

Prognostic Potential of Alternative Splicing Markers in Endometrial Cancer

Qian Wang,^{1,6} Teng Xu,^{2,3,6} Yu Tong,¹ Jianbo Wu,⁴ Weijian Zhu,⁴ Zhongqiu Lu,⁵ and Jianchao Ying^{4,5}

¹Department of Clinical Laboratory, Wenzhou People's Hospital, The Third Clinical Institute Affiliated to Wenzhou Medical University, Wenzhou, China; ²Research Center for Translational Medicine, Shanghai East Hospital, Tongji University School of Medicine, Shanghai, China; ³Institute of Translational Medicine, Baotou Central Hospital of Inner Mongolia Medical University, Baotou, China; ⁴Central Laboratory, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China; ⁵Institute of Emergency Medicine, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

Alternative splicing (AS), an important post-transcriptional regulatory mechanism that regulates the translation of mRNA isoforms and generates protein diversity, has been widely demonstrated to be associated with oncogenic processes. In this study, we systematically analyzed genome-wide AS patterns to explore the prognostic implications of AS in endometrial cancer (EC). A total of 2,324 AS events were identified as being associated with the overall survival of EC patients, and eleven of these events were further selected using a random forest algorithm. With the implementation of a generalized, boosted regression model, a prognostic AS model that aggregated these eleven markers was ultimately established with high performance for risk stratification in EC patients. Functional analysis of these eleven AS markers revealed various potential signaling pathways implicated in the progression of EC. Splicing network analysis demonstrated the notable correlation between the expression of splicing factors and AS markers in EC and further determined eight candidate splicing factors that could be therapeutic targets for EC. Taken together, the results of this study present the utility of AS profiling in identifying biomarkers for the prognosis of EC and provide comprehensive insight into the molecular mechanisms involved in EC processes.

INTRODUCTION

Despite improvements in screening, diagnosis, curative resection, and preventive strategies, endometrial cancer (EC) is still the most common gynecologic malignancy in developed countries,¹ and the incidence of EC is rising because of increasing obesity of the female population.² Multiple other risk factors have been identified, including long-lasting endogenous or exogenous hyperestrogenism (polycystic ovary, tamoxifen therapy, anovulation, nulliparity), hypertension, and diabetes mellitus.³ Most cases of EC are diagnosed in early stages, since abnormal uterine bleeding is the presenting symptom in 90% of cases, and the final histopathological subtyping and grading based on the hysterectomy specimen are considered the gold standards for correct risk classification of patients for metastatic spread and recurrent disease.⁴ To the best of our knowledge, EC generally has a favorable prognosis, with a 5-year overall survival (OS) reaching 80%, mainly because most women are diagnosed at an early stage and are managed by surgery alone with a low risk of recurrence.⁵ However, the 5-year survival

rate of patients with stage III and IV disease is dramatically decreased, ranging from 42% to 79%.⁶ Therefore, although relatively few women with EC experience recurrence, it accounts for most EC-related deaths. This high incidence and poor prognosis have led tumor markers of EC to become a developing area of research that may help to predict treatment response and patient prognosis.

Developments in high-throughput genomic technologies have opened a new era in cancer genomic research. With the application of RNA sequencing in recent years, gene expression and genomic profiling of EC have been sufficiently evaluated.⁷ Alternative splicing (AS) is an important post-transcriptional regulatory mechanism that regulates the translation of mRNA isoforms and generates protein diversity. Over 95% of human genes undergo AS and encode splice variants in normal physiological processes.⁸ Therefore, dysregulation of AS can affect essential biological processes and thus drive disease-associated pathophysiology.⁹ Emerging data demonstrated that aberrant AS events were closely associated with cancer progression, metastasis, therapeutic resistance, and other oncogenic processes.¹⁰ Thus, cancer-specific splice variants may be used as diagnostic, prognostic, and predictive biomarkers, as well as therapeutic targets.

Due to technical limitations, the effect or functions of AS events in EC have been individually studied in only a small number of cases. A previous study identified the exon 6-skipping mRNA splicing isoform of YT521 as a potential independent prognostic factor for patients with EC.¹¹ In another study, whereas estrogen receptor alpha (ER α /ESR1) expression is regulated by AS, among which ER α D7 is a dominant negative variant, it was determined experimentally that increased expression of ER α D7 was characterized as a prognosticator toward an improved clinical outcome.¹² Additionally, Ouyang et al.¹³

Received 1 May 2019; accepted 25 October 2019;
<https://doi.org/10.1016/j.omtn.2019.10.027>.

⁶These authors contributed equally to this work.

Correspondence: Jianchao Ying, Central Laboratory, The First Affiliated Hospital of Wenzhou Medical University, Nanbaixiang, Wenzhou 325000, China.

E-mail: yingjc@wmu.edu.cn

Correspondence: Zhongqiu Lu, Institute of Emergency Medicine, The First Affiliated Hospital of Wenzhou Medical University, Nanbaixiang, Wenzhou 325000, China.

E-mail: lzq640815@163.com



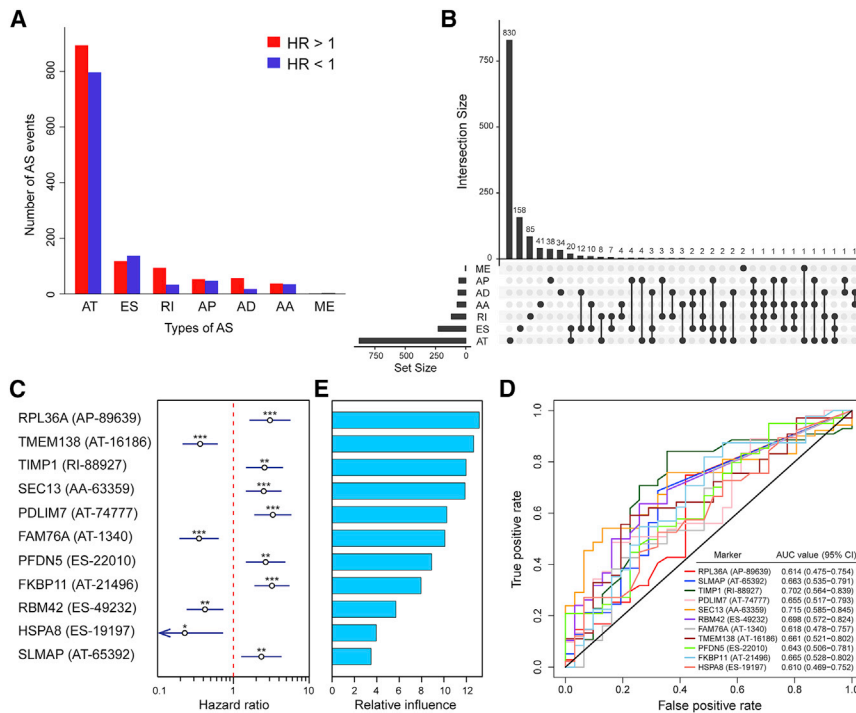


Figure 1. Identification of the Prognostic AS Markers in the Training Cohort

(A) Survival-associated AS events in EC. Number of positively survival-associated (HR < 1) and negatively survival-associated (HR > 1) AS events in EC. (B) UpSet plot of intersections and aggregates among diverse types of survival-associated AS events in EC. One gene may have more than one type of AS event to be associated with patient survival. (C) Forest plot of HRs of the eleven AS markers. (*p < 0.05, **p < 0.01, ***p < 0.001). (D) ROC curves for the eleven AS markers in the testing cohort. (E) Relative influence of the selected AS markers calculated by GBM.

144 genes, 301 alternate acceptor (AA) sites in 257 genes, 235 alternate donor (AD) sites in 173 genes, and 11 mutually exclusive exons (MEs) in 11 genes.

To explore the prognostic utility of an AS signature in EC, AS events associated with OS were identified by fitting univariate Cox proportional hazard regression models in the training cohort. Consequently, 2,324 AS events in 1,290 genes were determined with p values < 0.05 (Figure 1A), including 1,255 negatively survival-associated AS events (hazard ratio [HR] > 1) and 1,069 positively survival-associated AS events (HR < 1). The UpSet plot was generated to visualize the intersecting sets between different genes and AS events (Figure 1B), indicating that one gene might have more than one survival-associated AS event. It is noteworthy that six types of AS in *RPS9*, including AA, AD, AP, AT, ES, and RI, were all associated with OS in EC patients.

Variable Selection and Prognostic Model Construction for EC

A total of 532 potential prognostic AS events (with area under the curve [AUC] values > 0.6), assessed by receiver operating characteristic (ROC) analysis in the training cohort, were retained for further variable selection. By conducting the random forest variable hunting (RFVH) algorithm, a panel of eleven AS events was finally selected as prognostic AS markers (Figure 1C; Table 1). The ability of each AS marker in the OS prediction of EC patients was then demonstrated by ROC curve (Figures 1D and S1) and Kaplan-Meier curve (Figures S2 and S3) analyses.

Subsequently, the percent spliced in (PSI) level of these eleven AS markers in the training cohort was used to construct the prognostic models by implementing the generalized boosted regression model (GBM), least absolute shrinkage and selection operator (LASSO), and multivariate Cox regression algorithms. Based on the PSI levels of these markers, the survival risk scores for each patient were calculated from these models. ROC curve analyses demonstrated that all three models performed well in both cohorts (all AUC > 0.75) (Figures 2A and 2B). Notably, the GBM had the highest AUC value

demonstrated the potential clinical significance of the interaction of two splicing regulators, hnRNP G and hTra2- β 1, in EC patients, opening a door for pharmaceutical targeting options of splicing in future cancer treatment strategies.

Currently, machine learning approaches are increasingly applied in the screening of molecular biomarkers and the construction of prediction classifiers.^{14,15} With the combination of system biology, the prognostic model can recognize specific patterns of diseases and distinguish patients with different survival risks. Furthermore, machine learning can identify candidate biomarkers without bias and effectively improve the sensitivity and specificity of the model.¹⁶ With the rapid accumulation of gene expression data, the public databases provide a rich source for the investigation of AS patterns in EC. Thus, in this study, we systematically analyzed the genome-wide AS patterns and combined them with machine learning to explore the potential prognostic implications of AS in EC.

RESULTS

Identification of Survival-Associated AS Events in EC

After the preprocessing procedure, the mRNA splicing data of the entire EC cohort from The Cancer Genome Atlas (TCGA) SpliceSeq database (<https://bioinformatics.mdanderson.org>),¹⁷ which was enrolled in this study, contains 7,614 AS events in 3,261 genes. Alternate terminator (AT) was the most frequent splice type among the seven AS types, followed by exon skip (ES) and retained intron (RI). Specifically, there were 5,251 ATs in 2,301 genes, 941 ESs in 673 genes, 507 RIs in 414 genes, 368 alternate promoters (APs) in

Table 1. Eleven AS Markers Included in the Prognostic Model of EC

AS ID	Splice Type	Exons	Gene Symbol	PSI Level Association with Poor Prognosis	Candidate Splicing Factor
89639	AP	2	<i>RPL36A</i>	high	<i>RAE1</i>
65392	AT	23.2	<i>SLMAP</i>	high	<i>POM121</i>
88927	RI	3.4	<i>TIMP1</i>	high	<i>NUP153</i>
74777	AT	9	<i>PDLIM7</i>	high	<i>RAE1</i>
63359	AA	10.1	<i>SEC13</i>	high	<i>LSM7</i>
49232	ES	3.2:4:5:6.1:6.2:6.3:7:8:9.1	<i>RBM42</i>	low	<i>RAE1</i>
1340	AT	10	<i>FAM76A</i>	low	<i>CCDC12</i>
16186	AT	4	<i>TMEM138</i>	low	<i>RBM39</i>
22010	ES	2:3	<i>PFDN5</i>	high	<i>CSTF1</i>
21496	AT	7.2	<i>FKBP11</i>	high	<i>LSM7</i>
19197	ES	8.2:9:10.1	<i>HSPA8</i>	low	<i>PRPF18</i>

(0.889, 95% confidence interval [CI]: 0.833–0.945) (Figure 2A) compared with the other two models in the training cohort. Therefore, the GBM that aggregated eleven AS markers was chosen as the optimal prognostic model in this study, and the relative influence of each marker was calculated in the meantime, which indicated their variable importance in the GBM (Figure 1E).

The eleven AS prognostic model was further validated in the testing cohort with an AUC of 0.802 (95% CI: 0.695–0.901) (Figure 2B). Additionally, the patients in the training cohort were divided into two risk groups based on the optimal cut-off point value (−3.319) (Figure 2C) that was determined by the survminer package. As shown in Figure 2D, a significant difference between the OS for patients in these two risk groups was observed by plotting Kaplan-Meier curves (HR = 13.18, $p < 0.001$). An analogous situation was observed in the testing cohort as expected (HR = 4.37, $p < 0.001$) (Figure 2E). Moreover, in comparison with single AS marker, this combination model exhibited an improvement in predictive performance from the ROC curve (Figures 1D and S1) and Kaplan-Meier curve (Figures S2 and S3) analyses. These findings demonstrated that this eleven AS model might be used to predict the prognoses of EC patients.

Performance Evaluation of the Prognostic AS Model

Several clinical variables potentially associated with the prognosis of EC, including age, International Federation of Gynecology and Obstetrics (FIGO) stage, histological grade, and histological type, together with the AS model, were included in univariate and multivariate Cox regression analyses using testing and entire EC cohorts. The results indicated the relatively high prognostic significance of the AS model, as well as the FIGO stage (all $p < 0.05$) (Table 2). To evaluate the effectiveness of the AS model among patients in different FIGO stages, survival analysis was further performed in subsets of patients stratified by FIGO stage. Strikingly, EC patients could be successfully separate into high-risk and low-risk subgroups in both the early (FIGO I/II stage) (Figure 3A) and advanced (FIGO III/IV stage) (Figure 3B) stages by applying this model.

Next, the discrimination of the AS model and FIGO stage in survival analysis was further assessed by multiple methods. The time-dependent AUCs were plotted to demonstrate the 1- to 10-year OS prediction of the FIGO stage, AS model, and combined model comprised of the AS model and FIGO stage (Figure 3C). The AS model showed better predictive ability than the FIGO stage in either integrated AUC (IAUC) (Figure 3D) or concordance index (C-index) (Figure 3E) analyses. Remarkably, the combined model had a larger AUC than the FIGO stage and AS model alone, according to the IAUC analysis (Figure 3D), suggesting that the AS model might also be used to assist the FIGO stage in prognosis predictions for EC patients.

Characterization and Functional Analysis of the Eleven AS Markers

To investigate the effectiveness of AS markers in risk prediction of EC, a comparison of the PSI levels of these eleven AS markers between low- and high-risk EC groups was performed using the entire EC cohort. The PSI level distribution of each AS maker is significantly different between two risk groups (Figure 4A). The changes of three AS events were shown as examples in both SpliceSeq views and Integrative Genomics Viewer (IGV) plots (Figure S4). Regarding the characteristics of these AS markers (Table 1), higher PSI levels of seven markers were associated with shorter OS (HR > 1 in Figure 1A), whereas higher PSI levels of the remaining four markers were related to longer OS (HR < 1 in Figure 1A). Notably, although our study focused on AS markers associated with prognosis of EC, the PSI levels of eight AS markers of them showed significant differences between EC tissues and normal uterine tissues (Figure S5). These findings indicate that these eight markers may be not only related to prognosis of EC but also involved in the tumorigenesis of EC.

To investigate further the underlying biological roles of these eleven AS markers, we determined the corresponding eleven AS marker genes and predicted their interacting genes by performing gene-interaction analysis. A gene-interaction network was further

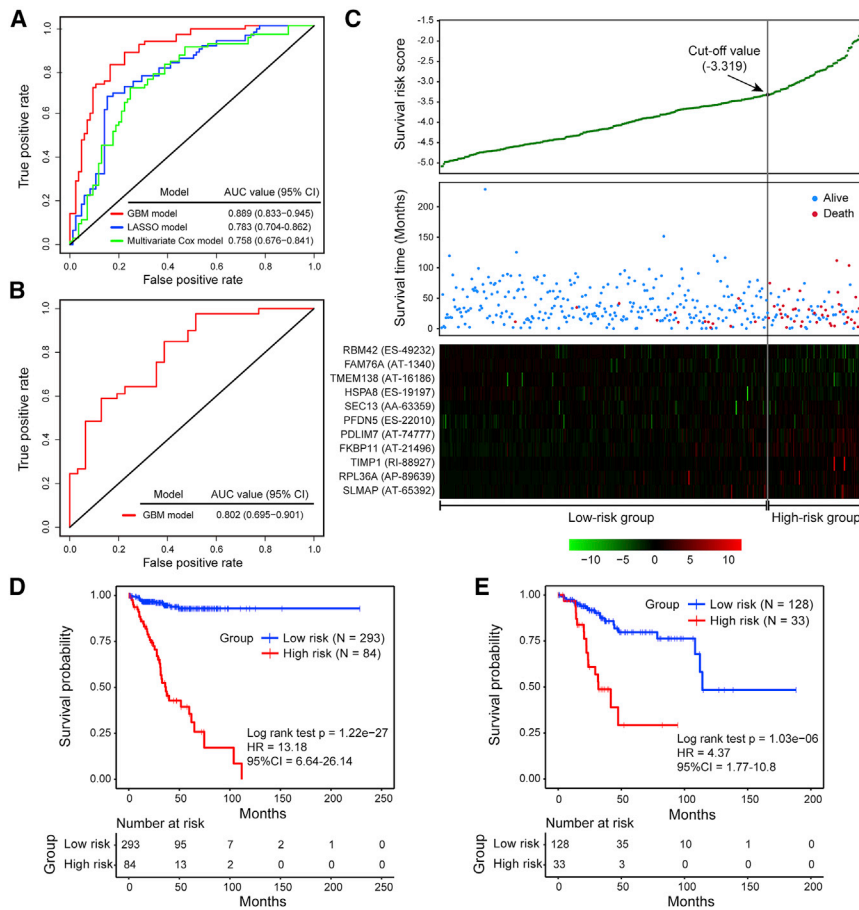


Figure 2. Construction and Validation of the Prognostic AS Model

(A) Optimal model selection based on ROC curves in the training cohort. ROC curves for the GBM, LASSO, and multivariate Cox models were generated for the 5-year OS predictions of EC. (B) ROC curve for the GBM was generated for the 5-year OS predictions of EC in the testing cohort. (C) The risk score analyses of EC patients in the training cohort were performed based on the GBM. Shown are distribution diagram of survival risk score of EC patients (top), survival status of EC patients (middle), and clustering heatmap of the PSI levels of eleven AS markers (bottom). The horizontal axis indicates the patients in order of risk score from low to high. The optimal cut-off point value (-3.319), shown as the gray straight line, was obtained from the training cohort to divide the patients into low- and high-risk groups both in the training and testing cohorts. (D and E) Kaplan-Meier curves for these two risk groups were then plotted to analyze the correlations between this model and the OS in the training (D) and testing (E) cohorts.

Correlation Analysis of the Eleven AS Marker Genes and Splicing Factors

To determine the splicing factors associated with these eleven AS markers in EC, AS-splicing factor correlation analysis was conducted using the entire EC cohort. The AS-splicing regulation network was further constructed based on the correlation coefficient calculated from Spearman's test, and the expression of 68 splicing factors was highly correlated with that of at least

constructed based on the high confidence (interaction score > 0.7), and a total of 1,174 genes interacted with at least one of the eleven genes (Figure 4B). Subsequently, Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses for these interacting genes were conducted (Table S2), which indicated that these genes were significantly associated with several cancer pathways, including pathways of colorectal cancer, prostate cancer, bladder cancer, pancreatic cancer, and renal cell carcinoma (all p values < 0.05) (Figure 4C). Moreover, a series of signaling pathways involved in cancer, such as the phosphatidylinositol 3-kinase (PI3K)-Akt signaling pathway, Hippo signaling pathway, FoxO signaling pathway, and p53 signaling pathway, were also observed (all p values < 0.05) (Figure 4C). In addition, we performed gene set enrichment analysis (GSEA) to elucidate the biological functions of the AS model (Table S3), which revealed that genes highly expressed in the high-risk group showed significant enrichment in multiple biological pathways, such as the ErbB signaling pathway, mismatch repair, and extracellular matrix (ECM)-receptor interaction, whereas the low-risk-related genes were associated with the pathway-related gene set, including chemokine signaling pathway, T cell receptor signaling pathway, and cell adhesion molecules (Figure 4D).

one of the eleven AS markers. Similarly, one AS event might also be regulated by multiple splicing factors (Figure 5A). The top correlation between AS PSI level and splicing factor expression was significantly negative for *FKBP11-LSM7* ($p = 2.44E-38$), *TMEM138-RBM39* ($p = 2.79E-37$), *RBM42-RAE1* ($p = 4.95E-24$), *HSPA8-PRPF18* ($p = 2.62E-16$), and *SEC13-LSM7* ($p = 2.51E-15$) and significantly positive for *RPL36A-RAE1* ($p = 2.83E-15$), *SLMAP-POM121* ($p = 3.48E-14$), *FAM76A-CCDC12* ($p = 1.91E-13$), *TIMP1-NUP153* ($p = 5.00E-13$), *PDLIM7-RAE1* ($p = 2.15E-23$), and *PFDN5-CSTF1* ($p = 4.39E-16$). For instance, the correlation between splicing factor *LSM7* and AT of *FKBP11* is shown in Figure 5B, and the low expression of *LSM7* was associated with poor survival of patients by performing Kaplan-Meier survival analysis ($p = 2.04E-04$) (Figure 5C).

DISCUSSION

This study focused on the identification of prognostic AS markers to explore the utilization of AS signatures in predicting the prognosis of EC patients. In variable screening procedures, the machine learning method was applied, which can identify candidate biomarkers without bias instead of simply selecting the most significant variables. As a result, eleven AS events were selected as prognostic AS markers

Table 2. Univariable and Multivariable Cox Regression Analyses of Potential Prognostic Variables for EC Patients

Variables		Test EC Cohort		Entire EC Cohort	
		HR (95% CI)	p Value	HR (95% CI)	p Value
Univariable Analysis					
Age	>60 versus ≤60	1.99 (0.90–4.36)	0.087	2.11 (1.26–3.53)	4.70E–03
FIGO stage	advanced stage versus early stage	5.43 (2.66–11.11)	3.58E–06	3.96 (2.61–6.01)	8.61E–11
Histologic grade	high grade versus low grade	3.54 (1.47–8.52)	4.73E–03	3.42 (1.99–5.87)	8.20E–06
Histological type	MSE versus EEA	3.96 (1.32–11.92)	1.42E–02	2.86 (1.22–6.69)	1.56E–02
	SEA versus EEA	3.54 (1.76–7.13)	3.94E–04	2.88 (1.87–4.43)	1.69E–06
AS model	high risk versus low risk	4.81 (2.41–9.64)	9.10E–06	8.93 (5.76–13.87)	<2E–16
Multivariable Analysis					
Age	>60 versus ≤60	–	–	1.18 (0.68–2.05)	0.56
FIGO stage	advanced stage versus early stage	3.75 (1.74–8.05)	7.08E–04	3.03 (1.93–4.75)	1.46E–06
Histologic grade	high grade versus low grade	1.54 (0.55–4.26)	0.41	1.55 (0.84–2.86)	0.16
Histological type	MSE versus EEA	2.24 (0.69–7.32)	0.18	1.26 (0.52–3.08)	0.61
	SEA versus EEA	1.39 (0.60–3.24)	0.44	0.74 (0.44–1.25)	0.26
AS model	high risk versus low risk	2.70 (1.23–5.94)	0.013	7.31 (4.42–12.09)	9.99E–15

Advanced stage, I/II stage; early stage, III/IV stage; high grade, G3; low grade, G1/G2; EEA, endometrioid endometrial adenocarcinoma; MSE, mixed serous and endometrioid; SEA, serous endometrial adenocarcinoma.

that might be used to predict the survival of EC patients. Corresponding to these AS markers, we further obtained eleven AS-marker genes that included *RPL36A*, *SLMAP*, *TIMP1*, *PDLIM7*, *SEC13*, *RBM42*, *FAM76A*, *TMEM138*, *PFDN5*, *FKBP11*, and *HSPA8*. Although the implication of most genes in EC progression is unclear, several genes have been reported to be associated with cancer processes in previous studies. For instance, *RPL36A*, representing an overexpression in hepatocellular carcinoma, has been reported to be related to tumor cell proliferation and may be a potential target for anticancer therapy.¹⁸ *SLMAP* encodes tail-anchored protein, and isoforms of *SLMAP*, derived from AS, are targeted to cell membrane, mitochondria, and the microtubule organization center.¹⁹ *SLMAP* has been reported to be an implication in mitosis and cell growth and may be important for normal cell growth and to promote proliferation of giant cell tumor stromal cells.^{20,21} The tumor inhibitor of metalloproteinase, encoded by *TIMP1*, is involved in the process of tumor cell invasion through the ECM.²² Previous studies have demonstrated an association between the relatively high *TIMP1* expression and the poor prognosis of various types of cancer, including non-small cell lung cancer, breast cancer, colon cancer, and pancreatic cancer.^{22–25} Interestingly, a retrospective study on a large cohort of primary breast cancer patients provided evidence that the combined expression of full-length *TIMP1* mRNA and its splice variant lacking exon 2 is associated with good prognosis, which is contrary to the findings of other previous studies.²⁴ The *TIMP1* splicing variant identified in our study is affected in exon 3 and associated with poor prognosis in EC patients, which may indicate that different AS markers play different roles in cancer. An alternative *PFDN5* variant, representing a significant overexpression in malignant thyroid tissues, has been demonstrated to be associated with thyroid tumorigenesis.²⁶ To date, *FKBP11* has not been reported to be linked to cancer, but another *FKBP* family

gene (named *FKBP7*) was highly expressed in melanoma tissue and significantly associated with poor survival, which indicated that *FKBP* members may have strong potential as new therapeutic targets or diagnostic markers in melanoma.²⁷ *HSPA8* has been reported in several types of cancers, such as pancreatic cancer, breast cancer, and EC.^{28–30} It is noteworthy that *HSPA8* was significantly upregulated in EC cells, as confirmed by immunoblot analysis, indicating that *HSPA8* plays a vital role in the development of EC and might be a candidate biomarker for EC.²⁹

To understand further the functional mechanisms behind the prognostic values of these markers, we determined that 1,174 genes interacted strongly with these eleven AS-marker genes by performing gene-interaction analysis. These interacting genes were significantly enriched in several cancer pathways, as well as other signaling pathways involved in cancer in enrichment analysis, such as the PI3K-Akt signaling pathway, Hippo signaling pathway, FoxO signaling pathways, p53 signaling pathway, and transforming growth factor β (TGF- β) signaling pathway. The PI3K signaling pathway, one of the most frequently altered pathways in human cancer, plays a critical role in tumor initiation and progression and has been demonstrated to be activated in the majority of EC cases.^{31,32} Moreover, inhibition of the PI3K/Akt pathway could reverse progesterin resistance in EC, which is the main obstacle to successful conservative therapy in EC patients, indicating that the PI3K/Akt signaling pathway may shed new light on the potential treatment and prognosis of EC.³³ A previous study revealed that the FoxO pathway is involved in breast cancer initiation;³⁴ however, little is known about the role of the FoxO signaling pathway in EC. The Hippo pathway is crucial in human cancer, and the degradation of the Hippo pathway has been reported to occur in a broad range of cancers, including lung cancer, prostate

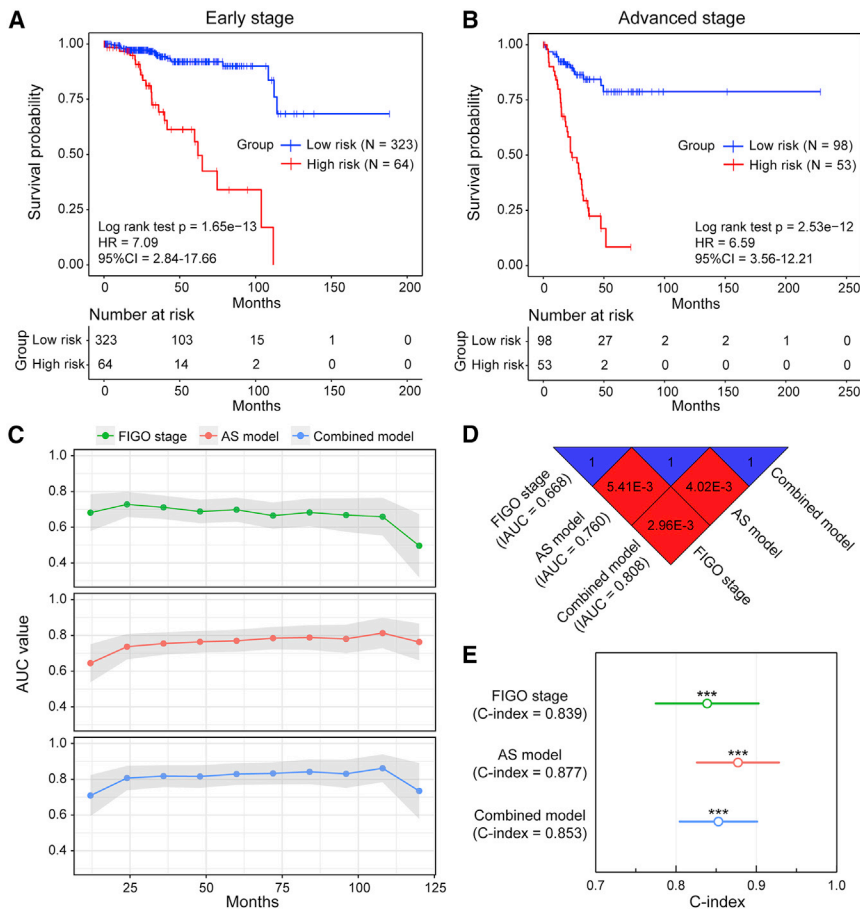


Figure 3. Comparison of Survival Prediction Power of the AS Prognostic Model with FIGO Stage

(A and B) Stratification analysis of the AS model by FIGO stage. EC patients with early (FIGO I/II stage) and advanced stages (FIGO III/IV stage) were divided into low- and high-risk groups using the AS model, respectively. By plotting Kaplan-Meier curves, the prognostic capability for EC patients with early (A) and advanced (B) stages was evaluated individually. (C) The time-dependent AUCs for 1- to 10-year OS prediction of FIGO stage, AS model, and combined model. (D) Comparison of the integrated AUC of FIGO stage, AS model, and combined model. The entry values of the figure represent the p values calculated from the Wilcoxon rank sum test for the comparison between larger IAUC and smaller IAUC. (E) Forest plot of C-index values of FIGO stage, AS model, and combined model (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

cancer, and EC, and is often correlated with poor patient prognosis.^{6,35,36} The p53 pathway is a common oncogenic pathway in EC and many other tumor types, and it has been demonstrated that several markers of the p53 pathway could improve stratification and prognosis of EC.³⁷ The TGF- β signaling pathway is a key network in cell signaling that controls vital processes, including apoptosis and tumorigenesis, and the abnormal regulation of the TGF- β pathway can contribute to a broad range of cancers.³⁸ Given that EC patients were divided into two risk groups by our AS model in the entire cohort, functional investigation of differentially expressed genes between them would be useful to explore specific pathways involved in EC development processes. GSEA identified several molecular pathways associated with cancer, including the ErbB signaling pathway, chemokine signaling pathway, and ECM-receptor interaction. These findings could facilitate our further understanding of the metabolic pathways involved in EC and contribute to the development of new targeted anti-cancer therapies of EC. Nevertheless, the relationships between these pathways and EC require experimental verification.

To the best of our knowledge, it has been determined that the process of splicing is regulated precisely by splicing factors through binding to

splicing regulatory elements of specific genes.³⁹ Therefore, we constructed an AS-splicing regulation network to explore the correlation of eleven AS markers and splicing factors. A total of 68 highly correlated splicing factors were identified to be associated with survival in EC, indicating that they may influence oncogenic processes by regulating the AS of several downstream target genes at the same time. Furthermore, we determined eight candidate splicing factors, including *LSM7*, *RAE1*, *POM121*, *NUP153*, *CCDC12*, *RBM39*, *CSTF1*, and *PRPF18*, which significantly affect these AS markers and that could provide potential therapeutic targets for the treatment of EC. These findings will also help elucidate the underlying mechanisms of AS in the development of EC.

Beyond that, we attempted to construct an optimal prognostic model that could be used to predict prognosis in EC patients. Although each of the eleven AS markers showed a certain prognostic value, the AS-combined model, aggregating multiple markers, outperformed the single AS marker alone, which is consistent with the results of numerous previous studies.^{40,41} In this study, the model construction was carried out by the application of several machine learning and statistics algorithms. Although the multivariate Cox model and LASSO were widely used for model construction in most previous studies, especially on AS,^{15,42} the GBM in this study performed better than other algorithms and was chosen as the final prognostic model. We conclude that it is necessary to implement multiple algorithms for model construction, which may contribute to obtaining the ideal model with optimal performance.

More importantly, the possibility of overfitting of the model has been considered in this study and was mainly controlled in three aspects. First, for variable selection, a univariate prescreening procedure and machine learning-based RFVH method were applied for dimension

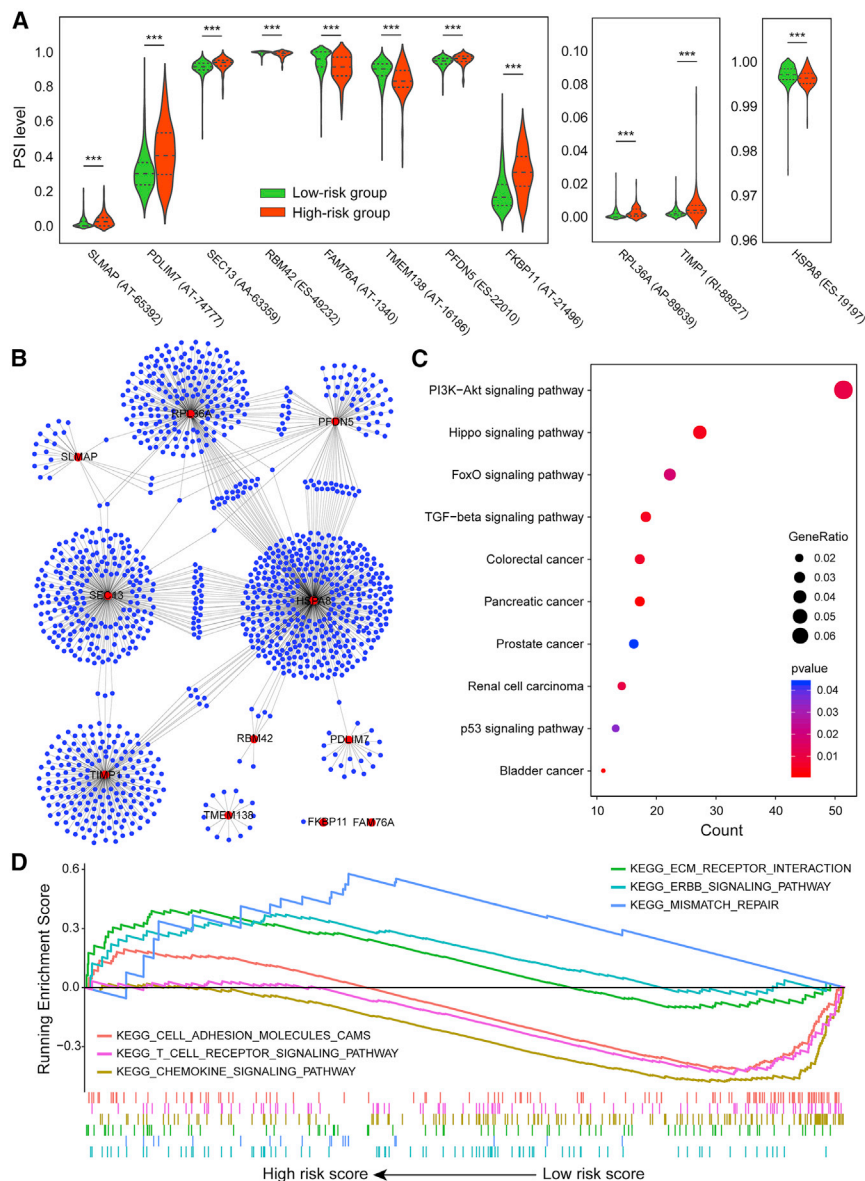


Figure 4. Functional Analysis of AS-Marker Genes
 (A) The PSI levels of eleven AS markers in 421 low-risk patients and 117 high-risk patients. The distributions of the PSI level data are represented by violin plots, and the dashed lines indicate the quartiles. p values were calculated by Mann-Whitney U test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). (B) Visualization of the interaction between the eleven AS-marker genes and 1,174 genes. The red circle indicates the AS-marker genes, and the blue circle represents the other interacting genes. (C) KEGG functional enrichment of these interacting genes in EC. Ten significant pathways involved in cancer are displayed. (D) GSEA delineates biological pathways correlated with risk scores. Several enrichment results with significant associations between high- and low-risk groups are shown.

was derived from a validation cohort of 506 EC patients from TCGA database, which was also used for variable selection and model construction. Therefore, a cohort, independent of both studies, would be a better way to compare the performance of these two models. Nevertheless, by contrast, our AS model exhibited increased AUC values (AUC > 0.8 in both training and testing cohorts).

Further evaluation procedures for this model were performed using testing and entire cohorts, and the prognostic model was demonstrated to be an independent prognostic factor for predicting OS in EC patients. Similar to previous studies,^{15,40} the FIGO staging system, one of the most adopted classifications for the treatment and prognosis for EC patients,⁴³ also exhibited high prognostic significance in this study. Remarkably, this model is suitable for prognosis prediction under different FIGO stages and can further distinguish patients with an elevated risk of mortality stratified by the FIGO stage. In addition, the survival prediction power of this AS model was further compared with the FIGO stage, demonstrating that this model has higher accuracy and might assist the FIGO stage in prognosis prediction for EC patients. However, the prognostic implication of this AS model for EC clearly requires validation through further functional experiments and clinical trials.

reduction, so as to make the model more capable of generalization and combat overfitting. Second, for model construction, the CV was employed to estimate the optimal number of iterations in the GBM algorithm, which could reduce the possibility of the overfitting in model selection. Last but not least, the testing or validation cohort used for model validation was often absent in several previous studies on AS, which may lead to overfitting of the model and not guarantee the validity of the model in other samples. Understandably, it is difficult to obtain additional large-scale samples; thus, we set up the testing cohort by splitting the entire cohort randomly for validation and evaluation. Prior to our studies, Gao et al.⁴² has proposed a new AS-based prediction model for EC, which achieved good prognostic performance (AUC = 0.758). In the study of Gao et al.,⁴² the AUC value

Overall, we identified eleven prognostic AS markers and constructed a prognostic AS model that could efficiently facilitate survival prediction for EC patients and guide the application of rational therapy in clinical practice. This study also provided insight into the underlying mechanisms involved in the development and progression of EC.

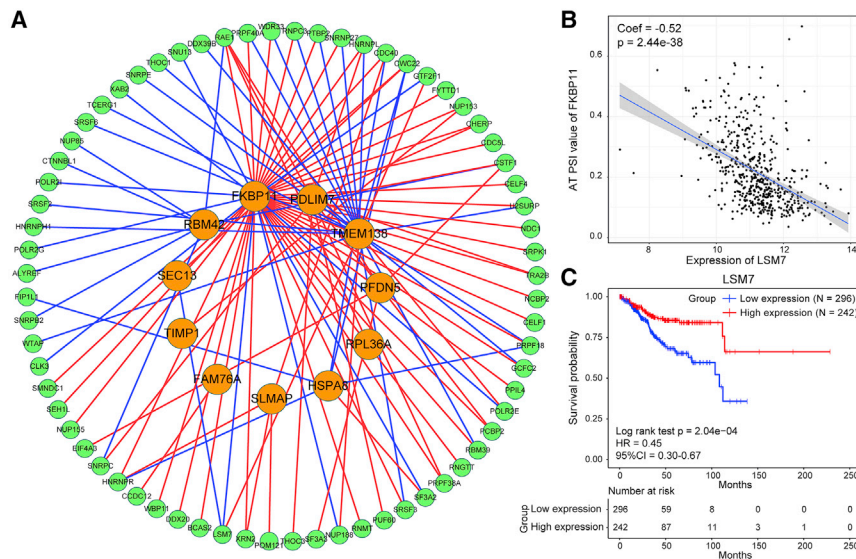


Figure 5. Construction of the AS-Splicing Factor Correlation Network

(A) Cytoscape visualization of the correlation of 11 AS markers and 68 splicing factors (Spearman's correlation coefficient > 0.30, $p < 0.05$). AS markers and splicing factors are represented with orange and green dots, respectively. The positive/negative correlation between the expression of splicing factors and PSI values of AS is denoted with red/blue lines. (B) Dot plot of the correlation between the expression of *LSM7* and the AT PSI value of *FKBP11*. (C) Low expression (blue line) of splicing factor *LSM7* was significantly associated with poor OS in EC.

MATERIALS AND METHODS

Data Sources and Data Processing

mRNA splicing data of the EC cohort were obtained from TCGA SpliceSeq database (<https://bioinformatics.mdanderson.org>), which included seven common types of AS events: ES, ME, RI, AP, AT, AD, and AA.^{17,44} The PSI value,⁴⁵ a common, intuitive ratio for quantifying splicing events from 0 to 1, was calculated for each sample and every possible splice event. In detail, PSI is the ratio of normalized read counts, indicating inclusion of a transcript element over the total normalized reads for that event (both inclusion and exclusion reads).⁴⁶ The corresponding clinical parameters and expression profile data (reads counting with HTSeq) were retrieved from TCGA database (<https://portal.gdc.cancer.gov/>). Patients without complete information (i.e., survival time, age, FIGO stage, histological grade, and histological type) were removed, and a total of 538 EC patients were finally included in this study (Table S1). To avoid the impact of missing values on subsequent analysis, PSI values for any AS events that did not exist across all 538 samples were also excluded. Splicing factor genes in the mRNA splicing pathway were obtained from Reactome (<https://reactome.org/>) and PathCards database (<https://pathcards.genecards.org/>). The entire cohort was randomly split into training ($n = 377$) and testing cohorts ($n = 161$) at a 7:3 ratio (Table S1). The training cohort was mainly used for variable/marker selection and model construction, whereas the testing cohort was only used for validation and evaluation of the model.

Identification of Prognostic AS Markers in EC

To remove excessive noise and accelerate the computational procedure, a univariate prescreening procedure (univariate Cox regression) was performed on the training cohort, which was generally conducted prior to the application of any variable selection method.⁴⁷ The “surv_cutpoint” function (survminer package) is an outcome-oriented method providing a value of a cut-point that corresponds to the most significant relation with survival using the maximally

selected rank statistics from the maxstat package and was employed to determine the optimal cut-off point for an AS event or prognostic model. The patients were then divided into high-risk and low-risk groups by cut-off point value for each AS event. Kaplan-Meier survival curves and log rank tests were used to assess the differences in OS of these two groups. HR and p values were calculated to compare survival curves by using the survival package. The timeROC package in R, which allows for time-dependent ROC curve estimation with censored data, was used to generate AUC of the ROC curve and estimate the sensitivity and specificity of these AS events. RfVH, a variable selection method suitable for high-dimensionality data,⁴⁸ was implemented in the randomForestSRC package and used for marker selection with an iteration procedure, according to minimal depth and variable importance scores at each iteration step. After 100 Monte Carlo iterations, the AS events were ranked by the frequency of occurrence, and the average number (P) of selected AS events per iteration was also determined. The top P ranked AS events were finally selected as AS markers. With the use of the SciPy package in Python, the Mann-Whitney U test was performed to examine the differential PSI level of AS markers between high- and low-risk groups of EC patients, as well as EC and normal uterine tissues.

Construction and Evaluation of the Prognostic Model of EC

With the aggregation of the PSI level of AS markers selected above in the training cohort, three approaches for statistics and machine learning, including the GBM, LASSO, and multivariate Cox regression, were employed to create AS-combined models for predicting the prognosis of EC patients. In detail, GBM, an implementation of boosting for the Cox proportional hazard model, implements extensions to Freund and Schapire's AdaBoost algorithm and Friedman's gradient boosting machine in the gbm package. Boosting is the process of iteratively adding basis functions in a greedy fashion, such that each additional basis function further reduces the selected loss function.⁴⁹ The 10-fold CV was conducted to calculate an estimate of generalization error for each boosting iteration, and the optimal number of boosting iterations was determined by the minimum generalization error for reducing the possibility of the over-fitting in model selection. An eleven AS model was constructed, and the relative influence of each marker was also calculated to measure the

variable importance. The Cox model, regularized by the LASSO penalty, was conducted in the glmnet package. The optimal step was determined by the expected generalization error estimated from 10-fold CV, and a LASSO model was finally built based on the 8 of 11 markers. Additionally, multivariate Cox regression was applied to build a model and remove any AS markers that might not be independent factors in the model, and a five AS model was obtained.

Based on the prediction score from GBM, the optimal cut-off point value was calculated using the survminer package and was then used to stratify patients into distinct prognostic groups. Subsequently, the ROC curve and Kaplan-Meier curve were used to evaluate the performance of the model in prognosis prediction of EC. The Wilcoxon rank sum test implemented in the survcomp package was employed to compare any two IAUCs through the results of time-dependent ROC curves at the time points of 1 to 10 years. The C-index of the prognostic model was computed to assess their discrimination in survival analysis.

Gene Network Construction and Functional Enrichment Analysis

UpSet plot, a novel visualization tool for the quantitative analysis of interactive sets, was used to analyze the intersections among the seven types of AS. A gene-interaction network was constructed by importing the AS-marker genes into the STRING database (<https://string-db.org/>). KEGG pathway enrichment analyses of the interacted genes were performed using the clusterProfiler package.⁵⁰ Only the pathways with a p value threshold of <0.05 were considered to be significantly enriched functional categories. GSEA was performed to determine whether an *a priori* defined set of genes shows statistical significance, concordant differences between two biological states. In detail, with the use of corresponding gene-expression profiles of EC patients, differential expression analysis was performed with the DESeq2 package to rank all genes based on the fold change between two different risk groups of patients. Then, the entire ranked list was used to assess how the genes of each gene set are distributed across the ranked list. GSEA was conducted with the clusterProfiler package using the gene set of “c2.cp.kegg.v6.1.entrez” downloaded from the Molecular Signatures Database (MSigDB) database. Gene sets with a p value <0.05 and a q value <0.25 after performing 1,000 permutations were considered to be significantly enriched.

Correlation Analyses of AS Markers and Splicing Factor Genes

Correlations between the expression levels of splicing factors and the PSI levels of AS markers were analyzed by Spearman's test. A p value of <0.05 and a correlation coefficient of >0.30 were considered to be significant. The correlation network was then visualized by Cytoscape software.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.omtn.2019.10.027>.

AUTHOR CONTRIBUTIONS

J.Y. and Z.L. designed the study. Q.W., T.X. and Y.T. performed the data analysis. J.W. and W.Z. contributed to the interpretation of results. Q.W. and T.X. drafted the manuscript. J.Y. revised the manuscript. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare no competing interests.

ACKNOWLEDGMENTS

This work was supported by grants from the Natural Science Foundation of Zhejiang Province, China (grant number: LQ20H150004); the Science & Technology Project of Inner Mongolia Autonomous Region, China (grant number: 201802125); the National Natural Science Foundation of China (grant number: 81960381); and the start-up funds from the First Affiliated Hospital of Wenzhou Medical University (grant number: 2018QD014).

REFERENCES

- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424.
- Kitson, S., Ryan, N., MacKintosh, M.L., Edmondson, R., Duffy, J.M., and Crosbie, E.J. (2018). Interventions for weight reduction in obesity to improve survival in women with endometrial cancer. *Cochrane Database Syst. Rev.* 2, CD012513.
- Colombo, N., Preti, E., Landoni, F., Carinelli, S., Colombo, A., Marini, C., and Sessa, C.; ESMO Guidelines Working Group (2013). Endometrial cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* 24 (Suppl 6), vi33–vi38.
- Werner, H.M., Trovik, J., Marcickiewicz, J., Tingulstad, S., Staff, A.C., Engh, M.E., Oddenes, K., Rokne, J.A., Tjugum, J., Lode, M.S., et al. (2013). A discordant histological risk classification in preoperative and operative biopsy in endometrial cancer is reflected in metastatic risk and prognosis. *Eur. J. Cancer* 49, 625–632.
- Ouldamer, L., Bendifallah, S., Body, G., Touboul, C., Graesslin, O., Raimond, E., Collinet, P., Coutant, C., Lavoué, V., Lévêque, J., et al. (2016). Predicting poor prognosis recurrence in women with endometrial cancer: a nomogram developed by the FRANCOGYN study group. *Br. J. Cancer* 115, 1296–1303.
- Mitamura, T., Watari, H., Wang, L., Kanno, H., Kitagawa, M., Hassan, M.K., Kimura, T., Tanino, M., Nishihara, H., Tanaka, S., and Sakuragi, N. (2014). microRNA 31 functions as an endometrial cancer oncogene by suppressing Hippo tumor suppressor pathway. *Mol. Cancer* 13, 97.
- Chen, B.J., Byrne, F.L., Takenaka, K., Modesitt, S.C., Olzomer, E.M., Mills, J.D., Farrell, R., Hoehn, K.L., and Janitz, M. (2017). Transcriptome landscape of long intergenic non-coding RNAs in endometrial cancer. *Gynecol. Oncol.* 147, 654–662.
- Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457–463.
- Gamazon, E.R., and Stranger, B.E. (2014). Genomics of alternative splicing: evolution, development and pathophysiology. *Hum. Genet.* 133, 679–687.
- Climente-González, H., Porta-Pardo, E., Godzik, A., and Eyras, E. (2017). The Functional Impact of Alternative Splicing in Cancer. *Cell Rep.* 20, 2215–2226.
- Zhang, B., zur Hausen, A., Orłowska-Volk, M., Jäger, M., Bettendorf, H., Stamm, S., Hirschfeld, M., Yiqin, O., Tong, X., Gitsch, G., and Stickeler, E. (2010). Alternative splicing-related factor YT521: an independent prognostic factor in endometrial cancer. *Int. J. Gynecol. Cancer* 20, 492–499.
- Hirschfeld, M., Ouyang, Y.Q., Jaeger, M., Erbes, T., Orłowska-Volk, M., Zur Hausen, A., and Stickeler, E. (2015). HNRNP G and HTRA2-BETA1 regulate estrogen receptor alpha expression with potential impact on endometrial cancer. *BMC Cancer* 15, 86.

13. Ouyang, Y.Q., zur Hausen, A., Orlowska-Volk, M., Jäger, M., Bettendorf, H., Hirschfeld, M., Tong, X.W., and Stickeler, E. (2011). Expression levels of hnRNP G and hTra2-beta1 correlate with opposite outcomes in endometrial cancer biology. *Int. J. Cancer* 128, 2010–2019.
14. Sinha, M., Jupe, J., Mack, H., Coleman, T.P., Lawrence, S.M., and Fraley, S.I. (2018). Emerging Technologies for Molecular Diagnosis of Sepsis. *Clin. Microbiol. Rev.* 31, e00089-17.
15. Ying, J., Xu, T., Wang, Q., Ye, J., and Lyu, J. (2018). Exploration of DNA methylation markers for diagnosis and prognosis of patients with endometrial cancer. *Epigenetics* 13, 490–504.
16. Ko, E.R., Yang, W.E., McClain, M.T., Woods, C.W., Ginsburg, G.S., and Tsalik, E.L. (2015). What was old is new again: using the host response to diagnose infectious disease. *Expert Rev. Mol. Diagn.* 15, 1143–1158.
17. Ryan, M., Wong, W.C., Brown, R., Akbani, R., Su, X., Broom, B., Melott, J., and Weinstein, J. (2016). TCGASpliceSeq a compendium of alternative mRNA splicing in cancer. *Nucleic Acids Res.* 44 (D1), D1018–D1022.
18. Kim, J.H., You, K.R., Kim, I.H., Cho, B.H., Kim, C.Y., and Kim, D.G. (2004). Over-expression of the ribosomal protein L36a gene is associated with cellular proliferation in hepatocellular carcinoma. *Hepatology* 39, 129–138.
19. Guzzo, R.M., Sevinc, S., Salih, M., and Tuana, B.S. (2004). A novel isoform of sarcolemmal membrane-associated protein (SLMAP) is a component of the microtubule organizing centre. *J. Cell Sci.* 117, 2271–2281.
20. Chen, K., Yang, X., Wu, L., Yu, M., Li, X., Li, N., Wang, S., and Li, G. (2013). Pinellia pedatisecta agglutinin targets drug resistant K562/ADR leukemia cells through binding with sarcolemmal membrane associated protein and enhancing macrophage phagocytosis. *PLoS ONE* 8, e74363.
21. Fellenberg, J., Saehr, H., Lehner, B., and Depeweg, D. (2012). A microRNA signature differentiates between giant cell tumor derived neoplastic stromal cells and mesenchymal stem cells. *Cancer Lett.* 321, 162–168.
22. Fong, K.M., Kida, Y., Zimmerman, P.V., and Smith, P.J. (1996). TIMP1 and adverse prognosis in non-small cell lung cancer. *Clin. Cancer Res.* 2, 1369–1372.
23. D'Costa, Z., Jones, K., Azad, A., van Stiphout, R., Lim, S.Y., Gomes, A.L., Kinchesh, P., Smart, S.C., Gillies McKenna, W., Buffa, F.M., et al. (2017). Gemcitabine-Induced TIMP1 Attenuates Therapy Response and Promotes Tumor Growth and Liver Metastasis in Pancreatic Cancer. *Cancer Res.* 77, 5952–5962.
24. Sieuwerts, A.M., Usher, P.A., Meijer-van Gelder, M.E., Timmermans, M., Martens, J.W., Brünner, N., Klijn, J.G., Offenberg, H., and Foekens, J.A. (2007). Concentrations of TIMP1 mRNA splice variants and TIMP-1 protein are differentially associated with prognosis in primary breast cancer. *Clin. Chem.* 53, 1280–1288.
25. Song, G., Xu, S., Zhang, H., Wang, Y., Xiao, C., Jiang, T., Wu, L., Zhang, T., Sun, X., Zhong, L., et al. (2016). TIMP1 is a prognostic marker for the progression and metastasis of colon cancer through FAK-PI3K/AKT and MAPK pathway. *J. Exp. Clin. Cancer Res.* 35, 148.
26. Guimarães, G.S., Latini, F.R., Camacho, C.P., Maciel, R.M., Dias-Neto, E., and Cerutti, J.M. (2006). Identification of candidates for tumor-specific alternative splicing in the thyroid. *Genes Chromosomes Cancer* 45, 540–553.
27. Hagedorn, M., Siegfried, G., Hooks, K.B., and Khatib, A.M. (2016). Integration of zebrafish fin regeneration genes with expression data of human tumors in silico uncovers potential novel melanoma markers. *Oncotarget* 7, 71567–71579.
28. Tian, Y., Xu, H., Farooq, A.A., Nie, B., Chen, X., Su, S., Yuan, R., Qiao, G., Li, C., Li, X., et al. (2018). Maslinic acid induces autophagy by down-regulating HSPA8 in pancreatic cancer cells. *Phytother. Res.* 32, 1320–1331.
29. Shan, N., Zhou, W., Zhang, S., and Zhang, Y. (2016). Identification of HSPA8 as a candidate biomarker for endometrial carcinoma by using iTRAQ-based proteomic analysis. *OncoTargets Ther.* 9, 2169–2179.
30. Zagouri, F., Sergeantanis, T.N., Gazouli, M., Tsigginou, A., Dimitrakakis, C., Papaspyrou, I., Eleutherakis-Papaiakovou, E., Chrysikos, D., Theodoropoulos, G., Zografos, G.C., et al. (2012). HSP90, HSPA8, HIF-1 alpha and HSP70-2 polymorphisms in breast cancer: a case-control study. *Mol. Biol. Rep.* 39, 10873–10879.
31. Lien, E.C., Dibble, C.C., and Tokor, A. (2017). PI3K signaling in cancer: beyond AKT. *Curr. Opin. Cell Biol.* 45, 62–71.
32. Weigelt, B., Warne, P.H., Lambros, M.B., Reis-Filho, J.S., and Downward, J. (2013). PI3K pathway dependencies in endometrioid endometrial cancer cell lines. *Clin. Cancer Res.* 19, 3533–3544.
33. Gu, C., Zhang, Z., Yu, Y., Liu, Y., Zhao, F., Yin, L., Feng, Y., and Chen, X. (2011). Inhibiting the PI3K/Akt pathway reversed progesterin resistance in endometrial cancer. *Cancer Sci.* 102, 557–564.
34. Smit, L., Berns, K., Spence, K., Ryder, W.D., Zeps, N., Madiredjo, M., Beijersbergen, R., Bernards, R., and Clarke, R.B. (2016). An integrated genomic approach identifies that the PI3K/AKT/FOXO pathway is involved in breast cancer tumor initiation. *Oncotarget* 7, 2596–2610.
35. Liu, X., Sempere, L.F., Ouyang, H., Memoli, V.A., Andrew, A.S., Luo, Y., Demidenko, E., Korc, M., Shi, W., Preis, M., et al. (2010). MicroRNA-31 functions as an oncogenic microRNA in mouse and human lung cancer cells by repressing specific tumor suppressors. *J. Clin. Invest.* 120, 1298–1309.
36. Zhao, B., Li, L., Wang, L., Wang, C.Y., Yu, J., and Guan, K.L. (2012). Cell detachment activates the Hippo pathway via cytoskeleton reorganization to induce anoikis. *Genes Dev.* 26, 54–68.
37. Edmondson, R.J., Crosbie, E.J., Nickkho-Amiry, M., Kaufmann, A., Stelloo, E., Nijman, H.W., Leary, A., Auguste, A., Mileschkin, L., Pollock, P., et al. (2017). Markers of the p53 pathway further refine molecular profiling in high-risk endometrial cancer: A TransPORTEC initiative. *Gynecol. Oncol.* 146, 327–333.
38. Mirzaei, H., and Faghiloo, E. (2018). Viruses as key modulators of the TGF- β pathway; a double-edged sword involved in cancer. *Rev. Med. Virol.* 28.
39. Jyotsana, N., and Heuser, M. (2018). Exploiting differential RNA splicing patterns: a potential new group of therapeutic targets in cancer. *Expert Opin. Ther. Targets* 22, 107–121.
40. Ying, J., Wang, Q., Xu, T., and Lyu, J. (2018). Establishment of a nine-gene prognostic model for predicting overall survival of patients with endometrial carcinoma. *Cancer Med.* 7, 2601–2611.
41. Li, Y., Sun, N., Lu, Z., Sun, S., Huang, J., Chen, Z., and He, J. (2017). Prognostic alternative mRNA splicing signature in non-small cell lung cancer. *Cancer Lett.* 393, 40–51.
42. Gao, L., Xie, Z.C., Pang, J.S., Li, T.T., and Chen, G. (2019). A novel alternative splicing-based prediction model for uteri corpus endometrial carcinoma. *Aging (Albany N.Y.)* 11, 263–283.
43. Morice, P., Leary, A., Creutzberg, C., Abu-Rustum, N., and Darai, E. (2016). Endometrial cancer. *Lancet* 387, 1094–1108.
44. Matlin, A.J., Clark, F., and Smith, C.W. (2005). Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.* 6, 386–398.
45. Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476.
46. Ryan, M.C., Cleland, J., Kim, R., Wong, W.C., and Weinstein, J.N. (2012). SpliceSeq: a resource for analysis and visualization of RNA-Seq data on alternative splicing and its functional impacts. *Bioinformatics* 28, 2385–2387.
47. Wasserman, L., and Roeder, K. (2009). High-Dimensional Variable Selection. *Ann. Stat.* 37 (5A), 2178–2201.
48. Ishwaran, H., Kogalur, U.B., Gorodeski, E.Z., Minn, A.J., and Lauer, M.S. (2010). High-Dimensional Variable Selection for Survival Data. *J. Am. Stat. Assoc.* 105, 205–217.
49. Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.* 29, 1189–1232.
50. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287.