



Research Article

Whole Genome Analysis and Prognostic Model Construction Based on Alternative Splicing Events in Endometrial Cancer

Caixia Wang ^{1,2}, Mingjun Zheng,^{1,2} Shuang Wang,^{1,2} Xin Nie,^{1,2} Qian Guo,^{1,2} Lingling Gao,^{1,2} Xiao Li,^{1,2} Yue Qi,^{1,2} Juanjuan Liu,^{1,2} and Bei Lin ^{1,2}

¹Department of Obstetrics and Gynaecology, Shengjing Hospital Affiliated to China Medical University, Liaoning, China

²Key Laboratory of Maternal-Fetal Medicine of Liaoning Province, Key Laboratory of Obstetrics and Gynecology of Higher Education of Liaoning Province, Liaoning, China

Correspondence should be addressed to Bei Lin; linbei88@hotmail.com

Received 9 April 2019; Accepted 4 June 2019; Published 2 July 2019

Academic Editor: Marcelo A. Soares

Copyright © 2019 Caixia Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objectives. A growing body of evidence has shown that aberrant alternative splicing (AS) is closely related to the occurrence and development of cancer. However, prior studies mainly have concentrated on a few genes that exhibit aberrant AS. This study aimed to determine AS events through whole genome analysis and construct a prognostic model of endometrial cancer (EC). **Methods.** We downloaded gene expression RNAseq data from UCSC Xena, and seven types of AS events from TCGA SpliceSeq. Univariate Cox regression was employed to analyze the prognostic-related alternative splicing events (PASEs) and splicing factors; multivariate Cox regression was conducted to analyze the effect of risk score (All) and clinicopathological parameters on EC prognosis. An underlying interaction network of PASEs of EC was constructed by Cytoscape Reactome FI, GO, and KEGG pathway enrichment was performed by DAVID. ROC curves and Kaplan-Meier analysis were used to assess the diagnostic value of prognostic model. The correlation between PASEs and splicing factors was analyzed by GraphPad Prism; then a network was constructed using Cytoscape. **Results.** In total, 28,281 AS events in EC were identified, which consisted of 1166 PASEs. RNPS1, NEK2, and CTNNB1 were the hub genes in the network of the top 600 PASEs. The area under the curve (AUC) of risk score (All) reached 0.819. Risk score (All) together with FIGO stage, cancer status, and primary therapy outcome success was risk factors of the prognosis of EC patients. Splicing factors YBX1, HNRNPDL, and HNRNPA1 were significantly related to the overall survival (OS). The splicing network indicated that the expression of splicing factors was significantly correlated with percent-splice-in (PSI) value of PASEs. **Conclusion.** We constructed a model for predicting the prognosis of EC patients based on PASEs using whole genome analysis of AS events and thereby provided a reliable theoretical basis for EC clinical prognosis evaluation.

1. Introduction

Alternative splicing (AS) is an important regulation mechanism in the process of mRNAs after transcription. Pre-mRNA produces different mRNAs through different splicing methods to translate into different proteins with versatile functions, which contributes to protein diversity [1]. AS is a ubiquitous biological process, approximately 95% of human genes undergo AS in physiological processes [2]. In recent years, studies have shown that AS plays an important role in the occurrence and development of cancer, and AS participates in the process of proliferation, apoptosis, and metastasis of tumor cells [3, 4]. AS is mainly regulated by

the spliceosome, which is the complex proteins composed of small nuclear ribonucleic acids (snRNA) and splicing factors [5]. A splicing factor is an important accessory protein that regulates pre-mRNA AS. Mutation and/or abnormal expression of splicing factors are closely related to the occurrence of abnormal AS [6]. In total, there are seven types of AS events listed in the SpliceSeq database, i.e., alternate acceptor site (AA), alternate donor site (AD), alternate terminator (AT), alternate promoter (AP), exon skip (ES), mutually exclusive exons (ME), and retained intron (RI) [7].

Endometrial cancer (EC) is one of the common gynecological malignant tumors. In recent years, older age (≥ 55 years old), obesity (≥ 25 kg/m²), diabetes, and tamoxifen

adjuvant therapy have led to an increase in the incidence of endometrial cancer (EC) [8]. According to cancer statistics in 2019, an estimated 61,880 new cases of EC will occur in the United States, proceeding to the first in gynecological malignant tumors [9]. EC is a hormone-dependent cancer: research has shown that AS of estrogen receptor α (ER α) and progesterone receptor (PR) is closely related to the occurrence and development of EC [10]. A newly defined splicing factor, YT521, was found to promote the AS of vascular endothelial growth factor A (VEGF-A), thereby leading to an increase in splicing variant VEGF-165 and ultimately promoting EC invasion [11]. Until now, monitoring the diagnosis and prognosis of EC has been limited by a lack of biomarkers with both sensitivity and specificity. Therefore, the purpose of our study was to elucidate aberrant AS in EC, which would provide a theoretical basis for finding prognostic biomarkers and therapeutic targets for EC. We also analyzed PASEs through whole genome analysis and constructed prognostic model for EC.

In the present study, we obtained normalized RNA-seq data from UCSC Xena and seven types of AS events from TCGA SpliceSeq database, then performed a systematic analysis to figure out PASEs in EC, then we constructed prognostic model risk score (All) based on PASEs, and evaluated the diagnostic value of risk score (All) in assessing the prognosis of EC. In addition, splicing factors that were associated with prognosis of EC were identified using univariate Cox regression and a correlation network between splicing factors and PASEs was constructed. In summary, our study developed a new prognostic model for EC.

2. Materials and Methods

2.1. Data Extraction and Preprocessing. RNAseq-HTSeq-FPKM-UQ data was downloaded from UCSC Xena (<https://xenabrowser.net/>), which provided normalized RNAseq data. Except for 35 cases of adjacent normal samples, seven types of AS events of 545 ECs were downloaded from the TCGA SpliceSeq (<https://bioinformatics.mdanderson.org/TCGASpliceSeq/index.jsp>) database, and PSI value ranges from 0 to 1, which is used to quantify AS events. AS events with PSI range ≥ 0.5 were included in this study.

2.2. Building a Prognostic Model Based on PASEs. Patients with overall survival (OS) ≥ 30 days were included in this study. According to its median number, each parameter was divided into high-risk (\geq median number) and low-risk ($<$ median number) groups. Univariate Cox regression was performed on all seven types of AS events, and lnHR was used as the β -coefficient to construct a prognostic model for EC [12]. The prognostic model was constructed according to

$$\text{Risk score} = \sum_i^n \text{PSI}_i * \beta_i \quad (1)$$

2.3. Evaluation of the Diagnostic Value of Prognostic Models. A Kaplan-Meier survival curve was plotted using the “survival” package (version 2.43-3) to assess the effect of the

prognostic model on the 5-year OS of EC. A ROC curve was drawn using the “survivalROC” package (version 1.0.3) to assess the diagnostic value of prognostic models of EC

2.4. Functional and Pathway Enrichment Analysis. DAVID (<https://david.ncicrf.gov/>) is a database that integrates biological data and analysis tools and provides comprehensive annotation information. PASEs were analyzed using DAVID (version 6.8), including GO terms (Biological Process/Cellular Component/Molecular Function, BP/CC/MF) and KEGG pathways.

2.5. Construction of an Interaction and Correlation Network. To further explore the interaction of PASEs in EC, we constructed an interaction network using Cytoscape Reactome FI (version 3.6.0), and hub genes were selected according to the number of degrees. The splicing factors were downloaded from the TCGA SpliceAid2 (http://193.206.120.249/splicing_tissue.html) database (Table S1), and a scatter plot between prognostic-related splicing factor expression profiles and the PSI values of the selected PASEs was constructed in GraphPad Prism8.0. The correlation network between splicing factors and PASEs was constructed in Cytoscape (version 3.7.1).

2.6. Statistical Analysis. UpSetR (version 1.3.3) was used to quantify seven types of PASEs in EC. Statistical analyses were performed using R/Bioconductor (version 3.4.3) and GraphPad Prism (version 8.0). Two-tailed $P < 0.05$ was considered statistically significant. * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.

3. Results

3.1. Overview of AS Events in TCGA-EC. In total, there were 8140 genes, and 28281 AS events in 545 EC patients and the median number of AS events per gene was 3.474. Among the seven types of AS events, ES was the most common, followed by AT and AP. Specifically, there were 4604 genes in the 9744 ES events, 3411 genes in the 7796 ATs event, 1792 genes in the 4458 AP events, 1691 genes in the 2270 AA events, 1413 genes in the 2050 RI events, 1386 genes in the 1877 AD events, and 85 genes in the 86 ME events (Figure 1(a)).

It is worth noting that one gene can undergo multiple types AS events; therefore, UpSet picture was used to match the genes with AS events. The results showed that the most common event was ES followed by AT and AP; ME events occurred the least. Three (AA, AD, and ES) AS events observed were DMKN and ATXN2L (Figure 1(b)).

To better track the AS events, we named AS events using the gene name, AS type, and a unique ID number in AS event. For example, in RPLP0-ES-24731, RPLP0 represents the gene name, ES represents the type of AS event, and 24731 was the unique ID number of the AS event.

3.2. AS Events Associated with Prognosis of EC. To build a prognostic model, we needed to find AS events with good discrimination; thus, AS events with PSI range ≥ 0.5 were

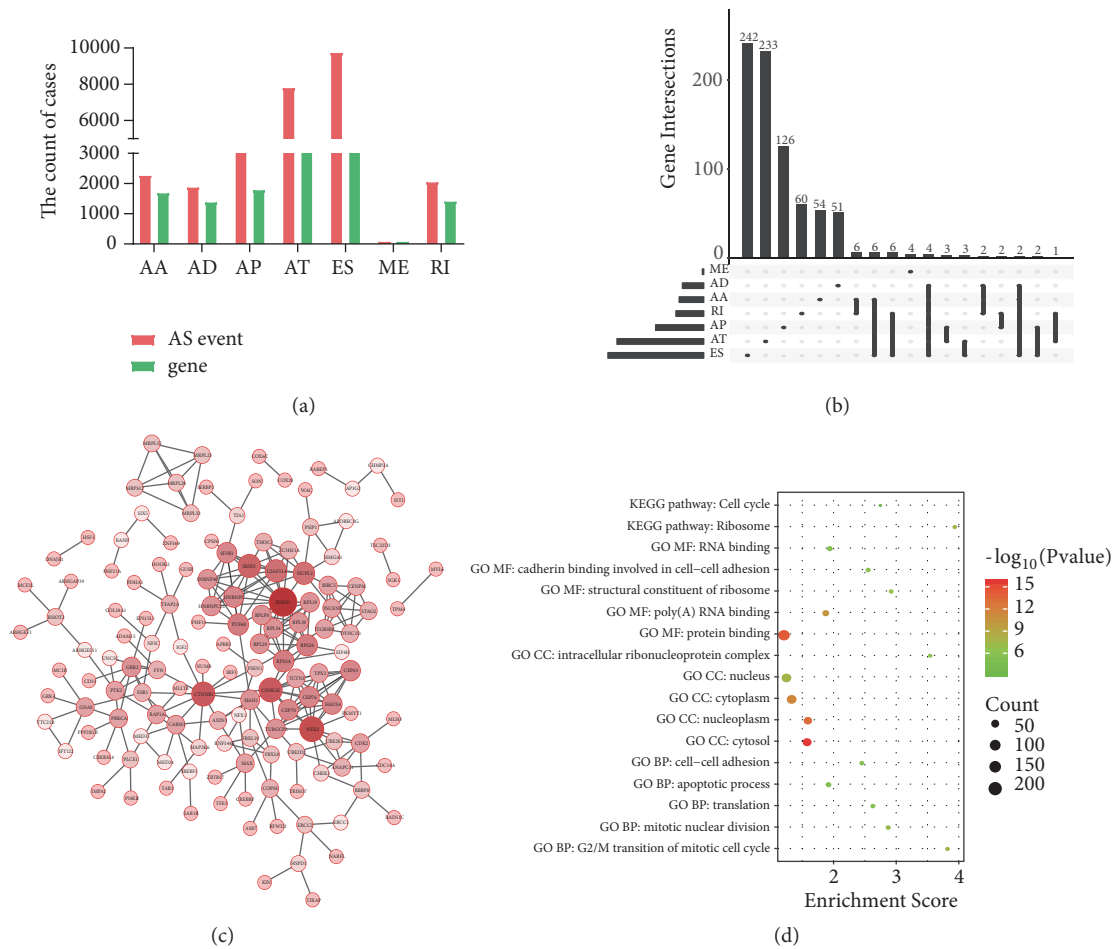


FIGURE 1: Upset and bioinformatic analysis of PASEs in EC. (a) AS events and genes in seven types of alternative splicing. (b) Upset plot of PASEs in EC. (c) Interaction network of PASEs constructed by cytoscape Reactome FI. (d) The bubble plot of Top5 GO (BP/CC/MF) and KEGG pathway.

included in this study. Ultimately, we obtained 5015 genes and 11709 AS events.

To explore the relationship between AS events and EC prognosis, we evaluated prognostic AS events using univariate Cox regression. The results showed that 1166 AS events were significantly associated with OS in EC ($P < 0.05$; Table S2). We selected top 600 PASEs and constructed an interaction network using Cytoscape Reactome FI. The results showed that the RNPS1, NEK2, and CTNNB1 genes were the hub nodes of this network (Figure 1(c)). Then we performed GO (BP/CC/MF) and KEGG pathway analysis using the DAVID database and found that 19 BPs were enriched, such as the G2/M transition of mitotic cell cycle, mitotic nuclear division, and translation, 12 CCs, 12 MFs ($P < 0.05$), and 2 KEGG signaling pathways ($P < 0.05$; Table S3), such as ribosome and cell cycle, and plotted the bubbles of Top5 GO (BP/CC/MF) and the KEGG pathways (Figure 1(d)).

3.3. Construction of Prognostic Model of EC. To investigate the relationship between AS events and EC survival outcomes, we constructed a prognostic model of EC based on

PASEs. The P values were sorted from small to large, and the top 4 genes for each AS event were selected to construct a prognostic model, while the top 6 genes were used to construct the risk score (All) prognostic model. We chose lnHR to construct a formula of prognostic models (Table 1).

$$\text{Risk score (All)} = (\text{PSI value of RPLP0-ES-24731} \times -9.093) + (\text{PSI value of ZNF586-AT-52338} \times 2.913) + (\text{PSI value of STK32C-AP-13486} \times -2.391) + (\text{PSI value of C4orf29-AT-70557} \times 4.133) + (\text{PSI value of C4orf29-AT-70558} \times -4.133) + (\text{PSI value of ANAPC11-ES-44217} \times 3.945).$$

According to the median number of risk score, EC patients were divided into high and low-risk groups, which indicated that the mortality rate was higher for EC patients in the high-risk group (red spots), which was associated with the lower OS (red and green spots; Figures 2(a)–2(h)). A ROC curve was used to evaluate the diagnostic value of the prognostic model; results suggest that the AUC of the risk score (All) is 0.819 followed by 0.805 for the risk score (ES) and 0.731 for the risk score (AT) (Figures 3(a)–3(h)). Compared with the prognostic model constructed by a single AS event, risk score (All) better predicted EC patients'

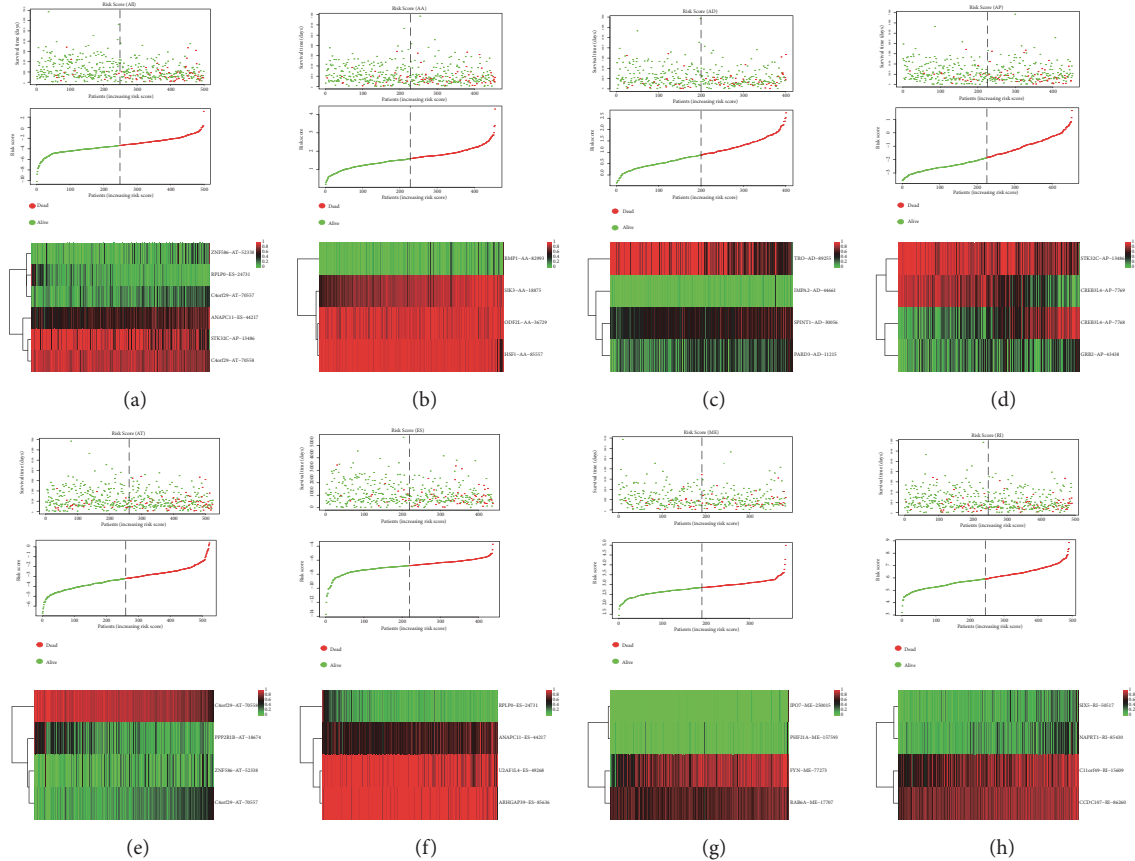


FIGURE 2: Construction of eight PASEs models for EC. (a–h) Risk scores for All, AA, AD, AP, AT, ES, ME, RI models in EC, respectively. Each individual plot (Top) represents the distribution of survival time and survival status of high- and low-risk groups. (middle) represents the distribution of patients in the high- and low-risk groups, (bottom) represents the PSI value heat map of the alternative splicing genes in the constructed model.

prognosis. We also plotted survival curves for all prognostic models (Figures 3(i)–3(p)) and found that the high-risk groups of all prognostic models were associated with poor prognosis in EC patients ($P < 0.05$).

In addition, we analyzed the relationship between clinicopathological parameters and EC prognosis. Age more than 55 years old and BMI more than 25 kg/m² were considered as high-risk factors. Univariate Cox regression analysis indicated that pathological type, FIGO stage, grade, cancer status, new tumor event after initial treatment, primary therapy outcome success, and risk score (All) were associated with poor prognosis of EC patients ($P < 0.05$). Multivariate Cox regression showed that only FIGO stage, cancer status, primary therapy outcome success, and risk score (All) were risk factors for the poor prognosis of EC ($P < 0.05$) (Table 2).

3.4. Splicing Factors Associated with EC Prognosis. Splicing factors are the executors of the AS events, and mutation or abnormal expression of splicing factors are related to the occurrence and development of cancer. To identify prognosis-associated splicing factors in EC, we performed univariate Cox regression analysis and obtained 10 splicing factors (FMR1, SRSF4, PCBP2, PTBP2, HNRNPDL,

YBX1, QKI, RBM5, HNRNPA1, and HNRNPF; $P < 0.05$), among which splicing factors SRSF4, HNRNPDL, RBM5, HNRNPA1, and HNRNPF were considered protective factors (Table S4). A correlation network constructed by Cytoscape indicated that 10 splicing factors (blue dots) were negatively (green lines) correlated with 125 favorable prognostic AS events (green dots), and positively correlated (red lines) with 143 AS events with poor prognosis (red dots; Figure 4(g); Table S5).

We constructed Kaplan-Meier survival curves of 10 splicing factors and found that only YBX1, HNRNPDL, and HNRNPA1 had significant effects on OS in EC patients (Figure S1). We plotted representative K-M survival curves of YBX1 and HNRNPDL ($P < 0.05$; Figures 4(a) and 4(b)). In addition, we generated a scatter plot between the expression of splicing factors and PSI value of AS events and found that splicing factor YBX1 was positively correlated with PSI value of DNAH9-AT-39292 and negatively correlated with PSI value of DNAH9-AT-39293 (Figures 4(c) and 4(e)). Splicing factor HNRNPDL showed different correlation between different splicing event types of the same gene CD33 ($P < 0.005$; Figures 4(d) and 4(f)). The above findings suggested that different splicing factors played different roles in different AS events.

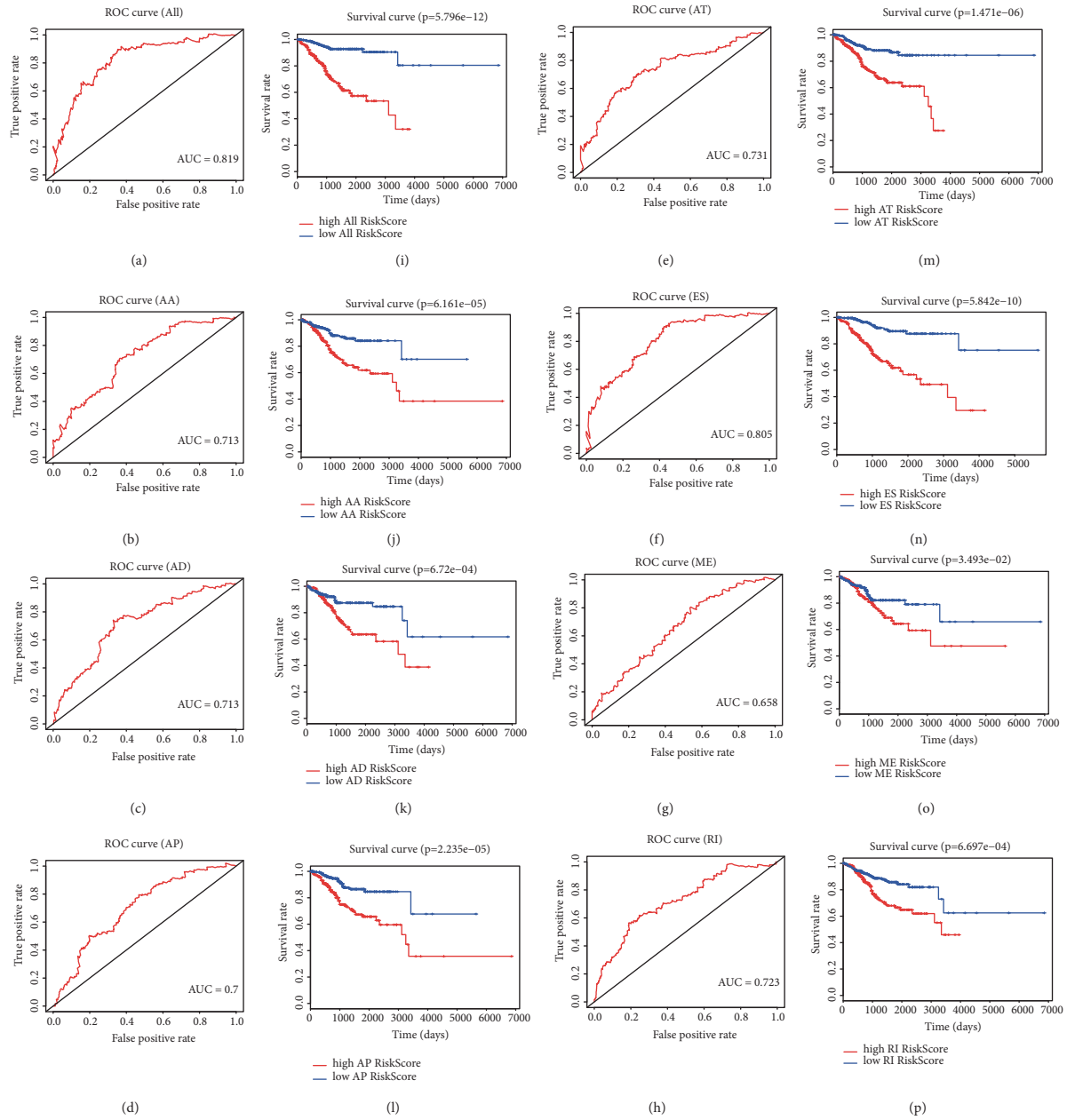


FIGURE 3: ROC curves and K-M survival curves for eight PASEs models in EC. (a-h) ROC curves for All, AA, AD, AP, AT, ES, ME, RI models. (i-p) K-M survival curves of All, AA, AD, AP, AT, ES, ME, RI models.

4. Discussion

AS is an important biological process for producing protein diversity. Abnormal AS events in tumors are closely related to tumor initiation and tumor progression. A gene can undergo different types of AS events and can be regulated by a variety of splicing factors, thus complicating the study of the regulatory networks between AS events and splicing factors. At present, research mainly detected AS of EC by using small sample sequencing [13]. However, studies employing whole genome analysis of AS events in EC have not been reported. In recent years, high-throughput sequencing technology can

better characterize abnormally mutated splices and splice sites; therefore, it is important to use bioinformatics techniques to analyze abnormal AS events in EC.

In the present study, we downloaded seven types of AS events from the TCGA SpliceSeq database. There were 8140 genes and 28281 AS events in EC, which indicates that AS events are ubiquitous in EC. 1166 PASEs were identified using univariate Cox regression analysis, and the underlying regulatory network was constructed using Cytoscape based on the top 600 PASEs. The results indicated that the RNPS1, NEK2, and CTNBN1 genes were hub nodes of the network. RNPS1 is a member of the SR protein family and functions

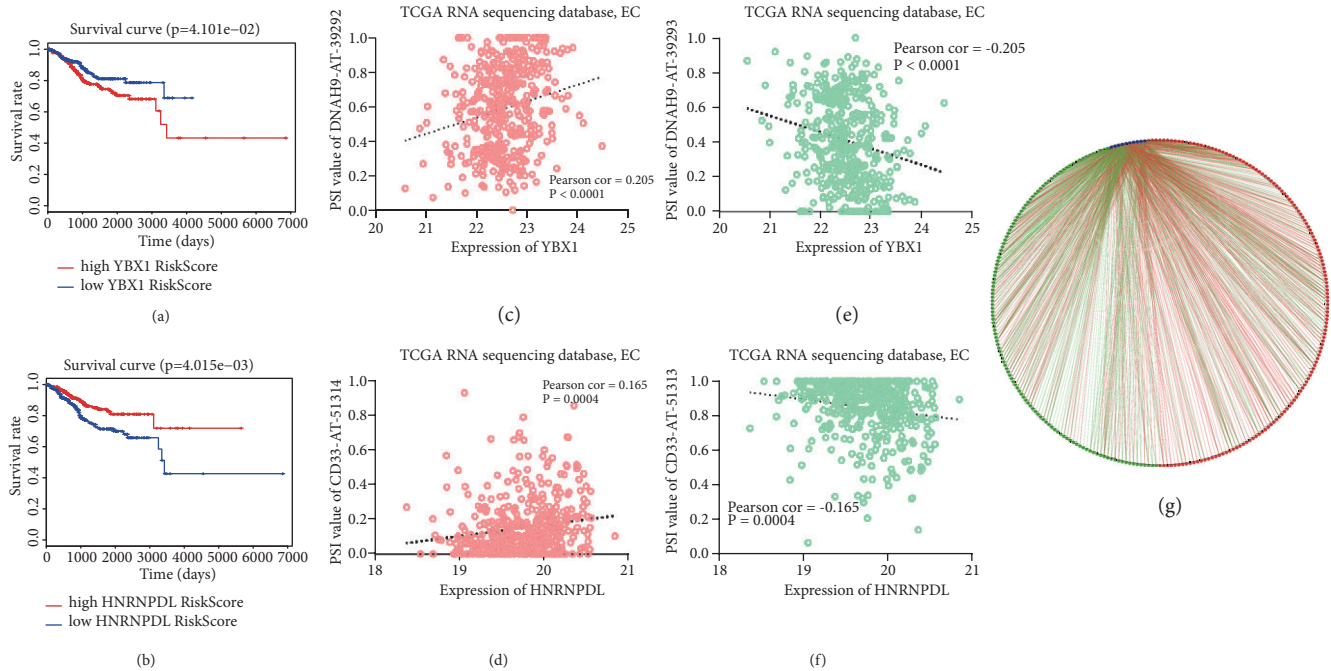


FIGURE 4: Correlation between prognosis-related splicing factors and PASEs in EC. (a, b) K-M survival curves of YBX1 and HNRNPDL, divided into high (red line), low (green line) group according to median expression value of splicing factor. (c, e) Correlation analysis between expression of YBX1 and PSI values of DNAH9-AT-39292, DNAH9-AT-39293. (d, f) Correlation analysis between expression of HNRNPDL and PSI values of CD33-AT-51314, CD33-AT-51313. (g) Correlation network between prognosis-related splicing factors and PSI values of alternative splicing events, blue dots represent splicing factors, green/red dots represent favorable/poor prognostic alternative splicing events, respectively, and red/green lines represent positive/negative correlations, respectively.

as an activator in the AS process [14, 15]. RNPS1 can inhibit abnormal splicing of pre-mRNA and plays an important role in quality control during pre-mRNA splicing [15, 16]. Nek2 is a serine/threonine protein kinase that is involved in the regulation of the cell cycle, and there exist three alternative splice variants, Nek2A, Nek2B, and Nek2C [17]. A prior study showed that Nek2C was highly expressed in breast cancer cells and promoted breast cancer cells invasion and migration [18]. In addition, we performed GO (BP/CC/MF) and KEGG pathway enrichment using DAVID and found that these genes were closely related to transition of mitotic cell cycle, mitotic nuclear division, ribosome, cell cycle, etc. From the above perspective, RNPS1, NEK2, and CTNBN1 were the hub genes in the identified network and might therefore be novel targets for EC treatment.

Recently, a prognostic model was constructed based on four miRNAs (miR-4758, miR-876, miR-142, and miR-190b) to evaluate the prognosis of EC patients [19]. However, the construction of a prognostic model for EC based on AS events has not been reported. To assess the diagnostic value of aberrant AS events in the prognosis of EC, we constructed prognostic models based on risk score (All) and seven types of AS event types (AA, AD, AP, AT, ES, ME, and RI). EC patients were divided into high-risk and low-risk groups according to the median number of each AS event, then we plotted Kaplan-Meier survival curves of risk score (All) and the risk scores of each of the seven types of AS events. The results showed that risk score (All) and seven types of AS

events were associated with poor prognosis in EC patients ($P < 0.05$). In addition, we also plotted ROC curves. The results show that the AUC of the risk score (All) is 0.819 followed by 0.805 for the risk score (ES) and 0.731 for the risk score (AT). Univariate Cox regression was used to analyze the effects of clinicopathological parameters and risk score (All) on the prognosis of EC patients, and multivariate Cox regression was performed to analyze the risk factors that were associated with prognosis. The results show that FIGO stage, cancer status, primary therapy outcome success, and risk score (All) were risk factors for EC patients. These results suggested that the risk score (All) model played an important role in the assessment of EC prognosis.

Abnormalities in splicing factor expression are also related to aberrant AS events. In this study, we obtained 10 splicing factors that are related to the OS of EC patients using univariate Cox regression. By plotting Kaplan-Meier survival curves, we found that only splicing factors YBX1, HNRNPDL, and HNRNPA1 were significantly associated with EC prognosis ($P < 0.05$). The constructed splicing network analyzes the association between splicing factors and PASEs in EC. The results suggested that poor prognostic AS events were positively correlated with the expression of splicing factors, while favorable prognostic AS events was negatively correlated with the expression of splicing factors. A splicing factor can precisely regulate the pre-mRNA splicing process by binding to the splicing regulatory sequence elements of a particular gene [20]. Y-box binding protein

TABLE I: Construction of prognostic model of EC.

Type	ID	β	HR	Lower	Upper	P value
All	RPLP0-ES-24731	-9.093	1.12E-04	1.15E-06	0.011	1.00E-04 ***
	ZNF586-AT-52338	2.913	18.42	4.238	80.06	1.02E-04 ***
	STK32C-AP-13486	-2.391	0.092	0.027	0.31	1.21E-04 ***
	C4orf29-AT-70557	4.133	62.39	7.408	525.417	1.44E-04 ***
	C4orf29-AT-70558	-4.133	0.016	1.90E-03	0.135	1.44E-04 ***
	ANAPC11-ES-44217	3.945	51.663	6.701	398.295	1.53E-04 ***
AA	ODF2L-AA-3672	2.0766	7.977	2.699	23.575	1.73E-04 ***
	HSF1-AA-85557	-3.667	0.026	3.61E-03	0.181	2.39E-04 ***
	BMP1-AA-82993	3.809	45.089	5.254	386.93	5.15E-04 ***
	SIK3-AA-18875	3.627	37.604	4.795	294.899	5.57E-04 ***
AD	IMPA2-AD-44661	4.107	60.745	4.778	772.247	1.55E-03 **
	SPINT1-AD-30056	2.654	14.215	2.724	74.17	1.64E-03 **
	TRO-AD-89255	-1.461	0.232	0.093	0.579	1.73E-03 **
	PARD3-AD-11215	2.814	16.673	2.828	98.309	1.88E-03 **
AP	STK32C-AP-13486	-2.391	0.092	0.027	0.31	1.21E-04 ***
	CREB3L4-AP-7768	1.379	3.972	1.942	8.125	1.58E-04 ***
	CREB3L4-AP-7769	-1.379	0.252	0.123	0.515	1.58E-04 ***
	GRB2-AP-43438	2.56	12.929	3.392	49.286	1.78E-04 ***
AT	ZNF586-AT-52338	2.913	18.42	4.238	80.06	1.02E-04 ***
	C4orf29-AT-70557	4.133	62.39	7.408	525.417	1.44E-04 ***
	C4orf29-AT-70558	-4.133	0.016	1.90E-03	0.135	1.44E-04 ***
	PPP2R1B-AT-18674	-4.078	0.017	2.01E-03	0.143	1.76E-04 ***
ES	RPLP0-ES-24731	-9.093	1.12E-04	1.15E-06	0.011	1.00E-04 ***
	ANAPC11-ES-44217	3.945	51.663	6.701	398.295	1.53E-04 ***
	U2AF1L4-ES-49268	-3.799	0.022	3.08E-03	0.163	1.77E-04 ***
	ARHGAP39-ES-85636	-4.678	9.29E-03	7.91E-04	0.109	1.98E-04 ***
ME	IPO7-ME-250015	3.358	28.742	4.152	198.957	6.69E-04 ***
	FYN-ME-77273	1.682	5.377	1.445	20.006	0.012 *
	PHF21A-ME-157593	3.419	30.549	1.735	537.811	0.019 *
	RAB6A-ME-17707	2.375	10.75	1.085	106.502	0.042 *
RI	SIX5-RI-50517	3.467	32.036	4.378	234.442	6.40E-04 ***
	C11orf49-RI-15609	2.418	11.219	2.8	44.957	6.41E-04 ***
	CCDC107-RI-86260	4.666	106.224	6.956	1622.054	7.95E-04 ***
	NAPRT1-RI-85430	2.911	18.371	2.992	112.814	1.67E-03 **

(YBX1) is a DNA/RNA-binding protein that is involved in gene transcriptional regulation and pre-mRNA splicing [21–23]. Allemand E [24] found that YBX1 interacted with the active form of PP2Cgamma, which is involved in spliceosome assembly, thus regulating CD44 splicing. In breast cancer, it has been shown that there is a significant difference in the expression of YBX1 between triple-negative (TN) and ER+ subtypes [25]. Additionally, another study found that high expression of YBX1 in ER+ breast cancer patients was associated with poor prognosis [26]. It is well-known that EC is a hormone-dependent tumor. EC is mainly divided into estrogen-dependent (type I) and non-estrogen-dependent (type II), and studies have shown that an ER- and PR-type is associated with poor prognosis of EC patients [27]. Therefore, we hypothesized that the differential expression of splicing factor YBX1 was closely related to the different types of endometrial cancer. Interestingly, our study showed

that YBX1 was both positively and negatively correlated with different splicing patterns of axonemal dynein heavy chain (DNAH9), which indicated that YBX1 might regulate the splicing of DNAH9. In summary, YBX1 might be diagnostic biomarker and therapeutic target for EC.

HNRNPDL and HNRNPA1 are members of the hnRNP family of RNA-binding proteins that are involved in RNA maturation and translation and in pre-mRNA AS [28, 29]. Studies have shown that the inhibition of HNRNPDL expression can result in increased expression of genes that are involved in cell proliferation and migration [28]. Consistent with this research, our results indicated that HNRNPDL served as a potent tumor suppressor gene. However, studies have shown that HNRNPA1 is highly expressed in gastric cancer cells, thus promoting gastric cancer proliferation, invasion, migration, and EMT [1, 30]. HNRNPA1 regulated AS of the androgen receptor (AR) and gave rise to a

TABLE 2: Univariate and multivariate Cox regression analysis of clinicopathological parameters and risk score (All) in 523 EC patients.

Variables	Group	Patient N=523	Univariate analysis		Multivariate analysis	
			HR (95% CI)	P value	HR (95% CI)	P value
Age			1.607(0.83-3.112)	0.16		
	<55	521				
	≥55	94				
BMI						
	<25	427				
	≥25	492	0.995 (0.576-1.718)	0.986		
Subtype						
	Endometrioid	91				
	Non-Endometrioid	401				
		523	2.993(1.96-4.57)	3.85E-07 ***	1.332(0.626-2.835)	0.456
Stage						
	I-II	393				
	III-IV	130				
		523	4.123(2.691-6.316)	7.55E-11 ***	2.832(1.362-5.888)	0.005**
Grade						
	Low grade (G1)	378				
	High grade (G2/G3)	145				
		523	11.692(2.876-47.53)	5.90E-04 ***	1.32 (0.533-3.274)	0.549
Cancer status						
	Tumor free	96				
	With tumor	427				
		495	9.018(5.817-13.98)	<2E-16 ***	2.282(1.094-4.763)	0.028 *
New tumor event after initial treatment						
	NO	417				
	YES	78				
		467	5.735 (3.54-9.292)	1.31E-12 ***	1.738(0.9-3.359)	0.1
Primary therapy outcome success						
	CRR+SD+PRR	387				
	PD	80				
		417	9.312(4.936-17.56)	5.54E-12 ***	4.15 (1.741- 9.892)	1.33E-03 **
Radiation						
	NO	401				
	YES	16				
		505	0.64(0.404-1.016)	0.0586		
Risk score (All)						
	Low	278				
	high	227				
		499	5.823 (3.308-10.25)	1.02E-09 ***	3.548(1.6-8.02)	2.05E-03 **
		250				
		249				

Abbreviations: BMI, Body Mass Index; CRR, Complete Remission/Response; SD, Stable Disease; PRR, Partial Remission/Response; PD, Progressive Disease.

splice variant AR-V7, which conferred drug resistance toward enzalutamide in cancer cells [31]. The aforementioned studies collectively indicated that high expression of HNRNPA1 could promote tumor progression. Our results indicated that high expression of the splicing factors HNRNPA1 was a protective factor for the prognosis of EC patients. We hypothesized that HNRNPA1 may have different expression patterns at the tissue level and might exert different biological functions. Further experimental verification of HNRNPA1 should be pursued in EC.

5. Conclusions

In summary, we analyzed the effects of abnormal AS events and splicing factors on the prognosis of EC through whole genome analysis and constructed a new clinical prognosis prediction model for EC based on risk score (All). In addition, we built a correlation network between splicing factors and PASEs, which was important to investigate the potential regulatory mechanisms between splicing factors and PASEs. Altogether, our study provided a new theoretical basis for prognostic evaluation and targeted therapy for EC.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare there are no competing interests.

Authors' Contributions

Caixia Wang conceived and designed the experiments, analyzed the data, contributed with reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft. Mingjun Zheng and Shuang Wang authored or reviewed drafts of the paper and approved the final draft. Xin Nie, Qian Guo, Lingling Gao, Xiao Li, Yue Qi, and Juanjuan Liu approved the final draft. Bei Lin authored or reviewed drafts of the paper, approved the final draft, revised, proofread, and provided ideas and guidance.

Acknowledgments

This work was supported by Shengjing Freelance Research Program [grant number 201804].

Supplementary Materials

Supplementary 1. Figure S1. Kaplan-Meier survival curves of 10 splicing factors in EC. (A-J) K-M survival curves of FMR1, SRSF4, PCBP2, PTBP2, HNRNPDL, YBX1, QKI, RBM5, HNRNPA1, and HNRNPF, respectively.

Supplementary 2. Table S1. Splicing factors and corresponding gene symbols.

Supplementary 3. Table S2. 1166 prognostic-related alternative splicing events with unique TCGA SpliceSeq AS id in EC.

Supplementary 4. Table S3. GO (BP/CC/MF) and KEGG pathway enrichment using the DAVID database.

Supplementary 5. Table S4. 10 prognosis-associated splicing factors in EC.

Supplementary 6. Table S5. Correlation matrix between PASEs and splicing factors.

References

- [1] A. C. Wong, J. E. Rasko, and J. J. Wong, "We skip to work: alternative splicing in normal and malignant myelopoiesis," *Leukemia*, vol. 32, no. 5, pp. 1081–1093, 2018.
- [2] K. Le, B. S. Prabhakar, W. Hong, and L. Li, "Alternative splicing as a biomarker and potential target for drug discovery," *Acta Pharmacologica Sinica*, vol. 36, no. 10, pp. 1212–1218, 2015.
- [3] L. Wan, W. Yu, E. Shen et al., "SRSF6-regulated alternative splicing that promotes tumour progression offers a therapy target for colorectal cancer," *Gut*, vol. 68, no. 1, pp. 118–129, 2019.
- [4] S. Oltean and D. O. Bates, "Hallmarks of alternative splicing in cancer," *Oncogene*, vol. 33, no. 46, pp. 5311–5318, 2014.
- [5] B. D. Wang and N. H. Lee, "Aberrant RNA splicing in cancer and drug resistance," *Cancers (Basel)*, vol. 10, no. 11, 2018.
- [6] A. Sveen, S. Kilpinen, A. Ruusulehto, R. A. Lothe, and R. I. Skotheim, "Aberrant RNA splicing in cancer; Expression changes and driver mutations of splicing factor genes," *Oncogene*, vol. 35, no. 19, pp. 2413–2427, 2016.
- [7] M. Ryan, W. C. Wong, R. Brown et al., "TCGASpliceSeq a compendium of alternative mRNA splicing in cancer," *Nucleic Acids Research*, vol. 44, no. D1, pp. D1018–D1022, 2016.
- [8] P. Morice, A. Leary, C. Creutzberg, N. Abu-Rustum, and E. Darai, "Endometrial cancer," *The Lancet*, vol. 387, no. 10023, pp. 1094–1108, 2016.
- [9] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: A Cancer Journal for Clinicians*, vol. 69, no. 1, pp. 7–34, 2019.
- [10] M. Skrzypczak, I. Merx, S. Schüler-Toprak et al., "Molecular profiling of estrogen receptor α and progesterone receptor transcript variants in endometrial cancer," *Steroids*, vol. 104, pp. 122–128, 2015.
- [11] B. Zhang, X. Shao, J. Zhou, J. Qiu, Y. Wu, and J. Cheng, "YT521 promotes metastases of endometrial cancer by differential splicing of vascular endothelial growth factor A," *Tumor Biology*, vol. 37, no. 12, pp. 15543–15549, 2016.
- [12] M. Siddiqui, C. Simillis, J. Bhoday et al., "A meta-analysis assessing the survival implications of subclassifying T3 rectal tumours," *European Journal of Cancer*, vol. 104, pp. 47–61, 2018.
- [13] S.-W. Jiang, H. Chen, S. Dowdy et al., "HE4 transcription- and splice variants-specific expression in endometrial cancer and correlation with patient survival," *International Journal of Molecular Sciences*, vol. 14, no. 11, pp. 22655–22677, 2013.
- [14] S. Kwon, E. Kim, and K. Baek, "RNPS1 is modulated by ubiquitin-specific protease 4," *FEBS Letters*, vol. 591, no. 2, pp. 369–381, 2017.
- [15] A. Mayeda, J. Badolato, R. Kobayashi, M. Q. Zhang, E. M. Gardiner, and A. R. Krainer, "Purification and characterization

- of human RNPS1: A general activator of pre-mRNA splicing," *EMBO Journal*, vol. 18, no. 16, pp. 4560–4570, 1999.
- [16] K. Fukumura, K. Inoue, and A. Mayeda, "Splicing activator RNPS1 suppresses errors in pre-mRNA splicing: A key factor for mRNA quality control," *Biochemical and Biophysical Research Communications*, vol. 496, no. 3, pp. 921–926, 2018.
- [17] W. Wu, J. E. Baxter, S. L. Wattam et al., "Alternative splicing controls nuclear translocation of the cell cycle-regulated Nek2 kinase," *The Journal of Biological Chemistry*, vol. 282, no. 36, pp. 26431–26440, 2007.
- [18] Z. Liu, Y. Wang, S. Wang, J. Zhang, F. Zhang, and Y. Niu, "Nek2C functions as a tumor promoter in human breast tumorigenesis," *International Journal of Molecular Medicine*, vol. 30, no. 4, pp. 775–782, 2012.
- [19] J. J. Oh, S. S. Byun, and S. E. Lee, "Genetic variations in VDR associated with prostate cancer risk and progression in a Korean population," *Gene*, vol. 533, no. 1, pp. 86–93, 2014.
- [20] Y. Lee and D. C. Rio, "Mechanisms and regulation of alternative Pre-mRNA splicing," *Annual Review of Biochemistry*, vol. 84, pp. 291–323, 2015.
- [21] T. Guo, J. Kong, Y. Liu et al., "Transcriptional activation of NANOG by YBX1 promotes lung cancer stem-like properties and metastasis," *Biochemical and Biophysical Research Communications*, vol. 487, no. 1, pp. 153–159, 2017.
- [22] U. Raffetseder, B. Frye, T. Rauen et al., "Splicing factor SRp30c interaction with Y-box protein-1 confers nuclear YB-1 shuttling and alternative splice site selection," *The Journal of Biological Chemistry*, vol. 278, no. 20, pp. 18241–18248, 2003.
- [23] M. van Dijk, A. Visser, K. M. L. Buabeng, A. Poutsma, R. C. van der Schors, and C. B. M. Oudejans, "Mutations within the LINC-HELLP non-coding RNA differentially bind ribosomal and RNA splicing complexes and negatively affect trophoblast differentiation," *Human Molecular Genetics*, vol. 24, no. 19, pp. 5475–5485, 2015.
- [24] E. Allemand, M. L. Hastings, M. V. Murray, M. P. Myers, and A. R. Krainer, "Alternative splicing regulation by interaction of phosphatase PP2C γ with nucleic acid-binding protein YB-1," *Nature Structural & Molecular Biology*, vol. 14, no. 7, pp. 630–638, 2007.
- [25] T. P. Stricker, C. D. Brown, C. Bandlamudi et al., "Robust stratification of breast cancer subtypes using differential patterns of transcript isoform expression," *PLoS Genetics*, vol. 13, no. 3, Article ID e1006589, 2017.
- [26] T. M. Campbell, M. A. Castro, K. G. de Oliveira, B. A. Ponder, and K. B. Meyer, "ER α binding by transcription factors NFIB and YBX1 enables FGFR2 signaling to modulate estrogen responsiveness in breast cancer," *Cancer Research*, vol. 78, no. 2, pp. 410–421, 2018.
- [27] D. Smith, C. J. R. Stewart, E. M. Clarke et al., "ER and PR expression and survival after endometrial cancer," *Gynecologic Oncology*, vol. 148, no. 2, pp. 258–266, 2018.
- [28] R. Z. Li, J. Hou, Y. Wei, X. Luo, Y. Ye, and Y. Zhang, "hnRNPD1 extensively regulates transcription and alternative splicing," *Gene*, vol. 687, pp. 125–134, 2019.
- [29] R. Roy, Y. Huang, M. J. Seckl, and O. E. Pardo, "Emerging roles of hnRNPA1 in modulating malignant transformation," *Wiley Interdisciplinary Reviews: RNA*, vol. 8, no. 6, 2017.
- [30] Y. Chen, J. Liu, W. Wang et al., "High expression of hnRNPA1 promotes cell invasion by inducing EMT in gastric cancer," *Oncology Reports*, vol. 39, no. 4, pp. 1693–1701, 2018.
- [31] N. Nadiminty, R. Tummala, C. Liu, W. Lou, C. P. Evans, and A. C. Gao, "NF- κ B2/p52:c-Myc:hnRNPA1 pathway regulates expression of androgen receptor splice variants and enzalutamide sensitivity in prostate cancer," *Molecular Cancer Therapeutics*, vol. 14, no. 8, pp. 1884–1895, 2015.