Gene 835 (2022) 146657

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

The comprehensive and systematic identification of BLCA-specific SF-regulated, survival-related AS events

Zhe Liu [a,1], Xudong Liu [b,c,1], Fang Liu [d,1], Hui Zhao [e], Yu Zhang [f], Yafan Wang [f], Ying Ma [f], Fuzhou Wang [a], Weitong Zhang [a], Olutomilayo Olayemi Petinrin [a], Zhongyu Yao [a], Jingbo Liang [g], Qian He [g], Dayun Feng [h,*], Lei Wang [f,i,*], Ka-Chun Wong [a,*]

[a] Department of Computer Science, City University of Hong Kong, Hong Kong, China
[b] School of Medicine, Chongqing University, Chongqing 400044, China
[c] Chongqing Key Laboratory of Translational Research for Cancer Metastasis and Individualized Treatment, Chongqing University Cancer Hospital, Chongqing 400030, China
[d] CYGNUS BIOSCIENCES (Beijing), Beijing 100176, China
[e] Key Laboratory of Growth Regulation and Translational Research of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou 310058, China
[f] College of Life Sciences, Xinyang Normal University, Xinyang 464000, China
[g] Department of Biomedical of Science, City University of Hong Kong, Hong Kong, China
[h] Department of Neurosurgery, Tangdu Hospital, The Fourth Military Medical University, Xi'an, Shaanxi 710038, China
[i] College of Medicine, Xinyang Normal University, Xinyang 464000, China

ARTICLE INFO

Edited by Erica Golemis

Keywords:
Alternative splicing (AS)
Splicing factor (SF)
Bladder urothelial carcinoma (BLCA)
TCGA
Prognostic analysis

ABSTRACT

Bladder urothelial carcinoma (BLCA) is a complex disease with high morbidity and mortality. Changes in alternative splicing (AS) and splicing factor (SF) can affect gene expression, thus playing an essential role in tumorigenesis. This study downloaded 412 patients' clinical information and 433 samples of transcriptome profiling data from TCGA. And we collected 48 AS signatures from SpliceSeq. LASSO and Cox analyses were used for identifying survival-related AS events in BLCA. Finally, 1,645 OS-related AS events in 1,129 genes were validated by Kaplan-Meier (KM) survival analysis, ROC analysis, risk curve analysis, and independent prognostic analysis. Finally, our survey provides an AS-SF regulation network consisting of five SFs and 46 AS events. In the end, we profiled genes that AS occurred in pan-cancer and five SFs' expression in tumor and normal samples in BLCA. We selected CLIP-seq data for validation the interaction regulated by RBP. Our study paves the way for potential therapeutic targets of BLCA.

1. Introduction

Bladder cancer is one of the top ten common cancers in China (Li et al., 2021). It has become one of the important disease burdens in our country due to its easy recurrence, easy metastasis, and limited treatment options. It seriously threatens the survival time and quality of life of patients. Urothelial carcinoma (UC) is common bladder cancer, accounting for more than 90% of all bladder cancer cases (Magi-Galluzzi et al., 2008). Urothelial cancer is further divided into bladder urothelial carcinoma (BLCA) and upper tract urothelial cancer, of which BLCA accounts for more than 90% of all bladder cancer cases (Magi-Galluzzi et al., 2008; Parker and Spiess, 2011). Currently, chemotherapy is still the main treatment for recurrent and metastatic urothelial carcinoma, but the efficacy is limited. In recent years, immunotherapy represented by PD-1/PD-L1 inhibitors has brought new therapeutic opportunities for patients (Stenehjem et al., 2018). The onset of urothelial carcinoma is

Abbreviations: AS, Alternative Splicing; BLCA, Bladder Urothelial Carcinoma; SF, Splicing Factor; TCGA, The Cancer Genome Atlas; OS, Overall Survival; ES, Exon Skip; AT, Alternative Terminator; AD, Alternative Donor Site; ME, Mutually Exclusive Exons; AP, Alternative Promoter; RI, Retained Intron; AA, Alternative Acceptor Site; ROC, Receiver Operating Characteristic; AUC, Area Under Curve; LASSO, Least Absolute Shrinkage and Selection Operator; PSI, Percent Spliced In; HR, Hazard Ratio; PCC, Pearson Correlation Coefficient; RBP, RNA Binding Protein; KM, Kaplan-Meier; LncRNA, Long Non-coding RNA; CR, Chromatin Regulator; miRNA, microRNA.

* Corresponding authors at: Department of Computer Science, City University of Hong Kong, Hong Kong, China.
E-mail addresses: dayunfmmu@163.com (D. Feng), wangleibio@126.com (L. Wang), kc.w@cityu.edu.hk (K.-C. Wong).

[1] These authors contributed equally.

https://doi.org/10.1016/j.gene.2022.146657
Received 9 April 2022; Received in revised form 27 May 2022; Accepted 2 June 2022
Available online 13 June 2022

insidious, and most patients are clinically advanced at the time of diagnosis. Platinum-based combined chemotherapy is the first-line standard treatment for advanced urothelial carcinoma, but the efficacy of chemotherapy is limited (Gómez De Liaño and Duran, 2018; Hanna, 2017). Thus, it is vital for us to identify biomarkers for diagnosing and treating BLCA.

AS, also known as differential splicing, refers to the fact that in the process of mRNA precursor to mature mRNA, different splicing methods allow the same gene to produce multiple different mature mRNAs, and ultimately produce different proteins (Lee and Rio, 2015; Montes et al., 2019). AS is an important mechanism for regulating gene expression and generating proteome diversity, and it is an important reason for the large differences in the number of eukaryotic genes and proteins (Black, 2003; Sciarrillo et al., 2020). AS frequently occurs in tumors and is closely related to tumor development. The study found that AS affects those protein gene families that are frequently mutated in tumors and alters protein–protein interactions in tumor-related signaling pathways, indicating that AS is also an important factor driving tumorigenesis (Black et al. 2019). For example, Paik et al. reported that oncogenic mutations in the proto-oncogene named MET can lead to exon 14 skip in lung cancer by nCounter Analysis System platform (Paik et al., 2015). Huang et al. suggested that *PCBP1* regulates the splicing of *APOC1* and *SPHK1* in hepatocellular carcinoma (HCC) by combining RNA-seq and eCLIP-seq data (S. Huang et al., 2021). However, few studies have systematically and comprehensively identified AS events associated with BLCA diagnosis and prognosis.

SFs are a class of protein factors involved in the splicing process of RNA precursors (Du et al., 2021). According to their functional roles, they can be divided into small nuclear ribonucleoprotein particle (snRNP) protein factors and non-snRNP protein factors (Gonçalves et al., 2017). Aberrant expression of SF can lead to altered AS of genes. In tumors, aberrant expression of SF may lead to the formation of specific cancer-promoting splicing isoforms, leading to cancer development (Koedoot et al., 2019). For instance, the intracellular protein T-cell Intracellular Antigen (*TIA1*) regulates *VEGF* isoform expression, angiogenesis, tumor growth, and bevacizumab resistance in colon cancer (Hamdollah Zadeh et al., 2015).

In this study, SF expression RNA-seq data and BLCA patient clinical data were downloaded from the TCGA database. AS events in BLCA and SFs information were collected from TCGA-SpliceSeq and SpliceAid 2 webserver, respectively. LASSO regression and Cox model were applied for inferring prognostic-associated AS events in BLCA. And these AS events were evaluated by KM survival analysis, ROC curve, risk curve, and independent prognostic analysis. Further, we combined SF expression levels data with AS events for building AS-SF regulatory network. And *TIA1* CLIP-seq data was used for validation. Finally, we comprehensively and systematically identified five SFs and 46 AS events.

2. Materials and methods

2.1. SF expression data and clinical data collection and pre-processing

The BLCA RNA-seq dataset (level 3) and clinical information were downloaded and integrated via the TCGAbiolinks R package (version: 2.22.4) (Colaprico et al., 2016; Mounir et al., 2019) from the TCGA data portal.

For obtained gene expression data, clusterProfiler R package (version: 4.0.2) (Wu et al., 2021; Yu et al., 2012)was utilized for gene symbol annotation (from Ensembl ID to official gene symbol). SpliceAid 2 webserver (https://www.introni.it/spliceaid.html) (Piva et al., 2012) was used for exporting SF genes.

2.2. AS events data collection and pre-processing

The AS events of BLCA were downloaded from TCGA SpliceSeq (https://bioinformatics.mdanderson.org/public-software/tcgasplicseq/) (Ryan et al., 2016) database, which is a tool for investigating alternative mRNA splicing in TCGA tumor and adjacent normal samples. Perform comparative analysis by producing splice graphs annotated with reading totals and percent spliced in (PSI) values for all potential splice events.

To generate as reliable a set of AS events as possible, we implemented a series of stringent filters: 1. Standard deviation (SD) filter: If the SD value of AS event is less than 0.1, then we remove the AS event. The standard deviation of an AS event in all samples is small, indicating that its fluctuation range is small, proving that this AS event has almost no effect on the patient survival time. 2. Data filtering: If the PSI value of an AS event is missing in more than 25% of the samples, then we remove the AS event. 3. Missing value imputation: fill missing PSI values (NA) with 0.

2.3. The performance of prognostic signatures

The survival R package (version: 3.2.13) was used to perform univariate Cox regression and LASSO logistic regression for screening the OS-associated AS events as prognostic signatures. The hazard ratio (HR) value was used for evaluating the relationship between the PSI value of an AS event and the risk of a patient. HR larger than one demonstrates this AS event is a risk AS event; that is, the large PSI value the high risk of the patient. While HR less than one demonstrates this AS event isn't a risk AS event, the larger PSI value the low risk of the patient. Finally, we selected the risk AS events based on the p-value < 0.05. For removing genes with high correlation and preventing overfitting of the model, we added lambda coefficient to LASSO regression and deleted genes with high correlation.

These AS events were included in the multivariate Cox regression model to construct the independent prognosis signature for BLCA. The risk score of each selected OS-related AS event was determined by the following formula:

$$\text{risk score (AS event)} = \sum_{i}^{n} \text{PSI}_i \times i\beta$$

Here, β represents the regression coefficient of multivariate Cox regression. Once we got the risk score of each patient, we can calculate the median risk score value of all patients. All patients were divided into two groups: the high-risk patient group (risk score $>$ median value) and the low-risk patient group (risk score $<$ median value). The p-value $<$ 0.05 and AUC > 0.65 were set for evaluating model performance.

2.4. The construction of the SF-AS regulatory network

The Pearson correlation analysis between the expression level of SF and the PSI value of AS event was performed for constructing the SP-AS regulatory relationship network. The cor.test() function in R program (version: 4.1.2) was used for getting the Pearson correlation coefficient (PCC) value and p-value. The Pearson correlation coefficient PCC value 0.6 and p-value 0.001 were set to the cutoff values to correlation and significance. In the end, the Cytoscape (version: 3.9.0) was utilized for visualizing and layout SF-AS regulatory network.

2.5. The validation of SF regulation and PPI interaction

In order to explore the regulation relationship between RBP and RNA, we downloaded the processed TIA1 CLIP-seq data from NCBI GEO via the accession number GSE94369 (three replicate samples: GSM2474165, GSM2474166, and GSM2474167). The mouse reference genome version is mm9. IGV software (version: 2.11.6) (Robinson et al., 2011) was used for visualizing the binding between RBP TIA1 and RNA SRSF7.

3. Results

3.1. AS events profile in BLCA

The flowchart of our research is illustrated in Supplementary Fig. 1. There are 433 patient samples with the transcriptome profiling FPKM values. And there are 412 patient samples with the clinical information (Supplementary Table 1). In summary, 56,602 Ensembl IDs are converted to 12,293 official gene symbols. Forty-eight out of 67 SF genes are selected as SF genes (Supplementary Table 2).

Here, we focus on the seven different types of AS events (Lee and Rio, 2015; Montes et al., 2019): exon skip (ES), mutually exclusive exons (ME), alternative promoter (AP), retained intron (RI), alternative terminator (AT), alternative donor site (AD), and alternative acceptor site (AA). The schematic diagram of the above seven types is plotted in Fig. 1A. In total, there are 10,434 different AS events in 4,659 genes. We detected 3,173 ESs in 1,163 genes, 3,009 APs in 880 genes, 1,934 ATs in 704 genes, 908 RIs in 302 genes, 668 AAs in 205 genes, 690 ADs in 199 genes, and 52 MEs in 14 genes (Fig. 1B). It shows that one gene might have two kinds of AS types on average (10434/4659). This result demonstrated that gene expression is diverse because of different AS types. ES is the dominant AS type among them, accounting for more than 30% of all AS events (3173/10434), which is consistent with the reported results (Bonnal et al., 2020).

3.2. Identification of prognosis-related AS events in BLCA

The Cox regression was used for identifying OS-related AS events. We defined prognosis-related AS events as the p-value < 0.05 and $|HR|>1$ (Fig. 2A).

In total, there are 1,645 OS-related AS events in 1,129 genes. We detected 418 ESs in 330 genes, 556 APs in 368 genes, 351 ATs in 201 genes, 154 RIs in 119 genes, 80 AAs in 53 genes, 77 ADs in 51 genes, and 9 MEs in 7 genes (Fig. 2B).

Then, we selected the top 20 prognosis-associated AS events in ES, ME, AP, AT, AD, AA, and RI (Fig. 3A-3G). These results show that only about eleven percent of AS events are significantly associated with the prognosis of BLCA (1129/10434).

3.3. The prognostic value BLCA-related AS events

We got the top 20 prognosis-related AS events for each AS type in BLCA. We didn't select and remove genes with a high correlation in this process. So, we performed LASSO regression analysis to avoid model overfitting problems (Supplementary Fig. 2).

After the LASSO analysis, we obtained ten AS events in ES type, six AS events in AA type, nine AS events in AD type, seven AS events in AP type, six AS events in AT type, six AS events in RI event, and five AS events in ME type. Fig. 4 and Supplementary Figs. 3-8 showed the risk curves, survival status, and risk heatmap. We split all patients into two equal numbers of groups: high-risk and low-risk groups according to the risk score value. If the risk score is larger than the median value of all risk score values, the patient is classified into the high-risk group. At the same time, the patient is classified into a low-risk group when its risk score is less than the median value of all risk score values. In the survival status Fig. 4B, we concluded that the death rate of patients increased as the large the risk score was. We identified high-risk AS events and low-risk AS events based on the color change from left to right in the heatmap. If the color is black to green, the AS event is a low-risk AS event. In contrast, the AS event is a high-risk AS event if the color is green to red. We also concluded them in Supplementary Table 1.

3.4. Univariate and multivariate independent prognostic analysis

In addition, an independent prognostic efficacy of the risk signatures in BLCA was performed for identifying independent prognostic factors.

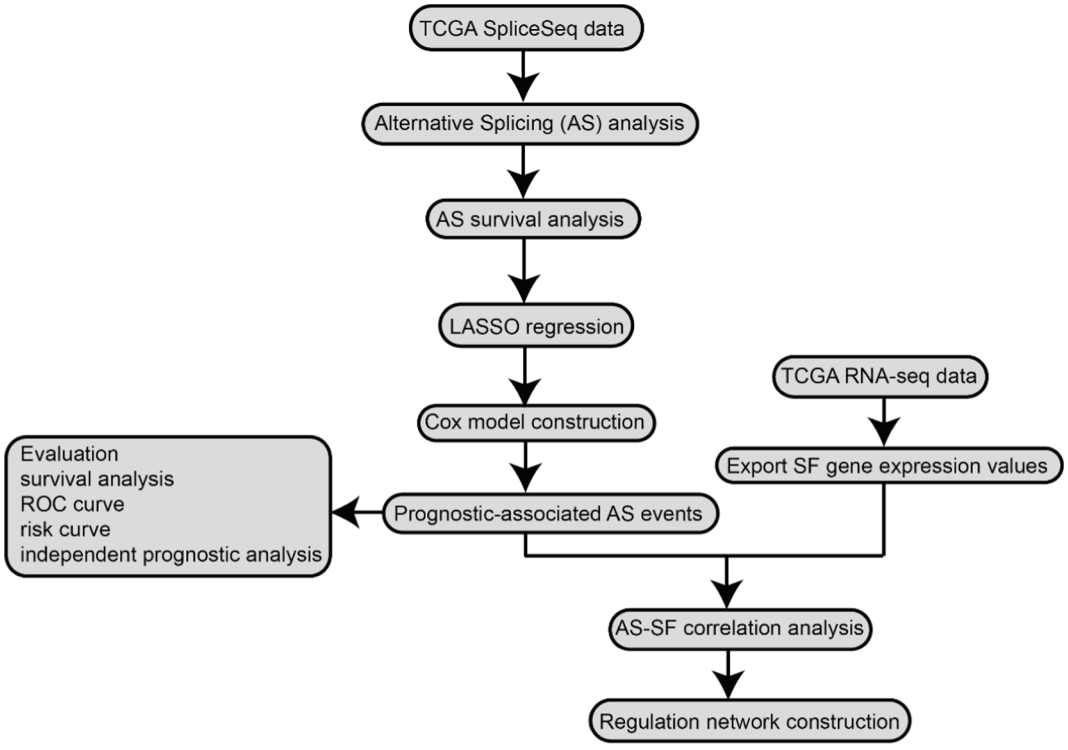

Fig. 1. The overview of AS events in BLCA. (A) The diagram of seven classic AS event models, including exon skip (ES), mutually exclusive exons (ME), alternative promoter (AP), retained intron (RI), alternative terminator (AT), alternative donor site (AD), and alternative acceptor site (AA). Exons are represented as red and blue blocks, introns as lines. (B) The Upset plot of the interactions how many genes are involved in each type of AS event in BLCA. For example, there are 203 genes involved in both ES and AP AS events. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

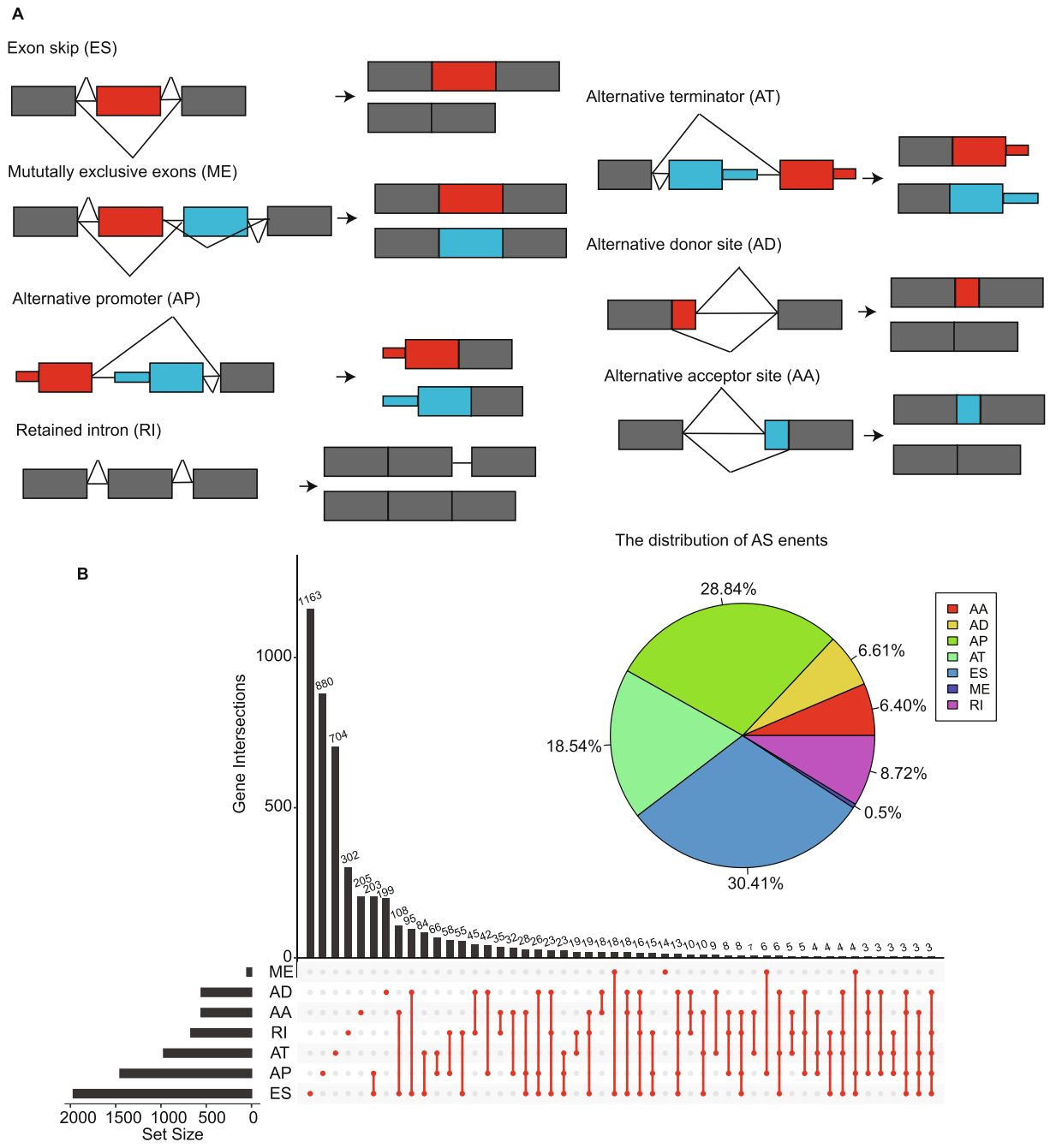

Fig. 2. OS-related AS events in BLCA. (A) The volcano plot indicates the OS-related AS events. The red dots represent prognosis-related AS, while the blue dots represent no significant prognosis-related AS. The x-axis and y-axis are z-score and -log10(p-value) from univariate Cox analysis, respectively. (B) The Upset plot of the interactions how many genes are involved in each type of AS event related to OS. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We divided patients into different groups based on the clinical characteristics: age, gender, disease stage, T (tumor), M (metastasis), and N (lymph node). The green forest plots of AA, AD, AP, AT, ES, RI, and ME suggested that age, gender, stage, T, N, and risk score are related to survival time and status. One of them can serve as an independent clinical characteristic by univariate independent prognostic analysis (p-value < 0.05, Supplementary Fig. 9A-9G).

Besides, we put all factors into the comparison at one time. After the multivariate independent prognostic analysis, we concluded that risk score is an independent prognostic factor (Supplementary Fig. 10A-10G).

Removing not significant factors and factors that can be instead, we identified an independent prognostic factor – risk score.

3.5. The evaluation of AS models in BLCA

In summary, prognostic-associated factors were obtained by univariate factor analysis. Then, the independent factor was obtained by multivariate factor analysis. Collectively, we constructed a convenient model for the prognosis of BLCA patients by only using risk scores. The KM survival analysis of the final signature (risk score) indicated that there was a notable difference in survival times between high-risk group and low-risk group (p-value < 0.05). Fig. 5A-5G showed the survival analysis results of these ES, ME, AP, RI, AT, AD, and AA.

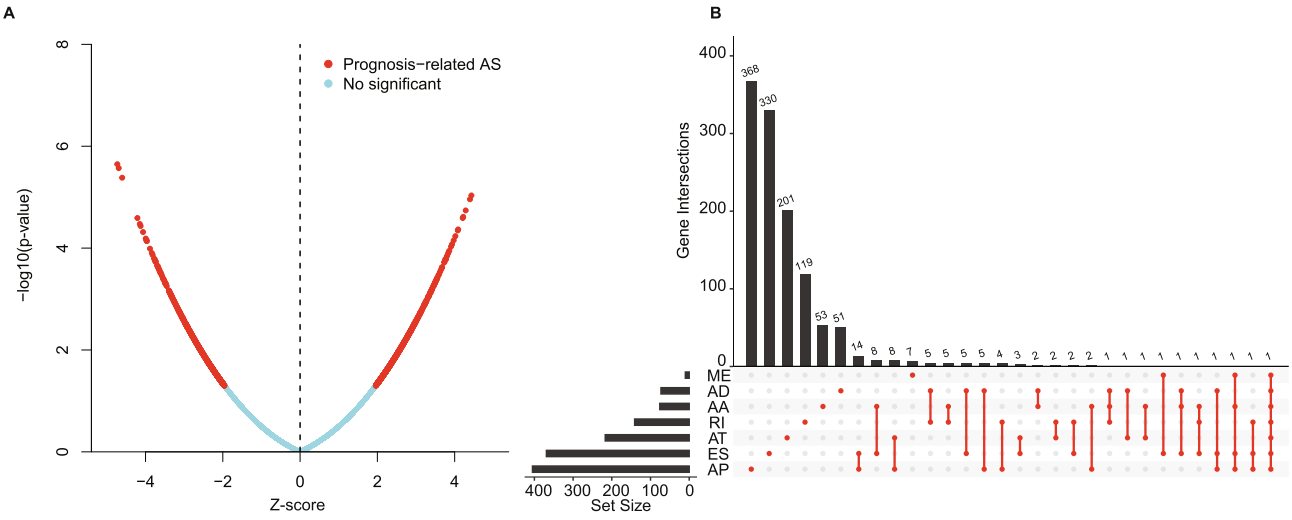

Fig. 3. The top 20 OS-related seven different kinds of AS events. (A) is for ES, (B) is for ME, (C) is for AP, (D) is for AT, (E) is for AD, (F) is for AS, and (G) is for RI. Note that: For the ME type, there are only nine (less than 20) significant AS events.

Moreover, the ROC curve is also used for assessing the efficiency among AS models. The range of AUC is from 0.5 to 1. The AS models have high AUC values (Fig. 6). The ROC curve demonstrated that the ES, AD, and AA constructed AS prognostic models are the top three models. AS events selected have good performance.

3.6. The construction of SF-AS regulatory network

A SF is a kind of protein involved in removing introns from strings of messenger RNA, so that the exons can bind together; the process takes place in spliceosomes (Ule and Blencowe, 2019). SF is essential to determine cell type and cause disease by regulating AS events (Cieply and Carstens, n.d; Zhang et al., 2008).

Pearson correlation analysis between the expression levels of SF and AS events was performed for selecting regulation pairs. After the filtering criteria: the |Pearson correlation coefficient (PCC) value| > 0.6 and p-value < 0.001, SF-AS regulatory pairs were determined. Fig. 7 is the visualization of the SF-AS regulatory network. The network consists of five SFs and 46 AS events (12 up AS events and 34 down AS events). The AS-up indicates it's a risk AS that is, the larger the PSI value of the AS with the high risk of the patient. The AS-down indicates it's not a risk AS; that is, the larger the PSI value of the AS with the low risk of the patient.

3.7. RBP regulation CLIP-seq data analysis

As was shown in Fig. 7, *HNRNPU* down-regulated (green line) good prognosis-related AS events (green dots), *TIA2A* and *TIA1* up-regulated (red line) bad prognosis-related ones (red dots), PCBP1 and RBM5 had various regulation functions. Supplementary Fig. 11 shows that two were significantly differentially expressed between tumor and adjacent normal BLCA samples with the p-value < 0.05: *TRA2A* and *TIA1*. At the same time, they were upregulated in tumor samples compared with normal samples.

TCGA SpliceSeq was used for providing an overview of AS events that our interested in cancers (Ryan et al., 2016). Fig. 8A shows the PSI values of AS event in exon 4.2, exon 4.3, exon 4.4, and exon 4.5 of the *SRSF7* gene that is regulated both by *TRA2A* and *TIA1* across 33 different cancer types. SRSF7-53280-RI belongs to RI AS event that occurred in exon 4.2, 4.3, 4.4, and 4.5 from this splice graph. Cross-tumor box plots show PSI data from all TCGA samples for 33 different tumor types and when available adjacent normal samples. Vagner S et al. provided the three replicates peak files in mammary tumor cells for TIA1 CLIP-seq data in the mm9 version. We visualized the TIA1 CLIP-seq peaks into IGV genomic visualization tool in reference genome mm9. In mammary tumor cell of mouse, we concluded that *TIA1* binding to *SRSF7* (Fig. 8B).

4. Discussion

Bladder cancer is a common malignant tumor of the urinary system (Zhang et al., 2021). BLCA is the predominant histological type of bladder cancer, accounting for 90% of all bladder cancers (Zhang et al., 2021).

Currently, some technologies and methods have applied for identifying biomarkers of BLCA. Long non-coding RNA (lncRNA), chromatin regulator (CR), protein-coding gene, microRNA (miRNA) can serve as the biomarker for predicting the prognosis of BLCA. In 2022, there is a study that found that *TERC* was significantly up regulated expressed in urinary exosomes from four BLCA patients compared with those from three healthy controls (Chen et al., 2022). And they summary that urinary exosome *TERC* is a diagnostic and prognostic biomarker for BLCA (Chen et al., 2022). Zhu et al. constructed and validated an 11 CRs-based model for predicting the survival status of BLCA patients (Zhu et al., 2022). Functional analysis suggested that these 11 CRs are related to immune checkpoint and immune cells infiltration (Zhu et al., 2022). And the eight small molecule drugs that sensitive to high-risk group were beneficial to treatment for BLCA (Zhu et al., 2022). *HMMR* was identified as a kind of protein coding gene of biomarker both in the expression level and in the independent prognostic level by Yang et al. (Yang et al., 2019). According to the TCGA data, Peng et al. identified a three-miRNA signature as a novel potential prognostic biomarker in BLCA (Peng et al., 2017).

However, studies about BLCA-specific SF-regulated, survival-related AS events are exceedingly rare. Furthermore, the experimental validation for AS-SF regulation relationship is also rare. Here, we proposed a mature and systematic pipeline for identifying prognostic related AS event. It realized non-invasive detection of BLCA.

In this study, we identified AS events and SF through computational biology methods from TCGA BLCA cohort data to explore the clinical significance of differential RNA splicing patterns. We detected 1,645 differentials AS events in 1,129 genes using univariate Cox regression analysis (Fig. 2B). Supplementary Table 3 describes about half of AS events were favorable prognostic factors (20 AS events, HR < 1), and about half of AS events were adverse prognostic factors (29 AS events, HR > 1). Only six of the top 20 significant OS-related AS events were favorable prognostic factors, and the rest 14 AS events were adverse

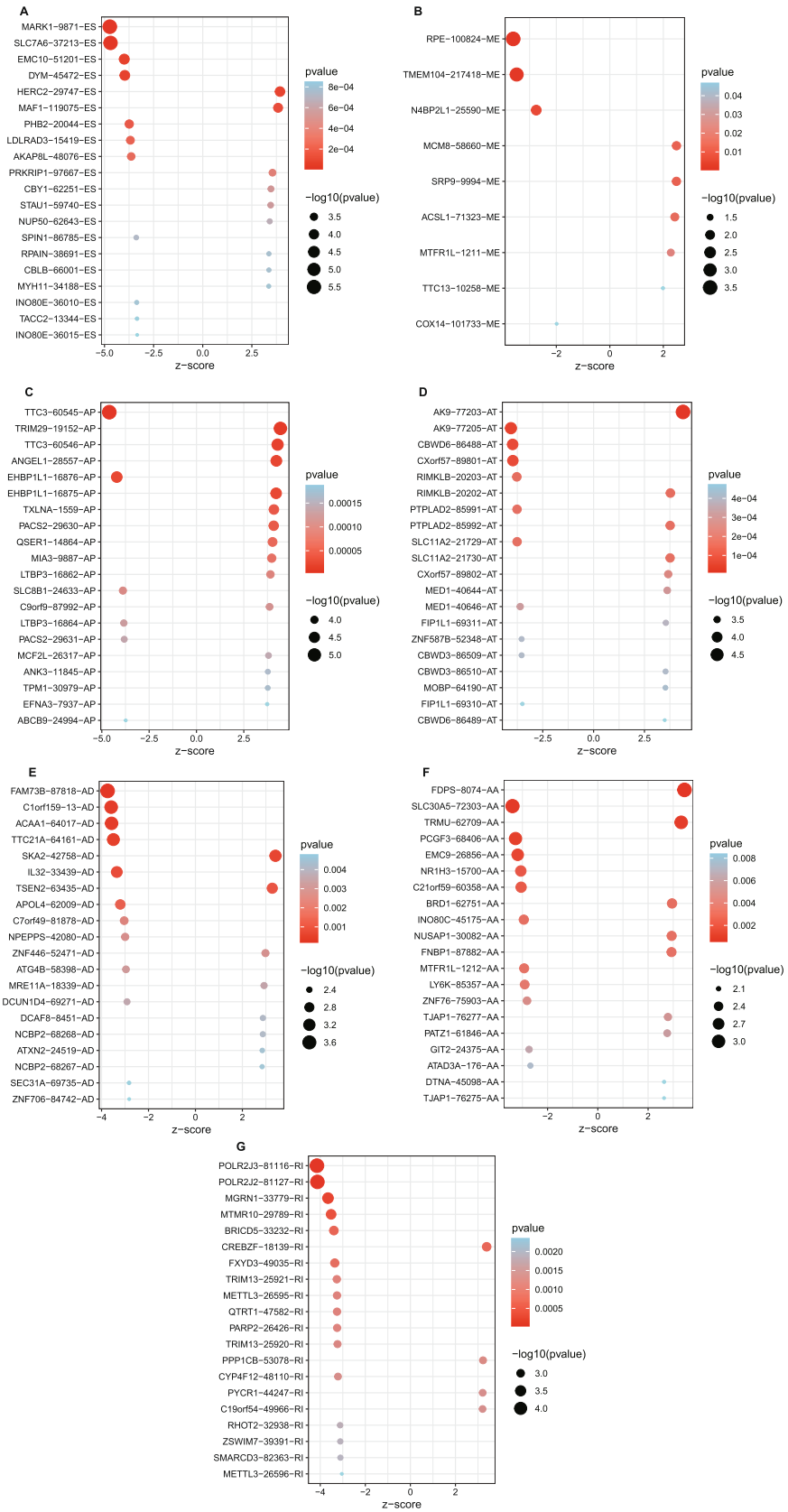

Fig. 4. The construction of prognostic models of ES type in BLCA. The x-axis is the patient sorted by the risk. (A) The y-axis is the risk score. The red dot indicates the risk score for high-risk patients (risk score > 1), while the green dot indicates the risk score for low-risk patients (risk score < 1). (B) The y-axis is the survival time. The red dot indicates the dead patient, while the green shows the alive patient. (C) The y-axis is the AS events belonging to the ES type. The color indicates the risk score of each AS event in each patient. The red color indicates the high risk, while the green shows the low risk. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

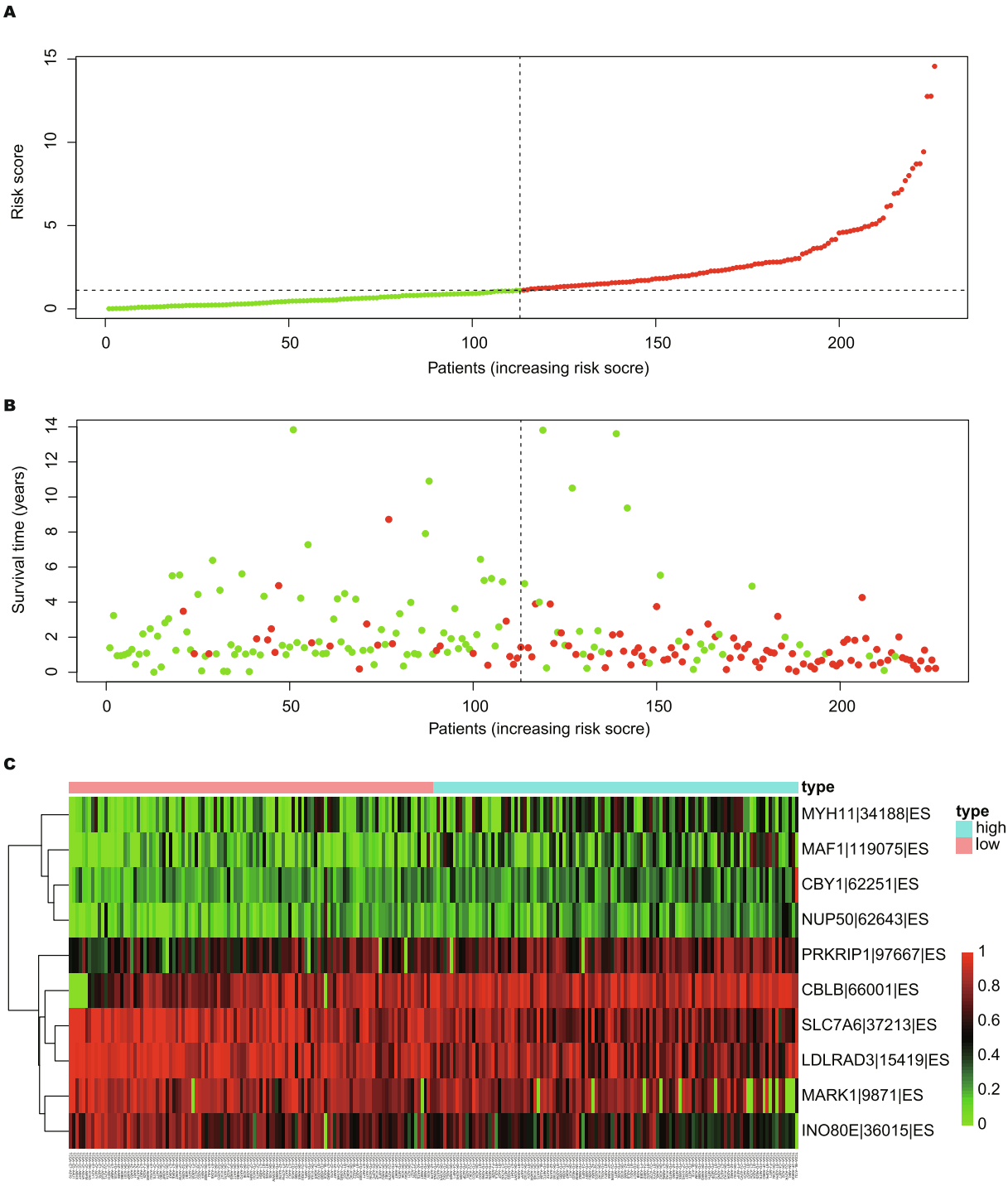

Fig. 5. The K-M curves of the prognostic predictors about the OS characteristics of seven different kinds of AS events. (A) is for ES, (B) is for ME, (C) is for AP, (D) is for RI, (E) is for AT, (F) is for AD, and (G) is for AA. The p-value < 0.05 shows a significant difference between the low and high groups.

prognostic factors. We expected that protein AS regulated may function in the development of BLCA, however, further studies are needed to confirm our suggestion.

Prognostic-related AS events were evaluated by KM survival curve, risk curve, ROC curve, and independent prognostic analysis. The KM survival analysis indicated that there was a notable difference in survival times between high-risk group and low-risk group (p-value < 0.05). Removing high-related AS events, we obtained ten AS events in ES type, six AS events in AA type, nine AS events in AD type, seven AS events in AP type, six AS events in AT type, six AS events in RI event, and five AS events in ME type (Fig. 4 and Supplementary Figs. 3-8). We concluded that the death rate of patients increased as the large the risk score was. Two groups of AS events were obtained for ES type: low-risk group (AS events occurred in *MYH11*, *MAF1*, *CBY1*, and *NUP50*) and high-risk group (AS events occurred in *PRKRIP1*, *CBLB*, *SLC7A6*, *LDLRAD3*, *MARK1*, and *INO80Es*). Myosin heavy chain 11 (*MYH11*), encoded by the *MYH11* gene, is a smooth muscle myosin that belongs to the myosin heavy chain family. Previous studies have proved that the mutations of *MYH11* lead to BLCA by involving in cell adhesion, cell migration, and tumor suppression pathway (Nie et al., 2020; Ning and Deng, 2017). For AA type, AS events located in in *FDPS*, *BRD1*, *EMC9*, *PATZ1*, *C21orf59*, and *DTNA* genomic regions are the risk factors that affect the survival

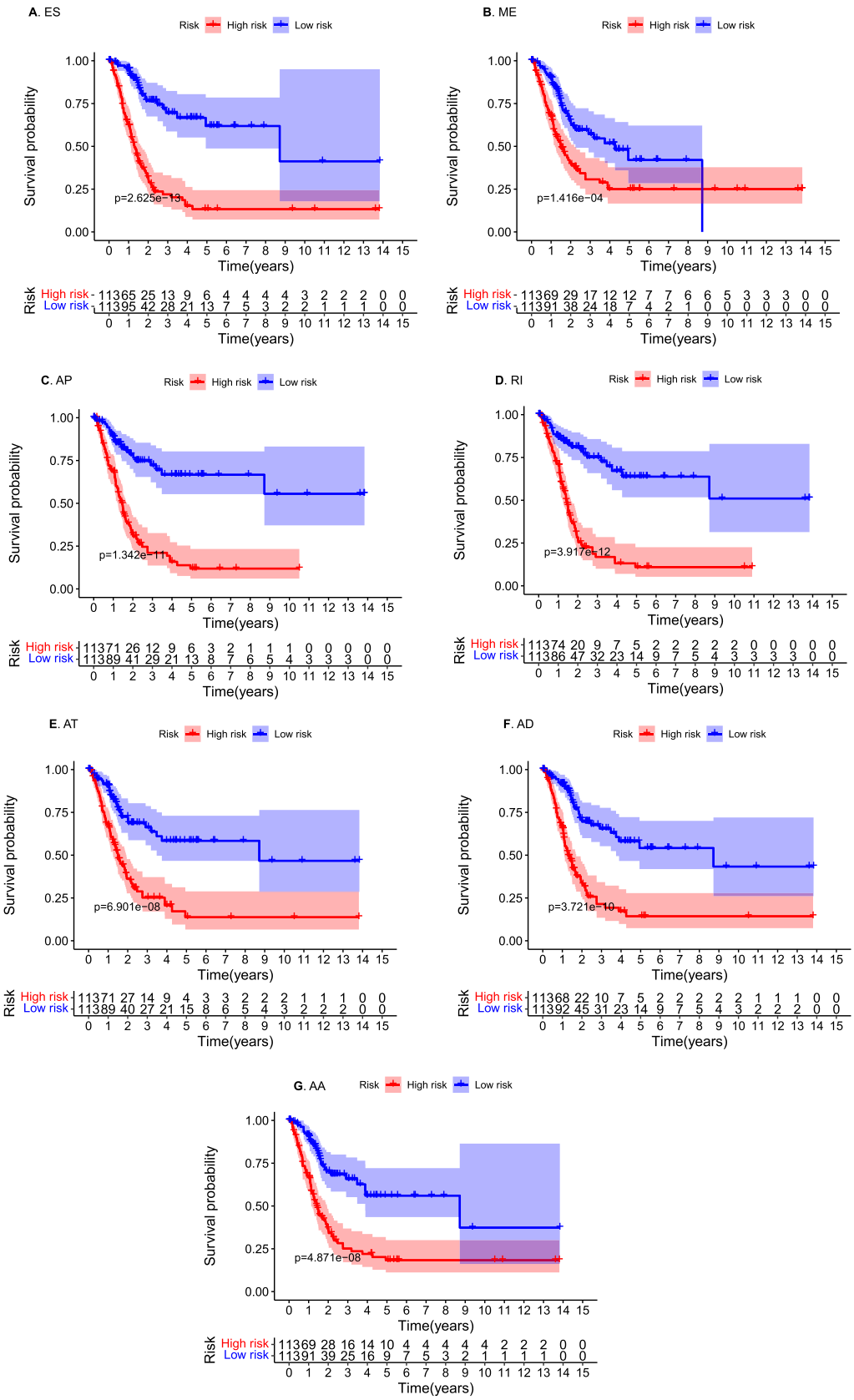

Fig. 6. The ROC curves for seven kinds of AS events. ROC curves are built by clinical features. All models constructed are satisfied with the model prediction efficiency (AUC > 0.65). The top 3 AUC values are ES (AUC = 0.836), AD (AUC = 0.782), and AA (AUC = 0.756).

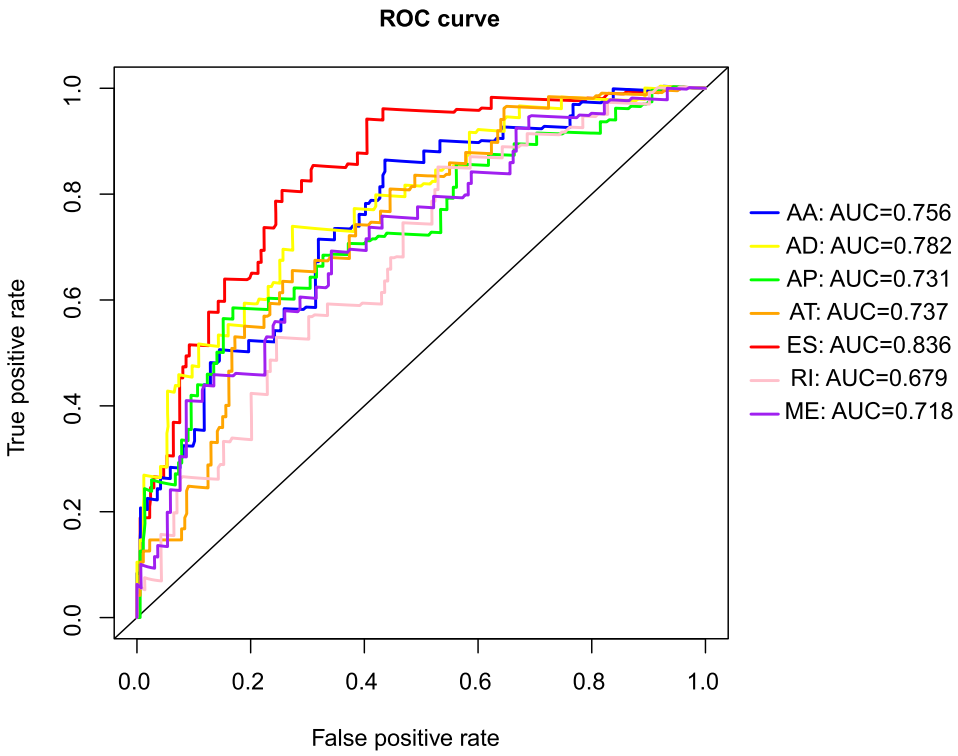

Fig. 7. The correlation network of SF and AS events. The purple triangle indicates SF; the green circle indicates the AS events with good prognostic states (HR < 1), the red circle indicates the AS events with poor prognostic states (HR > 1). The red line represents this relationship as a positive regulation (PCC > 0.6), while the green line represents this regulation as a negative regulation (PCC < -0.6). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

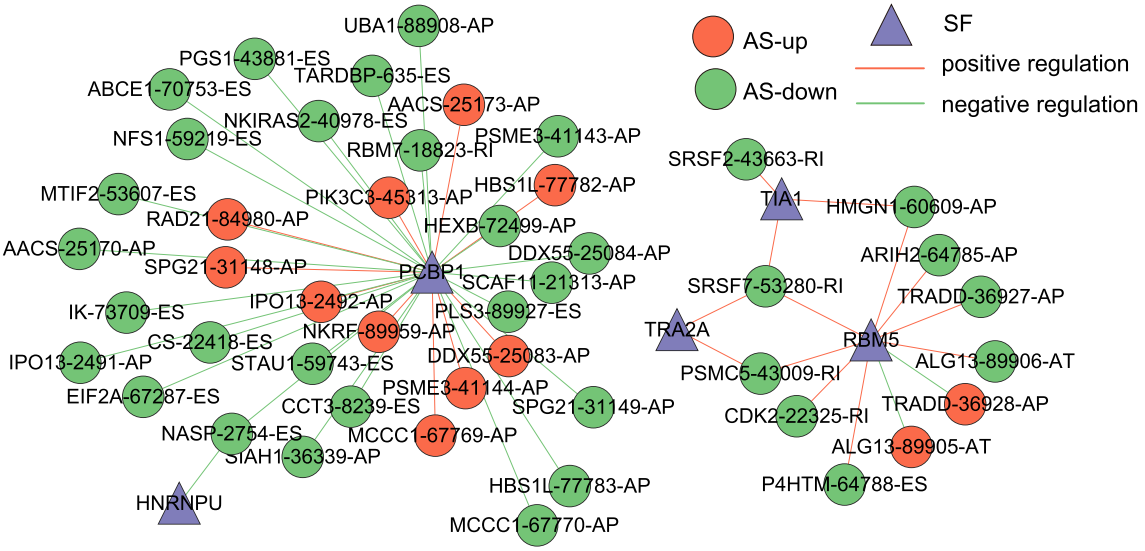

Fig. 8. Representative AS events and SFs. (A) The PSI values of SRSF7-53280-RI across 33 TCGA cancers. The red box indicates the PSI values in cancer samples, while the green box indicates the PSI values in normal samples. There are RI ASs occurred in SRSF7 from exon 4.1 to exon 4.6. (B) The IGV visualization of TIA1 CLIP-seq around SRSF7 genomic position. There are three replicates' experiments. The value below the peak represents the number of reads. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

times of BLCA patients. Xu et al. reported that *DTNA* is a significant gene by univariate Cox regression model (P < 0.005) (Xu et al., 2019). For AD type, AS events located in *IL32*, *ATXN2*, *SKA2*, *TSEB2*, *APOL4*, *TTC21A*, *NCBP2*, *ACAA1*, and *ZNF706* genes are the risk factors that affect the survival times of BLCA patients. ACAA1 expression is positively correlated with CD4 + T cell infiltration. The copy number variation of *ACAA1* was negatively associated with CD4 + T cell polarization. Feng et al. suggested that *ACAA1* significantly correlated with 13 out of 20 types of cancer, including bladder cancer. Cancers with higher *ACAA1* expression level displayed higher OS, while those with reduced *ACAA1* expression had worst outcomes (Feng and Shen, 2020). They concluded that *ACAA1* acts as a tumor suppressor, by altering the nutrient configuration and immune suppression (Sciarrillo et al., 2020). For AP type, AS events located in *TPM1*, *TXLNA*, *PACS2*, *C9orf9*, *TRIM29*, *MIA3*, and *EFNA3* genes are the risk factors that affect the survival times of BLCA patients. Zhang et al. reported that compared with noncancerous tissues, *TPM1* related to cell cycle, cell proliferation, cell movement, receptor signaling, and viral carcinogenesis, which was significantly downregulated in BLCA (Zhang et al., 2020). For RI type, AS events located in *MGRN1*, *FXYD3*, *PYCR1*, *C19orf54*, *CREBZF*, and *PPP1CB* genes are the risk factors that affect the survival times of BLCA patients. In 2011, scientists identified *FXYD3* has important biological

ramification for the genetic study of UC by mRNA expression levels and IHC validation, and they *FXYD3* may be a promising novel biomarker for the differential diagnosis of renal UC and a promising prognosis marker of UC from bladder (Zhang et al., 2011). Liu et al. identified 18 glucose metabolism-related, DNA methylation-related and survival-related genes, including pyrroline-5-carboxylate reductase 1 (*PYCR1*) (Liu et al., 2021). *PYCR1* expressions extremely correlated to their promotor methylation strengths as well as to tumor stages of bladder cancer patients (Liu et al., 2021). *PYCR1* was found to be able to promote bladder cancer cells' proliferation, migration, and evasion via cell functional experiments (Liu et al., 2021). For ME type, AS events located in *COX14*, *N4BP2L1*, *MTFR1L*, *RPE*, and *TMEM104* genes are the risk factors that affect the survival times of BLCA patients. In 2021, Huang et al. also reported AS events located *TMEM104* can serve as a prognostic-related signature (14). Fig. 5A-5G showed the survival analysis results of these ES, ME, AP, RI, AT, AD, and AA. Further analysis of the prediction model created by one type of AS pattern showed that ES events were more effective for distinguishing the survival outcome of BLCA patients than the predictor models built using the other six types of AS pattern (AUC = 0.836). Notably, seven prognostic prediction models performed well, with an AUC $>$ 0.65.

Finally, we constructed an AS-SF regulation network consisting of five SFs and 46 AS events. In which, *TIA1-RBM5* was selected for validation. IGV results *TIA1* ChIP-seq peaks bind the *RBM5* location.

5. Conclusion

Abnormal AS is widely regarded as a novel indicator of the carcinogenesis process, and SF plays a crucial role in this process. Therefore, we aimed to screen critical AS events and SFs that serve as biomarkers for the carcinogenesis and progression of BLCA. A prognostic model was constructed by different kinds of AS events, which showed significant effects in predicting OS times. However, this study has limitations; for example, we lack experimental validation. Yet, our computational analysis may give the views for researchers at the AS aspect and improve our understanding of AS events and BLCA.

Author contributions

The study conception and design were performed by KW, DF, and LW. Data collection and analysis were performed by ZL, XL, FL. The first draft of the manuscript was written by ZL, XL, and FL. HZ, YZ, YW, YM, FW, WZ, OP, ZY, JL, and QH were involved in grammar checking part. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Institutional Review Board Statement

Not applied.

Data Availability Statement

The public availability dataset of this research is from the accessibility database.

The mRNA expression and clinical data are from the TCGA database (https://portal.gdc.cancer.gov/).

The AS events data is from the TCGA SpliceSeq database (https://bioinformatics.mdanderson.org/public-software/tcgaspliceseq/).

Funding

This research was substantially sponsored by the research projects (Grant No. 32170654 and Grant No. 32000464) supported by the National Natural Science Foundation of China and was substantially supported by the Shenzhen Research Institute, City University of Hong Kong. The work described in this paper was substantially supported by the grant from the Health and Medical Research Fund, the Food and Health Bureau, The Government of the Hong Kong Special Administrative Region [07181426]. The work described in this paper was partially supported by the grants from City University of Hong Kong (CityU 11202219, CityU 11203520, CityU 11203221).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

RNA-seq data and the corresponding clinical information of the BLCA patients generated by the TCGA Research Network: https://www.cancer.gov/tcga. We thank the Tongyuan gene company for server computational resource support and Shengxin.Ren company for technical support.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2022.146657.

References

Black, A.J., Gamarra, J.R., Giudice, J., n.d. More than a messenger: Alternative splicing as a therapeutic target. Biochim. Biophys. Acta Gene Regul. Mech. 1862, 194395. https://doi.org/10.1016/j.bbagrm.2019.06.006.

Black, D.L., 2003. Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem 72 (1), 291–336. https://doi.org/10.1146/annurev.biochem.72.121801.161720.

Bonnal, S.C., López-Oreja, I., Valcárcel, J., 2020. Roles and mechanisms of alternative splicing in cancer - implications for care. Nat. Rev. Clin. Oncol. 17 (8), 457–474. https://doi.org/10.1038/s41571-020-0350-x.

Chen, C., Shang, A., Sun, Z., Gao, Y., Huang, J., Ping, Y., Chang, W., Gu, C., Sun, J., Ji, P., Yuan, Y.i., Lu, R., Li, D., Batista, W., 2022. Urinary Exosomal Long Noncoding RNA TERC as a Noninvasive Diagnostic and Prognostic Biomarker for Bladder Urothelial Carcinoma. J Immunol Res 2022, 1–9. https://doi.org/10.1155/2022/9038808.

Cieply, B., Carstens, R.P., 2015. Functional roles of alternative splicing factors in human disease: Functional roles of alternative splicing factors in human disease. WIREs RNA 6 (3), 311–326.

Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S., Malta, T.M., Pagnotta, S.M., Castiglioni, I., Ceccarelli, M., Bontempi, G., Noushmehr, H., 2016. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res 44, e71. https://doi.org/10.1093/nar/gkv1507.

Du, J.-X., Zhu, G.-Q., Cai, J.-L., Wang, B., Luo, Y.-H., Chen, C., Cai, C.-Z., Zhang, S.-J., Zhou, J., Fan, J., Zhu, W., Dai, Z., 2021. Splicing factors: Insights into their regulatory network in alternative splicing in cancer. Cancer Lett. 501, 83–104. https://doi.org/10.1016/j.canlet.2020.11.043.

Feng, H., Shen, W., 2020. ACAA1 Is a Predictive Factor of Survival and Is Correlated With T Cell Infiltration in Non-Small Cell Lung Cancer. Front. Oncol. 10 https://doi.org/10.3389/fonc.2020.564796.

Gómez De Liaño, A., Duran, I., 2018. The continuing role of chemotherapy in the management of advanced urothelial cancer. Ther. Adv. Urol. 10 (12), 455–480. https://doi.org/10.1177/1756287218814100.

Gonçalves, V., Pereira, J., Jordan, P., 2017. Signaling Pathways Driving Aberrant Splicing in Cancer Cells. Genes (Basel) 9 (1), 9. https://doi.org/10.3390/genes9010009.

Hamdollah Zadeh, M.A., Amin, E.M., Hoareau-Aveilla, C., Domingo, E., Symonds, K.E., Ye, X., Heesom, K.J., Salmon, A., D'Silva, O., Betteridge, K.B., Williams, A.C., Kerr, D.J., Salmon, A.H.J., Oltean, S., Midgley, R.S., Ladomery, M.R., Harper, S.J., Varey, A.H.R., Bates, D.O., 2015. Alternative splicing of TIA-1 in human colon cancer regulates VEGF isoform expression, angiogenesis, tumour growth and bevacizumab resistance. Mol. Oncol. 9, 167–178. https://doi.org/10.1016/j.molonc.2014.07.017.

Hanna, K.S., 2017. A Review of Immune Checkpoint Inhibitors for the Management of Locally Advanced or Metastatic Urothelial Carcinoma. Pharmacotherapy 37 (11), 1391–1405. https://doi.org/10.1002/phar.2033.

Huang, S., Luo, K., Jiang, L., Zhang, X.-D., Lv, Y.-H., Li, R.-F., 2021. PCBP1 regulates the transcription and alternative splicing of metastasis–related genes and pathways in hepatocellular carcinoma. Sci. Rep. 11, 23356. https://doi.org/10.1038/s41598-021-02642-z.

Koedoot, E., Smid, M., Foekens, J.A., Martens, J.W.M., le Dévédec, S.E., van de Water, B., 2019. Co-regulated gene expression of splicing factors as drivers of cancer progression. Sci. Rep. 9, 5484. https://doi.org/10.1038/s41598-019-40759-4.

Lee, Y., Rio, D.C., 2015. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu. Rev. Biochem. 84 (1), 291–323. https://doi.org/10.1146/annurev-biochem-060614-034316.

Li, X., Feng, J., Sun, Y., Li, X., 2021. An Exploration of the Tumor Microenvironment Identified a Novel Five-Gene Model for Predicting Outcomes in Bladder Cancer. Front. Oncol. 11 https://doi.org/10.3389/fonc.2021.642527.

Liu, Z., Sun, T., Zhang, Z., Bi, J., Kong, C., 2021. An 18-gene signature based on glucose metabolism and DNA methylation improves prognostic prediction for urinary bladder cancer. Genomics 113 (1), 896–907. https://doi.org/10.1016/j.ygeno.2020.10.022.

Magi-Galluzzi, C., Falzarano, S.M., Zhou, M., 2008. Urothelial Carcinoma and its Variants. Surg. Pathol. Clin. 1 (1), 159–209. https://doi.org/10.1016/j.path.2008.07.004.

Montes, M., Sanford, B.L., Comiskey, D.F., Chandler, D.S., 2019. RNA Splicing and Disease: Animal Models to Therapies. Trends Genet. 35 (1), 68–87. https://doi.org/10.1016/j.tig.2018.10.002.

Mounir, M., Lucchetta, M., Silva, T.C., Olsen, C., Bontempi, G., Chen, X.i., Noushmehr, H., Colaprico, A., Papaleo, E., Wang, E., 2019. New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. PLoS Comput. Biol. 15 (3), e1006701. https://doi.org/10.1371/journal.pcbi.1006701.

Nie, M.-J., Pan, X.-T., Tao, H.-Y., Xu, M.-J., Liu, S.-L., Sun, W., Wu, J., Zou, X., 2020. Clinical and prognostic significance of MYH11 in lung cancer. Oncol. Lett. 19, 3899–3906. https://doi.org/10.3892/ol.2020.11478.

Ning, X., Deng, Y., 2017. Identification of key pathways and genes influencing prognosis in bladder urothelial carcinoma. Onco Targets Ther. 10, 1673–1686. https://doi.org/10.2147/OTT.S131386.

Paik, P.K., Drilon, A., Fan, P.-D., Yu, H., Rekhtman, N., Ginsberg, M.S., Borsu, L., Schultz, N., Berger, M.F., Rudin, C.M., Ladanyi, M., 2015. Response to MET inhibitors in patients with stage IV lung adenocarcinomas harboring MET mutations causing exon 14 skipping. Cancer Discov. 5, 842–849. https://doi.org/10.1158/2159-8290.CD-14-1467.

Parker, J., Spiess, P.E., 2011. Current and emerging bladder cancer urinary biomarkers. ScientificWorldJournal 11, 1103–1112. https://doi.org/10.1100/tsw.2011.104.

Peng, F., Li, H., Xiao, H., Li, L., Li, Y., Wu, Y., 2017. Identification of a three miRNA signature as a novel potential prognostic biomarker in patients with bladder cancer. Oncotarget 8, 105553–105560. https://doi.org/10.18632/oncotarget.22318.

Piva, F., Giulietti, M., Burini, A.B., Principato, G., 2012. SpliceAid 2: a database of human splicing factors expression data and RNA target motifs. Hum. Mutat. 33 (1), 81–85. https://doi.org/10.1002/humu.21609.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29 (1), 24–26. https://doi.org/10.1038/nbt.1754.

Ryan, M., Wong, W.C., Brown, R., Akbani, R., Su, X., Broom, B., Melott, J., Weinstein, J., 2016a. TCGASpliceSeq a compendium of alternative mRNA splicing in cancer. Nucleic Acids Res. 44 (D1), D1018–D1022. https://doi.org/10.1093/nar/gkv1288.

Sciarrillo, R., Wojtuszkiewicz, A., Assaraf, Y.G., Jansen, G., Kaspers, G.J.L., Giovannetti, E., Cloos, J., 2020. The role of alternative splicing in cancer: From oncogenesis to drug resistance. Drug Resist Updat 53, 100728. https://doi.org/10.1016/j.drup.2020.100728.

Stenehjem, D.D., Tran, D., Nkrumah, M.A., Gupta, S., 2018. PD1/PDL1 inhibitors for the treatment of advanced urothelial bladder cancer. Onco Targets Ther 11, 5973–5989. https://doi.org/10.2147/OTT.S135157.

Ule, J., Blencowe, B.J., 2019. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. Mol. Cell 76 (2), 329–345. https://doi.org/10.1016/j.molcel.2019.09.017.

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L.i., Fu, X., Liu, S., Bo, X., Yu, G., 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (N Y) 2 (3), 100141. https://doi.org/10.1016/j.xinn.2021.100141.

Xu, Z., Wang, C., Xiang, X., Li, J., Huang, J., 2019. Characterization of mRNA Expression and Endogenous RNA Profiles in Bladder Cancer Based on The Cancer Genome Atlas (TCGA) Database. Med. Sci. Monit. 25, 3041–3060. https://doi.org/10.12659/MSM.915487.

Yang, D., Ma, Y., Zhao, P., Ma, J., He, C., 2019. Systematic screening of protein-coding gene expression identified HMMR as a potential independent indicator of unfavorable survival in patients with papillary muscle-invasive bladder cancer. Biomed. Pharmacother. = Biomed. Pharmacother. 120, 109433. https://doi.org/10.1016/j.biopha.2019.109433.

Yu, G., Wang, L.-G., Han, Y., He, Q.-Y., 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16 (5), 284–287. https://doi.org/10.1089/omi.2011.0118.

Zhang, C., Berndt-Paetz, M., Neuhaus, J., 2020. Identification of Key Biomarkers in Bladder Cancer: Evidence from a Bioinformatics Analysis. Diagnostics (Basel) 10 (2), 66. https://doi.org/10.3390/diagnostics10020066.

Zhang, C., Zhang, Z., Castle, J., Sun, S., Johnson, J., Krainer, A.R., Zhang, M.Q., 2008. Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. Genes Dev. 22 (18), 2550–2563. https://doi.org/10.1101/gad.1703108.

Zhang, M., Zhang, X., Yu, M., Zhang, W., Zhang, D.i., Zeng, S., Wang, X.i., Hu, X., 2021. A Novel Ferroptosis-Related Gene Model for Overall Survival Predictions of Bladder Urothelial Carcinoma Patients. Front. Oncol. 11 https://doi.org/10.3389/fonc.2021.698856.

Zhang, Z., Pang, S.-T., Kasper, K.A., Luan, C., Wondergem, B., Lin, F., Chuang, C.-K., Teh, B.T., Yang, X.J., 2011. FXYD3: A Promising Biomarker for Urothelial Carcinoma. Biomark Insights 6, 17–26. https://doi.org/10.4137/BMI.S6487.

Zhu, K., Liu, X., Deng, W., Wang, G., Fu, B., 2022. Identification of a chromatin regulator signature and potential candidate drugs for bladder cancer. Hereditas 159, 13. https://doi.org/10.1186/s41065-021-00212-x.