

## ORIGINAL ARTICLE

# Systematic profiling of alternative splicing signature reveals prognostic predictor for prostate cancer

Jiyu Zhao<sup>1</sup> | Luchen Chang<sup>2</sup> | Xianen Gu<sup>1</sup> | Jia Liu<sup>1</sup> | Bei Sun<sup>3</sup> | Xi Wei<sup>2</sup> 

<sup>1</sup>Department of Urology, ChuiYangLiu Hospital affiliated to Tsinghua University, Beijing, China

<sup>2</sup>Department of Diagnostic and Therapeutic Ultrasonography, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin, China

<sup>3</sup>Department of Outpatient Office, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China

## Correspondence

Bei Sun, Department of Outpatient office, Tianjin Medical University Cancer Institute and Hospital, Tianjin Hexi District, Ti Yuan Bei, Huan Huxi Road, 300060 Tianjin, China. Email: sunpei003@sina.com

Xi Wei, Department of Diagnostic and Therapeutic Ultrasonography, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin, China. Email: weixi@tmu.edu.cn

## Funding information

National Scientific Foundation of China, Grant/Award Number: 81771852

## Abstract

Alternative splicing (AS) provides the primary mechanism for producing protein diversity. There is growing evidence that AS is involved in the development and progression of cancers. The rapid accumulation of high-throughput sequencing technologies and clinical data sets offers an opportunity to systematically profile the relationship between mRNA variants and clinical outcomes. However, there is a lack of systematic analysis of AS in prostate cancer: Download RNA-seq data and clinical information from The Cancer Genome Atlas (TCGA) data portal. Evaluate RNA splicing patterns by SpliceSeq and calculate splicing percentage (PSI) values. Different expressions were identified as differently expressed AS events (DEAs) based on PSI values. Bioinformatics methods were used for further analysis of DEAs and their splicing networks. Kaplan-Meier, Cox proportional regression, and unsupervised cluster analysis were used to assess the correlation between DEAs and clinical characteristics. In total, 43 834 AS events were identified, of which 1628 AS events were differentially expressed. The parental genes of these DEAs played a significant role in the regulation of prostate cancer-related processes. In total, 226 DEAs events were found to be associated with disease-free survival. Four clusters of molecules with different survival modes were revealed by unsupervised cluster analysis of DEAs. AS events may be important determinants of prognosis and bio-modulation in prostate cancer. In this study, we established strong prognostic predictors, discovered a splicing network that may be a potential mechanism, and provided further validated therapeutic targets.

## KEYWORDS

alternative splicing, prognosis, prostate cancer, RNA-seq

## 1 | INTRODUCTION

Alternative splicing of RNA refers to the process by which mRNA precursors are cleaved at different sites to produce different

mRNA splicing bodies, which are then translated into different proteins with different functions. Alternative splicing is an important factor in the transcriptomics and proteomics diversification of eukaryotes. Studies have shown that 95% of human

**Abbreviations:** AA, alternate acceptor site; AD, alternate donor site; AP, alternate promoter; AR, androgen receptor; AR-V7, androgen receptor splice variant 7; AS, alternative splicing; AT, alternate terminator; AUC, area under the curve; CRPC, castration-resistant prostate cancer; DEAs, differently expressed alternative splicing events; DFS, disease-free survival; ES, exon skip; GO, gene ontology; HR, hazard ratio; M, metastasis; ME, mutually exclusive exons; N, lymph node; PSI, percent spliced in; RI, retained intron; ROC, receiver operating characteristic; T, tumor; TCGA, The Cancer Genome Atlas; VEGF, vascular endothelial growth factor.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. *Cancer Science* published by John Wiley & Sons Australia, Ltd on behalf of Japanese Cancer Association.

genes could be spliced.<sup>1</sup> In recent years, increasing evidence has suggested that splicing defects and the production of specific isomers are drivers for cancer.<sup>2-4</sup> Furthermore, AS is closely correlated with tumor proliferation, apoptosis, hypoxia, angiogenesis, immune escape, and metastasis.<sup>4,5</sup> Conversely, mutations or changes in the expression of the splicing factor may result in specific cancer-promoting splicing bodies, providing growth or survival advantages for the tumor cells.<sup>6</sup> Therefore, abnormal AS is considered to be another marker of cancer, and systematic studies of AS may provide potential biomarkers for malignant tumors.

Prostate cancer is one of the most common malignant tumors in the male reproductive system, occurring mainly in the elderly (age > 65 y)<sup>7,8</sup> and ranking second among male cancer-related deaths.<sup>9,10</sup> The importance of AS in the progression of prostate cancer has been demonstrated in many studies.<sup>11,12</sup> Growing evidence suggests that splicing events and dysregulation of cancer-specific splicing variants can serve as prognostic biomarkers and therapeutic targets for prostate cancer.<sup>13-15</sup> Therefore, identifying the link between splicing disorders and prostate cancer, especially accurate molecular analysis, is a key issue in ongoing cancer research. Therefore, there is an urgent need to exploit the genome-wide transcriptome approach to develop the potential prognostic value of AS events in patients with prostate cancer.

With the rapid accumulation of RNA-seq data in prostate cancer samples, TCGA project provides a rich source for AS model research. At the same time, TCGA database provides detailed clinical information. Therefore, it is possible to correlate prostate cancer-associated AS events in a larger population. RNA-seq has great advantages for the study of cancer, but requires efficient and reliable bioinformatics for data mining. The recently developed analysis tool, SpliceSeq, aligns RNA-seq readings with gene maps for accurate analysis of complex or low frequency AS events.<sup>16</sup>

We used RNA-seq data from TCGA program data to elucidate the role of differential alternative splicing patterns in 433 prostate cancer cohorts and further systematically examined the potential prognostic impact of prostate cancer-specific AS events on patient survival. The aim of this study was to explore differential RNA splicing patterns in prostate cancer and to elucidate the function of splicing variants as prognostic biomarkers. This research will help to develop new therapeutic targets for prostate cancer.

## 2 | MATERIALS AND METHODS

### 2.1 | Alternative splicing event in TCGA RNA-seq data

TCGA Data Portal (<https://portal.gdc.cancer.gov/projects>) provided RNA-seq data for TCGA prostate cancer. We used the SpliceSeq tool to analyze AS profiles and to assess the splicing pattern of genes

encoding protein in patients with prostate cancer. The PSI values, from 0 to 1, were used to quantify AS events and calculate 7 types of variable splicing events: ES, ME, RI, AP, AT, AD, and AA. The expression data of prostate cancer, relevant clinical information (DFS > 30 d), and variable splicing data were integrated. Finally, 433 prostate cancer samples and 46 normal control samples were selected for subsequent study and analysis.

### 2.2 | Differential analysis of alternative splicing events

In this study, a comparison was made between the prostate cancer samples and the adjacent normal AS samples to identify differentially alternative splicing events, and  $P < .05$  and  $|\text{Log FC}| > 1$  were defined as differential alternative splicing events. At the same time, we also analyzed the gene expression profiles of prostate cancer samples and normal cancer samples, and identified differentially expressed genes for further comparison.

### 2.3 | Survival analysis

In total, 433 patients with prostate cancer with DFS exceeding 30 d were included in the study of survival analysis. Patients were divided into 2 groups based on the median cutoff value for each parameter. Univariate Cox regression was performed on 7 types of candidate variable splicing events to identify the relationship between AS events and DFS. A  $P$ -value less than .05 was defined as an AS event associated with survival. At the same time, we used the clusterProfiler package to perform GO function enrichment analysis on survival-related AS events, and identified significantly related biological processes, cellular components, and molecular functions. To remove any gene that may not be an independent factor in prognostic predictors, multivariate Cox regression was further applied to the survival-related variable splicing events in 7 types. Finally, 7 different types of candidate independent prognostic AS events were combined to construct a final prognostic predictor. In addition, we plotted the ROC curve to compare the accuracy of prognostic models in each type of AS.

### 2.4 | Splicing related network construction

A list of 67 human splicing factors was created through hand-planned literature and database screening.<sup>17</sup> The expression of the splicing factor gene in the mRNA splicing pathway was derived from level 3 mRNA-seq data in TCGA. Survival analysis was used to identify survival-related splicing factors, and Spearman test was used to analyze the correlation between survival-related splicing factor gene expression and survival-related AS PSI values. When the  $P$ -value was less than .05, it was defined as a significant correlation and Cytoscape (v.3.6.0) software was used to construct the final interaction network

of AS events and splicing factors. At the same time, we used clusterProfiler to perform GO functional enrichment analysis on genes associated with AS networks to identify significantly correlated biological processes, cellular components, and molecular functions.

## 2.5 | Identify prognosis and analyze subtype-related clusters

Alternative splicing events occurred very differently at the individual level. To obtain robust classification, we used the unsupervised consensus approach implemented by Consensus Cluster Plus (R package) to identify molecular subtypes of prostate cancer. We analyzed survival of the identified molecular subtypes from the survival time and then identified the relationship between subtypes and survival, and further explored the relevant clinical information, trying to find a relationship between other clinical information and molecular subtypes. Details of this study design are illustrated as a flowchart in Figure 1.

## 3 | RESULTS

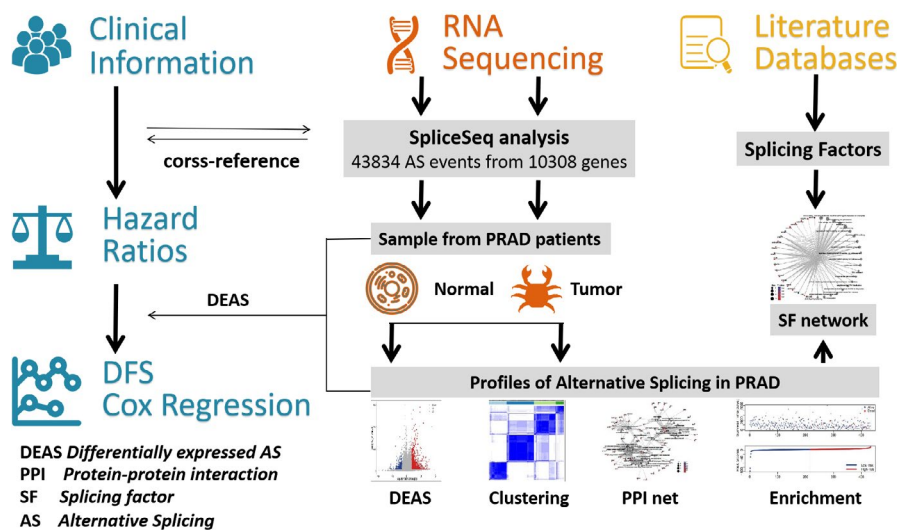
### 3.1 | Integrating alternative splicing events in prostate cancer

Of the 433 prostate cancer patients, 43 834 AS events associated with 10 308 genes were found. We detected 3503 AA-type AS events containing 2469 genes, 3088 AD-type AS events involving

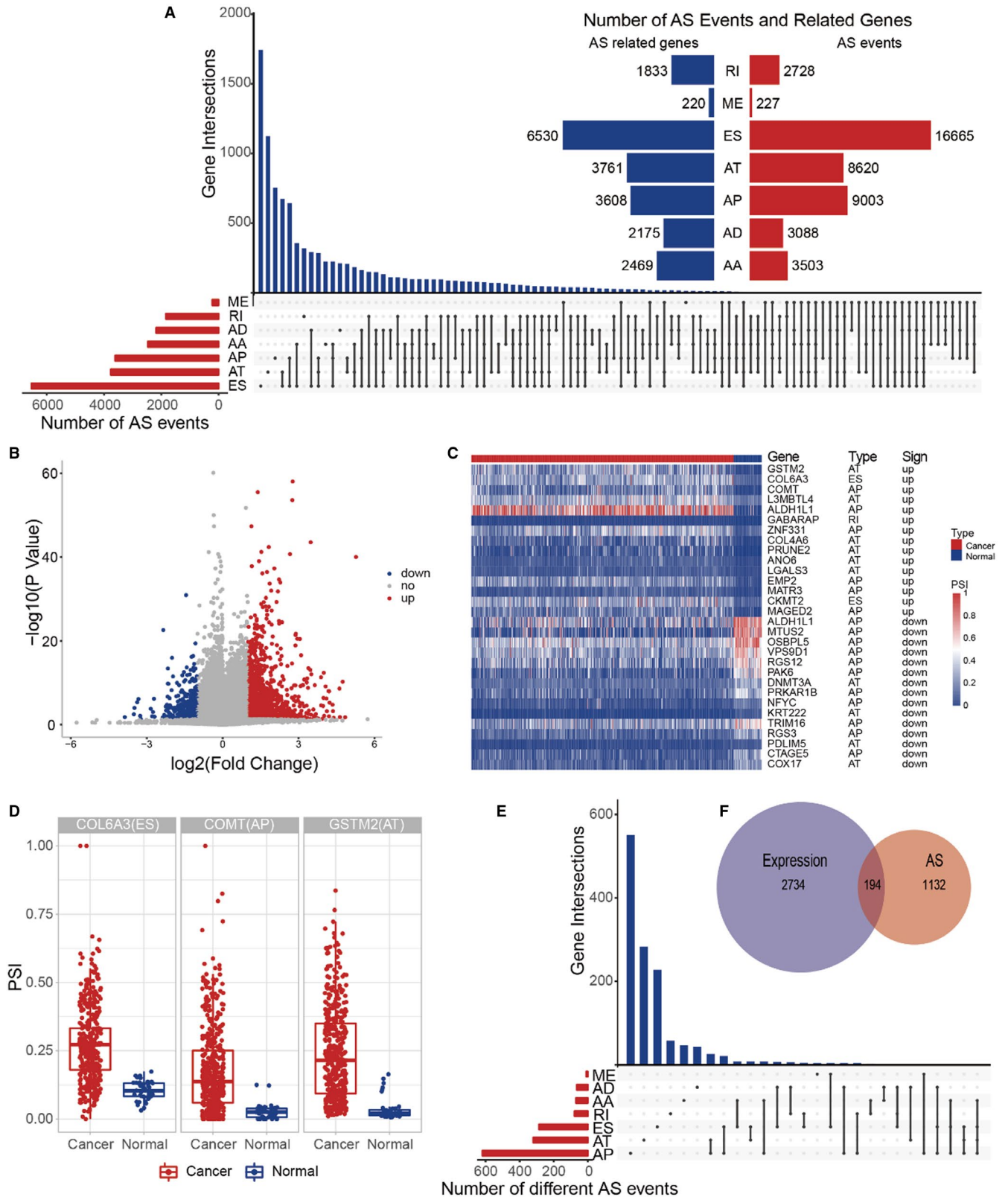
2175 genes, 9003 AP-type AS events including 3608 genes, 8620 AT-type AS events involving 3761 genes, 16 665 ES-type AS events containing 6530 genes, 227 ME-type AS events involving 220 genes, 2728 RI type of AS event containing 1833 genes, as shown in Figure 2A. The results indicated that one gene may have several types of mRNA splicing events, and one gene may be expressed by 4 AS types, while ES was the main type, because more than one-third of the AS events were ES events.

### 3.2 | Differential alternative splicing events in prostate cancer

Differential AS analysis revealed that 1628 differential AS events were identified in prostate cancer involving 1326 genes, of which 1338 were differentially upregulated AS events. Also, 290 were differentially downregulated AS events, with upregulated events in the main body, as shown in Figure 2B, in which the upregulated and downregulated TOP 15 differential mutagenic events-related expression and partial event cancer were compared with normal samples, as shown in Figure 2C,D and Table 1. Some genes could be expressed in up to 4 AS types, as shown in Figure 2E. Differential expression profiling revealed that 2928 differential expression genes were identified in prostate cancer. We compared the genes involved in the differential AS event with the differentially expressed genes, and found that the genes recognized in both had similarities, but there were also significant differences; the AS events could make up for shortcomings of differential expression analysis, as shown in Figure 2F.



**FIGURE 1** Flowchart for profiling the alternative splicing (AS) of prostate cancer using large-scale RNA-seq data. All data including clinical information were downloaded from The Cancer Genome Atlas (TCGA) data portal. We combined these datasets into a cohort to be analyzed. Through this process, PSI value was calculated for each AS events. Based on the data, we identified different alternative splicing events between prostate cancer and normal tissues. We then investigated the parent genes of these AS events by enrichment analysis. Next, we analyzed an interaction network of genes and a regulation network of different AS events. Finally, we assessed the prognostic value of 4 clusters using Kaplan-Meier analysis



**FIGURE 2** A, Overview of alternative splicing (AS) events and their interactions between genes. One gene may have several types of AS. B, Difference of AS events. Volcano plot visualizing the DEAs identified in prostate cancer. The red and blue points in the plot represent the differentially expressed AS events with statistical significance. C, Heat map of the AS events. The horizontal axis shows the clustering information of samples which were divided into 2 major clusters, and the clusters were normal tissue and paired tumor tissue, respectively (Top 15 events). D, Boxplot of the difference of AS events between normal tissue and paired tumor tissue (Top 3 events). E, UpSet plot of different AS events and their interactions between genes. F Venn diagram demonstrated the intersection set of gene and different alternative splicing events

**TABLE 1** The detailed information of the top 30 most different AS events

Symbol	AS type	Exons	MeanT	MeanN	logFC	P-value
Upregulated						
GSTM2	AT	13	0.240	0.035	0.832	9.45E-59
COL6A3	ES	6	0.273	0.104	0.417	3.18E-56
COMT	AP	3	0.179	0.027	0.828	2.68E-54
L3MBTL4	AT	19	0.330	0.150	0.342	4.45E-48
ALDH1L1	AP	1	0.660	0.059	1.046	3.08E-44
GABARAP	RI	1.2:1.3:1.4	0.021	0.006	0.549	3.94E-43
ZNF331	AP	1	0.285	0.099	0.460	5.96E-42
COL4A6	AT	52	0.135	0.021	0.801	2.03E-41
PRUNE2	AT	7	0.089	0.002	1.586	9.68E-41
ANO6	AT	25	0.087	0.032	0.432	9.55E-40
LGALS3	AT	6	0.076	0.034	0.346	1.54E-38
EMP2	AP	2	0.224	0.055	0.614	5.14E-38
MATR3	AP	1	0.084	0.025	0.527	3.69E-37
CKMT2	ES	2	0.283	0.096	0.472	7.88E-37
MAGED2	AP	3	0.167	0.036	0.670	2.21E-35
Downregulated						
ALDH1L1	AP	3	0.281	0.766	-0.435	1.19E-31
MTUS2	AP	7	0.140	0.711	-0.706	2.69E-23
OSBPL5	AP	1.1	0.345	0.728	-0.324	4.02E-20
VPS9D1	AP	1	0.266	0.597	-0.351	4.05E-17
RGS12	AP	7.1	0.226	0.529	-0.370	6.46E-17
PAK6	AP	2	0.112	0.450	-0.602	1.16E-16
DNMT3A	AT	5.2	0.127	0.278	-0.340	3.41E-15
PRKAR1B	AP	5	0.127	0.341	-0.427	2.22E-14
NFYC	AP	4	0.068	0.151	-0.349	2.87E-14
KRT222	AT	5.2	0.016	0.073	-0.652	3.91E-14
TRIM16	AP	6.1	0.251	0.529	-0.324	4.05E-14
RGS3	AP	24	0.111	0.267	-0.380	9.32E-14
PDLIM5	AT	4	0.031	0.069	-0.344	1.29E-13
CTAGE5	AP	1	0.092	0.267	-0.461	3.01E-13
COX17	AT	7	0.098	0.24	-0.391	.00

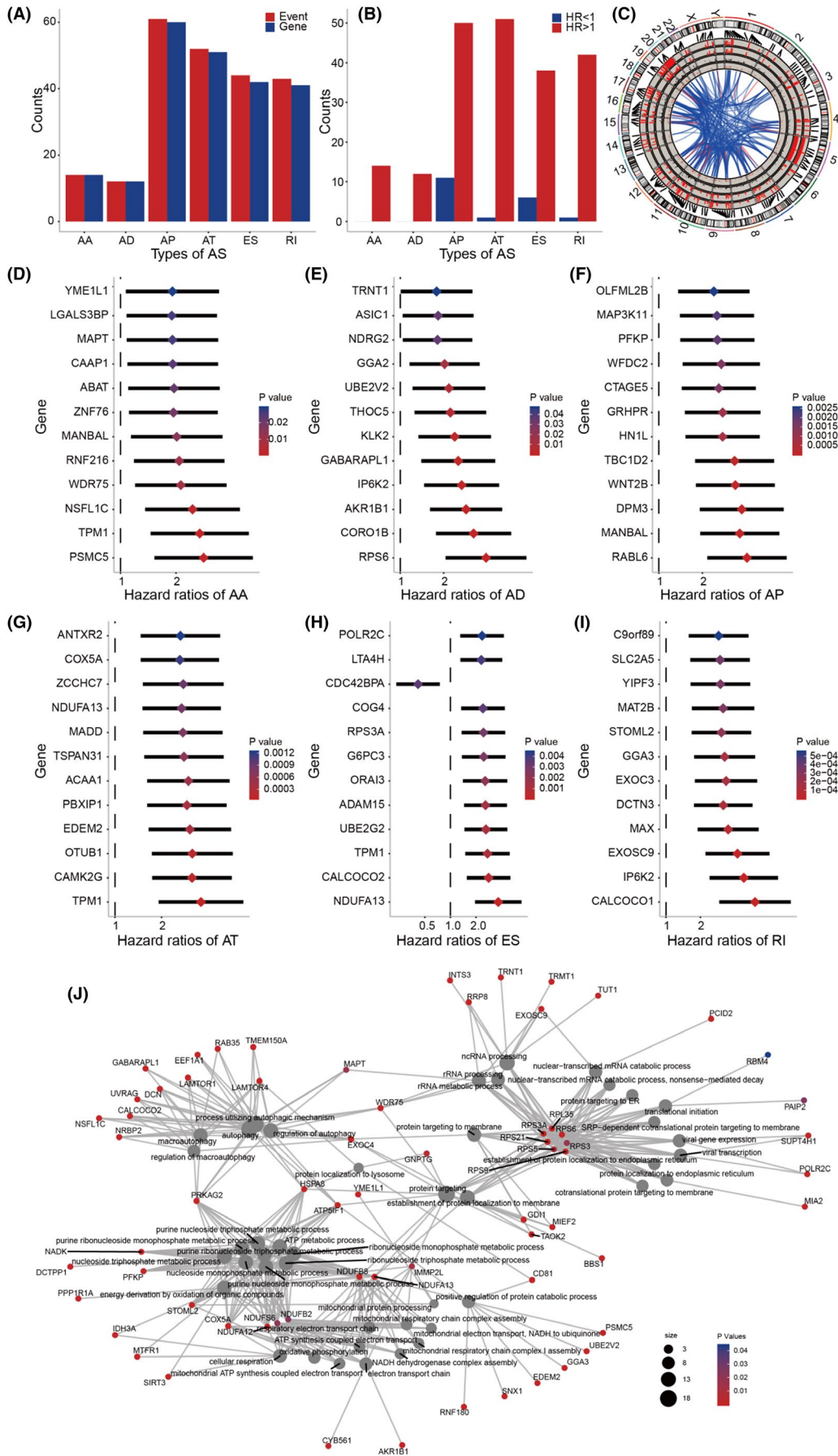
Note: The P-value was calculated by *t* test.

Abbreviations: logFC, log<sub>2</sub> fold change; MeanN, the mean PSI value in paired normal tissue; MeanT, the mean PSI value in prostate cancer tissue.

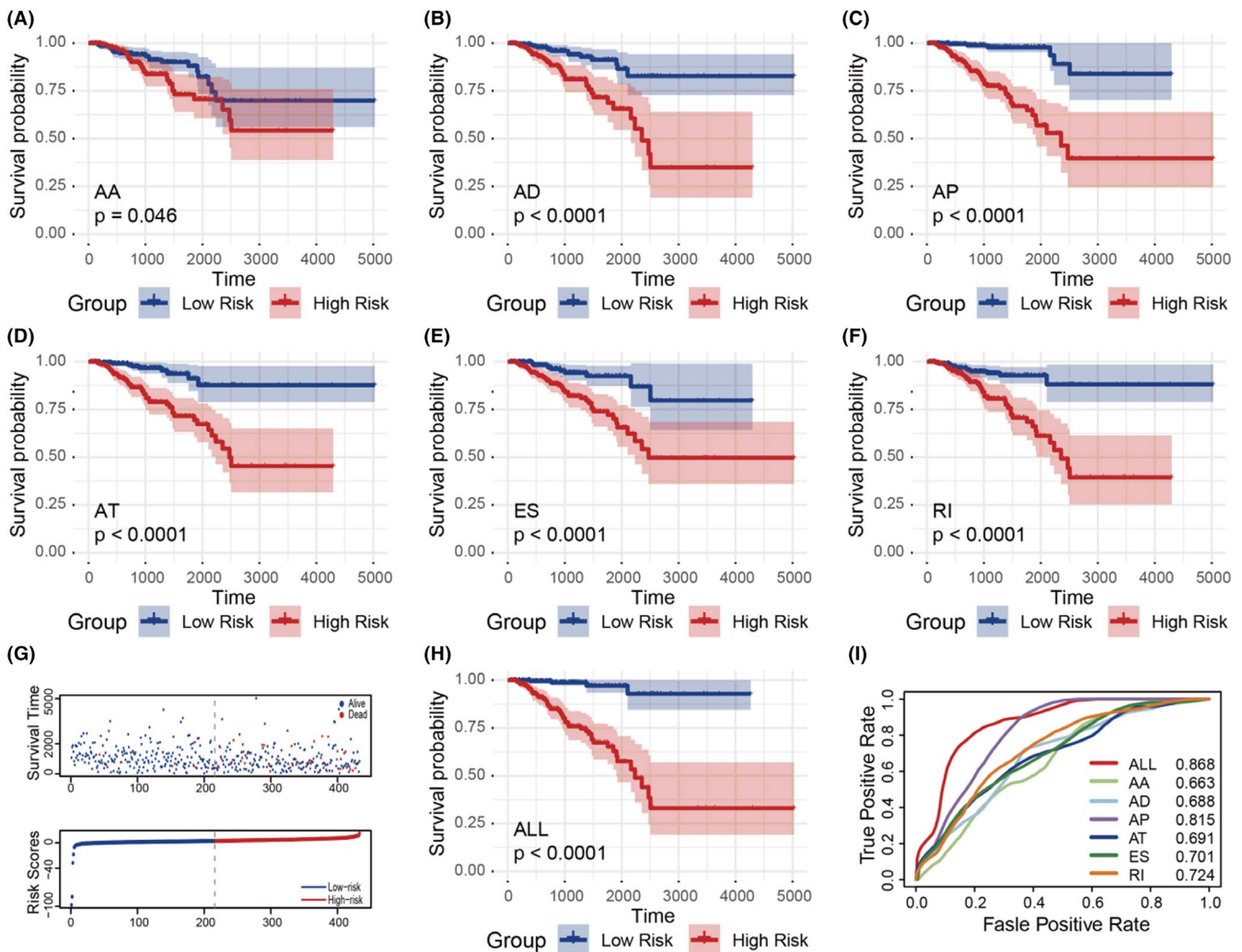
### 3.3 | Alternative splicing events associated with survival in prostate cancer

To investigate the prognostic value of AS events in patients with prostate cancer, we used univariate Cox regression analysis to assess the prognostic impact of differential AS events in patients with prostate cancer. We detected a total of 226 survival-related AS events ( $P < .05$ ) among the differential alternative splicing events, as shown in Figure 3A. It can be further seen, among the recognized survival-related AS events, that adverse factors ( $HR > 1$ ) were found more frequently than protective factors ( $HR < 1$ ), either for the whole

or for each type, as shown in Figure 3B. The Circos diagram shows details of survival-related AS events and their associated genes, as seen in Figure 3C. We drew forest maps by spotting the TOP 12 most important survival-related AS events among 6 AS types (AS events associated with survival were not found in the ME type), as shown in Figure 3D-I. Subsequently, GO function enrichment analysis was performed on 211 genes involved in survival-related AS events, and biological processes, cell components, and molecular functions ( $P < .05$ ) with significant concentrations of AS events were found; biological processes with significant representation are shown in Figure 3J.



**FIGURE 3** A, Number of prognosis-related alternative splicing (AS) events and involved genes. B, The distribution of different types of AS between beneficial to survival (HR < 1) and unfavorable to survival (HR > 1). C, Detail of AS events and related genes was shown in a Circos plot. The ribbons represent the potential interaction between the related gene of differentially expressed AS, and the thickness of the ribbons indicating the extent of the interaction strength. D-I, The detail of survival-related AS events and types of AS in Forest map (top 12). J, Gene network construction and functional enrichment analysis

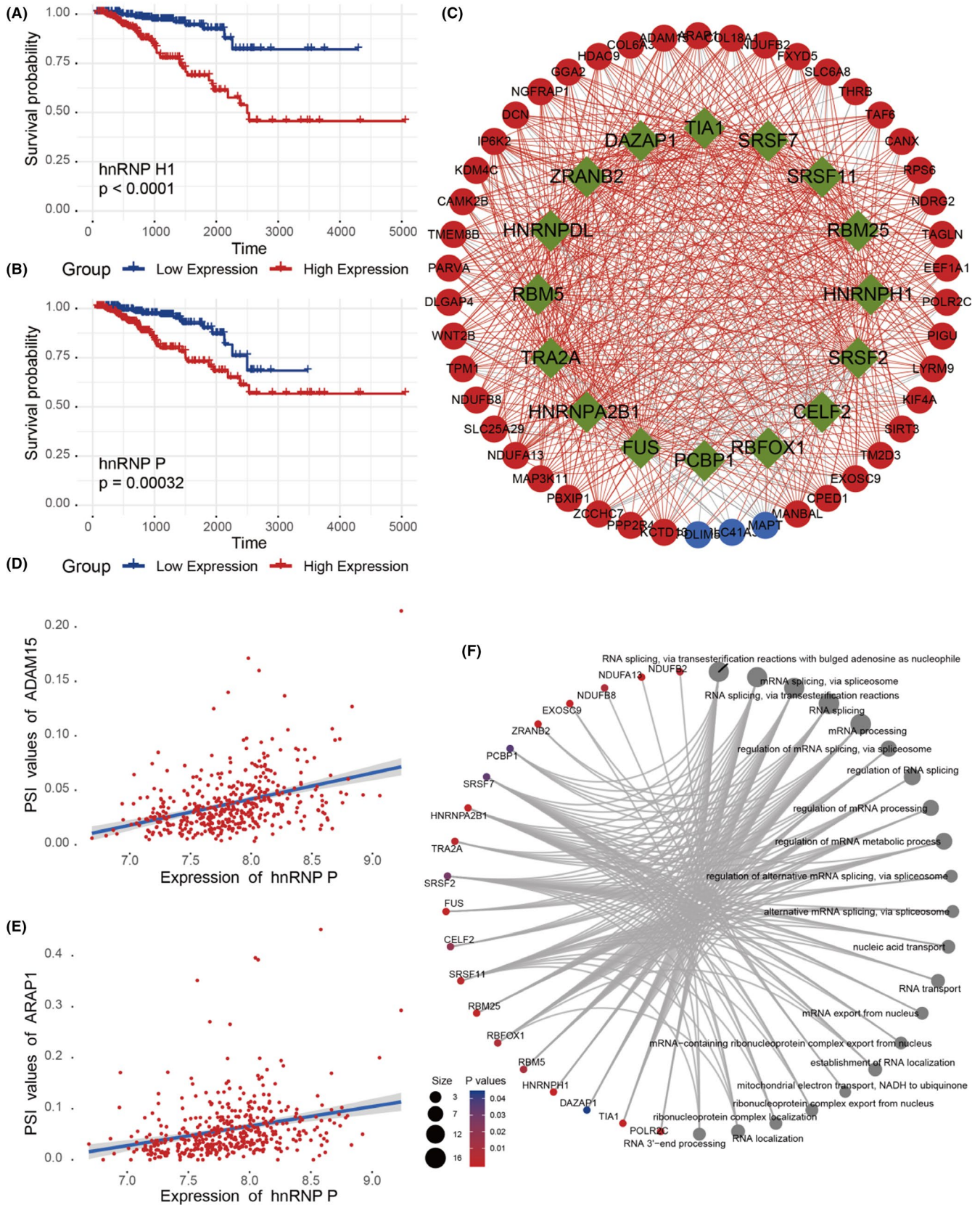


**FIGURE 4** A-F, Prognostic signatures based on all types of alternative splicing (AS) events in prostate cancer. G, Survival-death status and risk score distribution of each sample. H, ROC curves of prognostic predictors built by one type and all types of AS events in prostate cancer

### 3.4 | Prognostic factors for prostate cancer

To detect independent prognostic factors in patients with prostate cancer, we selected survival-related AS events as candidate factors, and multivariate Cox regression analysis was used to identify independent prognostic factors among the retained 6 AS types ( $P < .05$ ). One independent prognostic factor associated with AA, 4 independent prognostic factors associated with AD, 19 independent prognostic factors associated with AP, 11 independent prognostic factors associated with AT, 8 independent prognostic factors associated with ES, and 5 independent prognostic factors associated with RI were obtained. Subsequently, 6 different types of independent

prognostic AS events were combined to construct a final prognostic predictor. In our data analysis of each type of splicing pattern, the ability to predict the results of patients with prostate cancer using prognostic models constructed by different types of AS events is shown in Figure 4A-F. In particular, a prognostic model constructed by a single AP model showed the maximum prediction power among the 6 prognostic models (ROC = 0.815). In addition, a final prognostic predictor was constructed by combining 6 different types of candidate independent prognostic AS events (a total of 48 alternative splicing events). Notably, the final prognosis model showed good performance to predict power (ROC = 0.868), as shown in Figure 4G-I.



**FIGURE 5** A, B, Kaplan-Meier curves for splicing factors; C, Alternative splicing (AS) events network. The genes involved in downregulating splicing factors were represented with blue/green nodes. Red nodes were upregulating splicing factors. The positive/negative correlation between expressions of splicing factors and were represented with red/gray lines; D, E, Two cases for the relationship between PSI values of AS and the expression value of splicing factors; F, Gene network construction and functional enrichment analysis



### 3.5 | Interaction network between survival-related AS events and splicing factors

To determine which splicing factors were associated with survival-related AS events in prostate cancer, we performed a survival analysis of the splicing factor based on gene expression. The results showed that 16 splicing factors were significantly associated with overall survival, as seen in Figure 5A,B. Meanwhile, we also found that the vast majority (13/16) of survival-related splicing factors demonstrated that overexpression was associated with poor prognosis in patients. In addition, the correlation between the PSI values of prognosis-related AS events and the expression of survival-related splicing factors was investigated using Spearman test. Among them, 16 survival-related splicing factors (green dots) were significantly associated with 48 prognosis-related AS events (corresponding to 46 genes, 3 of which were downregulated and involved in alternative splicing events [blue dots] and 43 upregulated [red dots]) and 621 alternative splicing network interactions were constructed (the red line represents a positive correlation, the gray line represents a negative correlation, and the red line is dominant), as shown in Figure 5C. The correlation between the splicing factor hnRNP P and the AS event was reached 46 times, and genes such as *ADAM15*, *ARAP1*, *IP6K2*, *SLC25A29*, *SLC6A8*, *TAF6*, and *WNT2B* interacted with up to 16 splicing factors. The correlation between the splicing factor hnRNP P and *ADAM15* and *ARAP1* is shown in Figure 5D,E.

Subsequently, 62 genes involved in the AS network were profiled by GO functional enrichment analysis. Biological processes, cell components, and molecular functions ( $P < .05$ ) were significantly represented among AS events, biological processes are shown in Figure 5F.

### 3.6 | Molecular subtype cluster associated with prognosis

We further identified different AS patterns using unsupervised analysis of all samples based on AS events associated with prognosis. By combining the Elbow method to determine the optimal number of clusters, and based on the distribution of their consensus values ranging from 0 (white, samples never gathered together) to 1 (dark blue, samples always gathered together), we finally determined the 4 groups (C1, C2, C3, and C4) of samples as follows: C1 ( $n = 126$ , 29.1%), C2 ( $n = 102$ , 23.6%), C3 ( $n = 160$ , 36.9%), and C4 ( $n = 45$ , 10.4%), as shown in Figure 6A,B. Then we performed a Kaplan-Meier analysis to assess the relationship between clustering and prognosis. The results showed that clusters were associated with different survival patterns, with cluster 1 and cluster 2 correlating with poor outcomes in survival analysis, while cluster 3 and cluster 4 correlated with relatively good survival results, as shown in Figure 6C. At the same time, we further analyzed the relevant clinical information (such as survival time [DFS > 5 y or <5 y], survival status [alive or dead], T, N, M, age (age > 50 or < 50) and other information), finding that some related information, such as significant difference in

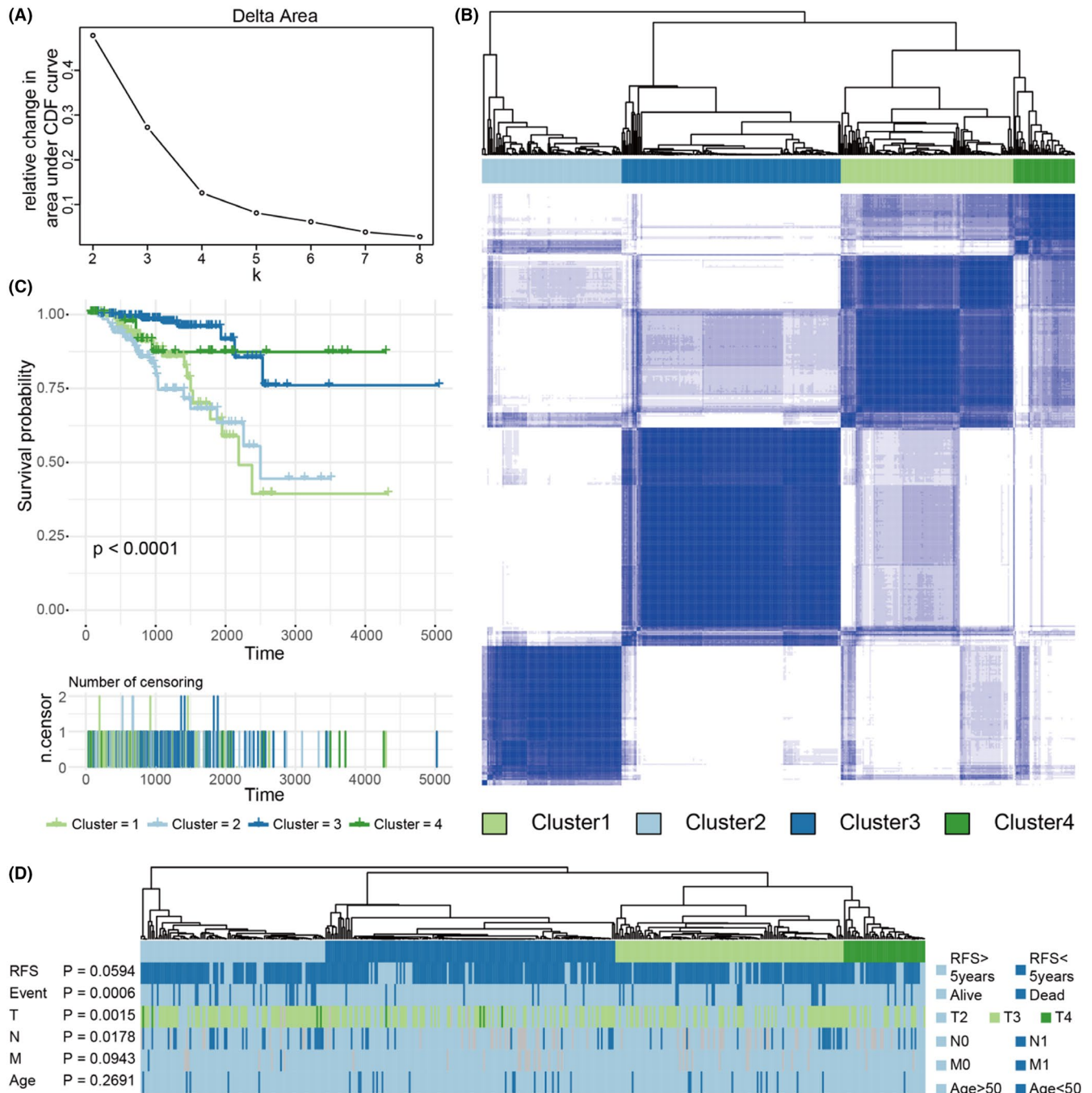
survival status (alive or dead), T, and N, in the 4 clusters was not randomly distributed (chi-square test,  $P < .05$ ), as shown in Figure 6D. Therefore, we could also identify molecular subtype clusters associated with prognosis through AS events.

## 4 | DISCUSSION

Most of the genes in the human genome are composed of multiple exons and introns, and the introns are cleaved to produce mature mRNA, which in turn encodes a protein.<sup>18</sup> The AS of RNA greatly enhances the diversity of the proteome, not only participating in normal physiological activities, but also playing a key role in the initiation or maintenance of human diseases, including prostate cancer.<sup>19-21</sup> Its development is complicated, especially in advanced prostate cancer. During the 14-30 mo after receiving hormonal therapy, most patients will be converted to CRPC. Many AS events are involved in this process. For example, splicing variants of AR are identified in metastatic tumors of CRPC patients,<sup>22-24</sup> the most widely studied is androgen receptor splice variant 7 (AR-V7), the expression of which is significantly associated with increased biochemical recurrence<sup>25</sup> and decreased survival in patients.<sup>26-28</sup> In addition, studies have shown the appearance of AS events in prostate cancer VEGF. The *VEGF* gene was thought to encode a single protein, but it is now very clear that the human *VEGF* gene generally contains 8 exons, producing at least 12 VEGF isomers by alternatively splicing within exons 6, 7 and 8,<sup>29</sup> resulting in proteins with different heparin-binding properties.<sup>30</sup> Furthermore, studies have shown that the ratio of different VEGF isoforms affects the formation of blood vessels in prostate cancer cells.<sup>31</sup> In addition, studies on abnormal AS in prostate cancer-related genes (such as *KLF6*, *BCL2L1*, *ERG*, etc) are also frequently reported.<sup>5,32,33</sup> Overall, the results confirmed that the AS expression of prostate cancer-associated genes and the balance of their splicing subtypes require wide and extensive research, because this approach could greatly expand the range of screening potential biomarkers and therapeutic targets.

In recent years, with the rapid development of high-throughput sequencing technologies and the continuous reduction in their cost, the diversity of AS maps for prostate cancer has achieved great success. Wang and colleagues<sup>34</sup> pointed out that differences in AA splicing variants of *PIK3CD*, *FGFR3*, *TSC2*, and *RASGRP2* showed stronger carcinogenicity. It was also pointed out that AS of the *PIK3CD-S* gene was closely related to poor prognosis of patients. Of course, the current TCGA database possesses a large amount of RNA-seq data from different cancer types. Through data mining, we can also systematically study RNA splicing disorders in cancer. By comparing different alternative splicing events between tumors and adjacent normal tissues, it has been established that survival-related AS events are associated with a variety of tumors, such as glioblastoma,<sup>35</sup> breast cancer,<sup>36</sup> lung cancer,<sup>37</sup> and colorectal cancer.<sup>38</sup>

Through the analysis of TCGA database, we identified AS events and regulatory splicing factors in prostate cancer to fully understand their differential RNA splicing pattern. In total, 1628 differentially



**FIGURE 6** A, Elbow and Gap statistic analysis for different numbers of clusters ( $k = 2-8$ ). B, Consensus matrix heatmap defined 4 clusters of samples for which consensus values range from 0 (in white, samples never clustered together) to 1 (dark blue, samples always clustered together). C, Kaplan-Meier survival analysis of patients within different clusters. D, Distribution of clinical information in 4 clusters

AS events were identified, of which 1338 were upregulated and 290 were downregulated, involving 1326 genes.

We then systematically studied the relationship between AS events and survival period of patients with prostate cancer. In total, 226 AS events were significantly associated with survival rate of patients with prostate cancer. Splicing events such as for *IP6K2*, *RPS6*, *DPM3*, *NDUFA13*, *TPM1*, and *WNT2B* play a key role in tumor biology. For example, tropomyosin is an important regulatory protein in muscle contraction. In nonmuscle cells, TPM1 acts

as a cell transformation inhibitor. Many of the splicing isoforms of TPM1, including the AT-splicing variant of TPM1 were found in our analysis to be involved in various types of tumors, such as head-and-neck cancer,<sup>39</sup> esophageal cancer,<sup>40</sup> breast cancer,<sup>41</sup> and lung cancer.<sup>42</sup> At the same time, for the first time, we found that high expression of the steroid-assisted regulatory gene *CALCOCO1* AS RI was significantly associated with poor prostate cancer prognosis. This result is similar to findings for breast cancer, which is also hormone related.<sup>43,44</sup>

In addition, we constructed a predictive model for each splicing type using multivariate Cox regression. The final predictive model for each splicing type performed quite well in distinguishing between patients with prostate cancer and their prognosis. The predictive models for each splicing type had different AUC values, and the prognostic model constructed by the AP model was the best among the 6 prognostic models, with the best predictive power (ROC = 0.815). We further combined 6 different types of candidate independent prognostic AS events to construct a final prognostic predictor with a predictive power (ROC = 0.868), which could provide patients with new predictive information and facilitate patient assessment.

Given that the regulation of key splicing factors may result in changes in widespread AS events, we further analyzed survival-related AS splicing factors and their associated networks and found that 16 splicing factors were significantly associated with overall patient survival. Interestingly, most of the splicing factor expression levels associated with poor survival were found for genes that were upregulated. Among them, the correlation between the splicing factor hnRNP P and the AS event was 46 times, which confirmed its key role in the development of prostate cancer, indicating that the results were consistent with the conclusions by Fei<sup>45</sup> and so on, and providing the possibility to explore the development and treatment of prostate cancer.

To the best of our knowledge, our study is the first to describe the use of unsupervised analysis to predict the prognosis of prostate cancer in different AS models, and furthermore to summarize different prognostic molecular subtypes. These subtypes were divided into 4 groups of molecular clusters, the first 2 subtypes of which were related to poor survival of prostate cancer, and another 2 subtypes of which indicated good survival. Furthermore, we further analyzed the relevant clinical information (such as survival time, survival status, TNM, Age, etc) and found that some related information was not randomly distributed, such as survival status (alive or dead) in 4 clusters. In addition, there were significant differences between T and N. This was similar to the phenomenon found in the study by Xiong and colleagues<sup>38</sup> in rectal cancer. Therefore, we could also identify molecular subtype clusters associated with prognosis through AS events. These findings enriched our knowledge of the molecular classification of prostate cancer and provided a large number of biomarker candidates and potential targets for the treatment of prostate cancer.

We describe here the most comprehensive and up-to-date study to assess prognostic factors and long-term survival outcomes for prostate cancer by analyzing abnormal AS variants, splicing factor regulatory networks, and related molecular cluster subtypes. Although the prognostic significance of these potential therapeutic targets for prostate cancer still requires further functional and clinical trials to validate the findings, this research enhances the potential of new clinical biomarkers and therapeutic approaches to help implement new treatment strategy for prostate cancer.

## ACKNOWLEDGMENTS

This work was supported by research grants from National Scientific Foundation of China (81771852). We are grateful to all patients who signed the consent form.

## DISCLOSURE

The authors have no conflicts of interest to declare.

## ORCID

Xi Wei  <https://orcid.org/0000-0003-4734-3900>

## REFERENCES

- Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet.* 2008;40:1413-1415.
- Venables JP. Aberrant and alternative splicing in cancer. *Can Res.* 2004;64:7647-7654.
- Kim E, Goren A, Ast G. Insights into the connection between cancer and alternative splicing. *Trends Genet.* 2008;24:7-10.
- Oltean S, Bates DO. Hallmarks of alternative splicing in cancer. *Oncogene.* 2014;33:5311-5318.
- David CJ, Manley JL. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev.* 2010;24:2343-2364.
- Sveen A, Kilpinen S, Ruusulehto A, Lothe RA, Skotheim RI. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene.* 2016;35:2413-2427.
- Salinas CA, Tsodikov A, Ishak-Howard M, Cooney KA. Prostate cancer in young men: an important clinical entity. *Nat Rev Urol.* 2014;11:317-323.
- Tabayoyong W, Abouassaly R. Prostate cancer screening and the associated controversy. *Surg Clin North Am.* 2015;95:1023-1039.
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin.* 2018;68:7-30.
- Etzioni R, Tsodikov A, Mariotto A, et al. Quantifying the role of PSA screening in the US prostate cancer mortality decline. *Cancer Causes Control.* 2008;19:175-181.
- Armenia J, Wankowicz SAM, Liu D, et al. The long tail of oncogenic drivers in prostate cancer. *Nat Genet.* 2018;50:645-651.
- Amin E, Oltean S, Hua J, et al. WT1 mutants reveal SRPK1 to be a downstream angiogenesis target by altering VEGF splicing. *Cancer Cell.* 2011;20:768-780.
- Wang Y, Bernhardt AJ, Cruz C, et al. The BRCA1-Δ11q alternative splice isoform bypasses germline mutations and promotes therapeutic resistance to PARP inhibition and cisplatin. *Can Res.* 2016;76:2778-2790.
- Li Y, Chan SC, Brand LJ, Hwang TH, Silverstein KAT, Dehm SM. Androgen receptor splice variants mediate enzalutamide resistance in castration-resistant prostate cancer cell lines. *Can Res.* 2013;73:483-489.
- Smith LD, Leme de Calais F, Raponi M, et al. Novel splice-switching oligonucleotide promotes BRCA1 aberrant splicing and susceptibility to PARP inhibitor action. *Int J Cancer.* 2017;140:1564-1570.
- Ryan MC, Cleland J, Kim R, Wong WC, Weinstein JN. SpliceSeq: a resource for analysis and visualization of RNA-Seq data on alternative splicing and its functional impacts. *Bioinformatics.* 2012;28:2385-2387.
- Giulietti M, Piva F, D'Antonio M, et al. SpliceAid-F: a database of human splicing factors and their RNA-binding sites. *Nucleic Acids Res.* 2013;41:D125-D131.
- Lee SC-W, Abdel-Wahab O. Therapeutic targeting of splicing in cancer. *Nat Med.* 2016;22:976-986.

19. Dehm SM, Schmidt LJ, Heemers HV, Vessella RL, Tindall DJ. Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. *Can Res*. 2008;68:5469-5477.
20. Busà R, Paronetto MP, Farini D, et al. The RNA-binding protein Sam68 contributes to proliferation and survival of human prostate cancer cells. *Oncogene*. 2007;26:4372-4382.
21. Carstens RP, Eaton JV, Krigman HR, Walther PJ, Garcia-Blanco MA. Alternative splicing of fibroblast growth factor receptor 2 (FGF-R2) in human prostate cancer. *Oncogene*. 1997;15:3059-3065.
22. Lu C, Luo J. Decoding the androgen receptor splice variants. *Trans Androl Urol*. 2013;2:178-186.
23. Antonarakis ES, Lu C, Wang H, et al. AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N Engl J Med*. 2014;371:1028-1038.
24. Efstathiou E, Titus M, Wen S, et al. Molecular characterization of enzalutamide-treated bone metastatic castration-resistant prostate cancer. *Eur Urol*. 2015;67:53-60.
25. Liu LL, Xie N, Sun S, Plymate S, Mostaghel E, Dong X. Mechanisms of the androgen receptor splicing in prostate cancer cells. *Oncogene*. 2014;33:3140-3150.
26. Hörnberg E, Ylitalo EB, Crnalic S, et al. Expression of androgen receptor splice variants in prostate cancer bone metastases is associated with castration-resistance and short survival. *PLoS One*. 2011;6:e19059.
27. Scher HI, Lu D, Schreiber NA, et al. Association of AR-V7 on circulating tumor cells as a treatment-specific biomarker with outcomes and survival in castration-resistant prostate cancer. *JAMA Oncol*. 2016;2:1441-1449.
28. Scher HI, Graf RP, Schreiber NA, et al. Nuclear-specific AR-V7 protein localization is necessary to guide treatment selection in metastatic castration-resistant prostate cancer. *Eur Urol*. 2017;71:874-882.
29. Perrin RM, Konopatskaya O, Qiu Y, Harper S, Bates DO, Churchill AJ. Diabetic retinopathy is associated with a switch in splicing from anti- to pro-angiogenic isoforms of vascular endothelial growth factor. *Diabetologia*. 2005;48:2422-2427.
30. Houck KA, Ferrara N, Winer J, Cachianes G, Li B, Leung DW. The vascular endothelial growth factor family: identification of a fourth molecular species and characterization of alternative splicing of RNA. *Mol Endocrinol*. 1991;5:1806-1814.
31. Catena R, Muniz-Medina V, Moralejo B, et al. Increased expression of VEGF121/VEGF165-189 ratio results in a significant enhancement of human prostate tumor angiogenesis. *Int J Cancer*. 2007;120:2096-2109.
32. Narla G, Difeo A, Reeves HL, et al. A germline DNA polymorphism enhances alternative splicing of the KLF6 tumor suppressor gene and is associated with increased prostate cancer risk. *Can Res*. 2005;65:1213-1222.
33. Hagen RM, Adamo P, Karamat S, et al. Quantitative analysis of ERG expression and its splice isoforms in formalin-fixed, paraffin-embedded prostate cancer samples: association with seminal vesicle invasion and biochemical recurrence. *Am J Clin Pathol*. 2014;142:533-540.
34. Wang B-D, Ceniccola K, Hwang S, et al. Alternative splicing promotes tumour aggressiveness and drug resistance in African American prostate cancer. *Nat Commun*. 2017;8:16161.
35. Pal S, Bi Y, Macyszyn L, Showe LC, O'Rourke DM, Davuluri RV. Isoform-level gene signature improves prognostic stratification and accurately classifies glioblastoma subtypes. *Nucleic Acids Res*. 2014;42:e64.
36. Suo C, Hrydziuszko O, Lee D, et al. Integration of somatic mutation, expression and functional data reveals potential driver genes predictive of breast cancer survival. *Bioinformatics*. 2015;31:2607-2613.
37. 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.
38. 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;36:183-195.
39. Liang Y, Song J, He D, et al. Systematic analysis of survival-associated alternative splicing signatures uncovers prognostic predictors for head and neck cancer. *J Cell Physiol*. 2019;234:15836-15846.
40. Huang G-W, Zhang Y-L, Liao L-D, Li E-M, Xu L-Y. Natural antisense transcript TPM1-AS regulates the alternative splicing of tropomyosin I through an interaction with RNA-binding motif protein 4. *Int J Biochem Cell Biol*. 2017;90:59-67.
41. Dube S, Thomas A, Abbott L, et al. Expression of tropomyosin 2 gene isoforms in human breast cancer cell lines. *Oncol Rep*. 2016;35:3143-3150.
42. Langer W, Sohler F, Leder G, et al. Exon array analysis using re-defined probe sets results in reliable identification of alternatively spliced genes in non-small cell lung cancer. *BMC Genom*. 2010;11:676.
43. Haiman CA, Garcia RR, Hsu C, et al. Screening and association testing of common coding variation in steroid hormone receptor co-activator and co-repressor genes in relation to breast cancer risk: the Multiethnic Cohort. *BMC Cancer*. 2009;9:43.
44. Wu D-Y, Ou C-Y, Chodankar R, Siegmund KD, Stallcup MR. Distinct, genome-wide, gene-specific selectivity patterns of four glucocorticoid receptor coregulators. *Nucl Recept Signal*. 2014;12:e002.
45. Fei T, Chen Y, Xiao T, et al. Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing. *Proc Natl Acad Sci USA*. 2017;114:E5207-E5215.

**How to cite this article:** Zhao J, Chang L, Gu X, Liu J, Sun B, Wei X. Systematic profiling of alternative splicing signature reveals prognostic predictor for prostate cancer. *Cancer Sci*. 2020;111:3020-3031. <https://doi.org/10.1111/cas.14525>