7.25E+00	4.58E+00	4.71E-06
	PDDC1.13753.AA	5.64E+03	1.02E+02	3.12E+05	2.05E+00	4.22E+00	2.46E-05
	PKMYT1 33331 AA	9 20E-03	7 13E-04	1 19E-01	1 30E+00	3 59E+00	3 27E-04
		0.202 00				0.002.00	
	RALGAPB.59379.AA	9.23E-03	8.86E-04	9.62E-02	1.20E+00	3.92E+00	8.91E-05
	TREX1.64674.AA	2.87E-02	5.79E-03	1.42E-01	8.16E-01	4.35E+00	1.34E-05
	TYMP62850 44	1 16F±01	2 86E±00	4 67E±01	7 13E-01	3 43E±00	5 95E-04
	11111.02000.244	1.102+01	2.002+00	4.07 2401	7.102 01	0.402400	0.002 04
AD	ACAD9.66685.AD	1.83E-02	1.80E-03	1.86E-01	1.18E+00	3.38E+00	7.28E-04
	ADAM15.7894.AD	7.85E+00	2.02E+00	3.04E+01	6.92E-01	2.98E+00	2.89E-03
		3 60E±03	5.61E+01	2 31E±05	2 12E±00	3 86E±00	1 15E-04
	ALINDI 10.43021.AD	0.00L+00	5.01L+01	2.012+00	2.122+00	5.00L+00	1.136-04
	NUP98.13998.AD	5.18E-02	8.85E-03	3.03E-01	9.01E-01	3.29E+00	1.01E-03
	OXA1L.26618.AD	6.32E+07	1.24E+04	3.21E+11	4.35E+00	4.13E+00	3.70E-05
		1 565 05	2 225 09	7 21 5 02	2145,00	2 525,00	4 205 04
	PGD.012.AD	1.50E-05	3.33E-00	7.3TE-03	3.14E+00	3.53E+00	4.200-04
	PIGG.68359.AD	8.25E-03	3.54E-04	1.93E-01	1.61E+00	2.98E+00	2.84E-03
AP	ABI2.57012.AP	1.75E-15	2.18E-23	1.40E-07	9.29E+00	3.66E+00	2.53E-04
		2 105 15	601504	6 405 07	0.075.00	2 205 . 00	6 07E 04
	COP55.84054.AP	2.10E-15	0.91E-24	6.40E-07	9.97E+00	3.39E+00	6.97E-04
	LRP2BP.71337.AP	1.15E-01	3.39E-02	3.92E-01	6.25E-01	3.46E+00	5.42E-04
	AES.46658.AP	1.78E+02	1.38E+01	2.31E+03	1.31E+00	3.97E+00	7.22E-05
		2 545 01	4.245.00		1.075.00	2 225,00	9 71E 04
	NFIA.47099.AF	3.54E+01	4.342+00	2.09E+02	1.07E+00	3.33E+00	0.7TE-04
	WDR37.10611.AP	9.93E+00	2.64E+00	3.73E+01	6.75E-01	3.40E+00	6.79E-04
	C11orf68.16947.AP	7.89E+01	9.84E+00	6.32E+02	1.06E+00	4.11E+00	3.90E-05
		1 245,01	2 105 00		7 475 01	2 495 .00	5 10E 04
	FAIVI228B.52812.AP	1.34E+01	3.10E+00	5.80E+01	7.47E-01	3.48E+00	5.10E-04
	MSANTD2.19240.AP	1.58E+01	3.48E+00	7.20E+01	7.73E-01	3.57E+00	3.54E-04
	TJAP1.76273.AP	1.13E-03	2.92E-05	4.36E-02	1.86E+00	3.64E+00	2.73E-04
		1 505 00	1 405 00	1.055.01	1.015.00	0.445.00	5 705 04
	CT101108.10948.AP	1.53E-02	1.42E-03	1.65E-01	1.21E+00	3.44E+00	5.73E-04
	CTNNBL1.59361.AP	5.56E-17	1.49E-26	2.07E-07	1.12E+01	3.33E+00	8.73E-04
	EPC1.11157.AP	4.59E+01	5.26E+00	4.01E+02	1.11E+00	3.46E+00	5.39E-04
		7.005.00			5 705 04		
	SLC29A1.76359.AP	7.66E-02	2.51E-02	2.34E-01	5.70E-01	4.51E+00	6.57E-06
AT	C4orf29.70557.AT	2.92E-03	9.28E-05	9.22E-02	1.76E+00	3.31E+00	9.19E-04
	LIMK2.61835.AT	7.20E+01	6.38E+00	8.13E+02	1.24F+00	3.46E+00	5.45E-04
				0.005.07		0.005.00	
	PTK2B.83154.AT	6.51E+04	1.61E+02	2.63E+07	3.06E+00	3.62E+00	2.96E-04
	DHRS12.25951.AT	1.42E+03	2.80E+01	7.21E+04	2.00E+00	3.62E+00	2.92E-04
	PTK2B.83155.AT	2.44E-05	1.21E-07	4.93E-03	2.71E+00	3.92E+00	8.82E-05
	THORA DOOD AT	0.005.04		1045 07	0.405.00	0.005.00	0.505.04
	TWIZD1.3232.AI	3.62E+04	7.14E+01	1.84E+07	3.18E+00	3.30E+00	9.58E-04
	ACPP.66802.AT	8.87E+00	2.73E+00	2.88E+01	6.02E-01	3.63E+00	2.86E-04
	MAP3K8.11123.AT	1.46E+01	3.11E+00	6.90E+01	7.91E-01	3.39E+00	6.91E-04
	TNEN4100 10107 AT	E 10E 00	0 545 01	4.075.05		0.045.00	1 055 04
	TMEM138.16187.AI	5.16E+03	6.54E+01	4.07E+05	2.23E+00	3.84E+00	1.25E-04
	FLCN.39460.AT	6.79E+01	1.25E+01	3.70E+02	8.65E-01	4.87E+00	1.09E-06
	HS2ST1.3691.AT	2.66E+02	2.06E+01	3.44E+03	1.31E+00	4.28E+00	1.89E-05
	FLON 00 404 AT	0.005.00	4.405.00	0.405.04	1 005 00	0.005.00	0.005.04
	FLGN.39461.AI	3.33E-02	4.48E-03	2.48E-01	1.02E+00	3.32E+00	8.96E-04
	HSPB7.787.AT	1.06E+02	9.23E+00	1.21E+03	1.24E+00	3.75E+00	1.80E-04
	HSPB7.788.AT	4.21E-03	2.97E-04	5.96E-02	1.35E+00	4.04E+00	5.25E-05
		0.007.00					
	HTRA3.68763.AI	9.39E-03	7.09E-04	1.24E-01	1.32E+00	3.54E+00	4.00E-04
	LRPAP1.68648.AT	2.34E+08	2.80E+03	1.95E+13	5.78E+00	3.33E+00	8.60E-04
	PTGER3.3415 AT	3.25E-02	4.46E-03	2.38E-01	1.01E+00	3.38E+00	7.32E-04
F.2			0.705.55		0.075		
ES	CFLAR.56797.ES	5.47E-62	2./2E-87	1.10E-36	2.97E+01	4.75E+00	2.08E-06
	CTNND1.15999.ES	2.27E-02	4.46E-03	1.16E-01	8.32E-01	4.55E+00	5.38E-06
	MFAP5 20197 FS	5.53E-02	1.41F-02	2.17E_01	6.98F-01	4.15F±00	3.37E-05
		0.002-02		2.17 -01	0.00L-01	-T. TOL+00	5.07 2-05
	SMARCE1.40881.ES	2.97E-55	1.18E-82	7.47E-28	3.22E+01	3.90E+00	9.61E-05
	ZNF568.49434.ES	4.53E-04	1.25E-05	1.64E-02	1.83E+00	4.21E+00	2.61E-05
	C16orf13 32021 ⊑€	2 185-02	A 15E-02	1 155-01	8 /7E_01	A 52E 00	
	0100110.02821.EO	2.100-02	4.102-03	1.100-01	0.47 E-01	4.JZE+UU	0.202-00
	DUS2.37175.ES	8.92E-06	3.32E-08	2.39E-03	2.85E+00	4.08E+00	4.60E-05
	SUN2.62261.ES	3.80E-07	8.39E-10	1.72E-04	3.12E+00	4.74E+00	2.14E-06
рı	KI E16 /6550 DI	1 01 - 40	2 70E 74	1 21 - 00	3015.01	3 625 00	2055 04
זרו	NEI 10.40002.HI	ı.⊎⊺⊏-4ŏ	2.195-14	1.312-22	3.04E+U1	J.U2E+UU	∠. 3 0E-04
	RPL13.38091.RI	2.05E+92	1.37E+28	3.08E+156	7.54E+01	2.82E+00	4.81E-03
	TIMP1.88928.RI	1.37E+65	2.14E+38	8.75E+91	3.15E+01	4.76E+00	1.91E-06





Figure S1 Mutated characteristics of driver genes. (A) A list of driver genes was produced via DriverDBv2 (http://driverdb.tms.cmu.edu.tw/ driverdbv2/). Bioinformatics tools are shown on the horizontal axis, and driver genes are shown on the vertical axis. (B) The expression of driver genes in the whole sarcoma cohort. The darker color represents a higher expression. (C) Illustrations of the mutation classes for driver genes. The truncating, missense and inframe subtypes are depicted in red, purple and green, respectively.



Figure S2 Correlations between splicing events and driver genes. (A) The heatmap of co-expression between mutated driver genes and the top 30 survival-associated splicing events. Blue to red represents the correlation coefficient from low to high. (B) The heatmap of significance between the top 30 survival-associated splicing events and driver genes through *t*-test. Blue to red represents P values from low to high.