SEC23A Regulated by Splicing Factor EIF3A may be Involved in the Development of Cervical Squamous Cell Cancer via Adherens Junction Pathway

Chen Li
Nantong Medical College: Nantong University Medical School
Fang Yan
Sun Yat-sen University Cancer Center
Ting Wan
Sun Yat-sen University Cancer Center
Haifeng Gu
Sun Yat-sen University Cancer Center
Gang Chen
Sun Yat-sen University Cancer Center
Qiang Li
Sun Yat-sen University Cancer Center
Hua Xian
Affiliated Hospital of Nantong University
Jingdun Xie (xiejd6@mail.sysu.edu.cn)
Sun Yat-sen University Cancer Center https://orcid.org/0000-0002-3820-8435

Primary research

Keywords: Alternative splicing, splicing factor, cervical squamous cell cancer, Spliceseq database, TCGA

Posted Date: November 5th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-101308/v1

Abstract

Background: Cervical squamous cell cancer is a common malignancy in women globally. Increasing studies have indicated that there was an indivisible association between alternative splicing and cancer. However, comprehensive analysis of alternative splicing is scarce in cervical squamous cell cancer.

Methods: Alternative mRNA splicing events data and clinical information of 216 cervical squamous cell cancer patients were downloaded from TCGA SpliceSeq database and TCGA website. We used identified survival-associated splicing events to construct prognostic signatures. Kaplan-Meier survival analysis and receiver operating characteristic curve were performed to evaluate the clinical value of prognostic signatures. A nomogram was carried out to quantitatively predict individuals' survival probability. Regulatory network between splicing factors and survival-associated splicing events was analyzed. Additionally, we performed Pearson's correlation analysis between survival-associated splicing events and pathways to explore potential downstream pathways. In order to verify the relationships between EIF3A and cell proliferation and migration, CCK8 and wound healing assays were conducted.

Results: After sorting out survival-associated splicing events, multivariable regression analysis was used to acquire 70 survival-associated splicing events that could be independent prognostic factors for overall survival in cervical squamous cell cancer. A nomogram was constructed with a concordance index of 0.82. EIF3A positively regulated SEC23A-27346-AP with highest correlation coefficient (P□0.001, R=0.69). The most significant KEGG pathways was adherens junction pathway (P=0.02, R=0.65). Cell proliferation and wound healing assays showed that EIF3A knockdown decreases cell proliferation and cell migration.

Conclusion: Prognostic signatures of survival-associated splicing events were independent prognostic factors for overall survival among cervical squamous cell cancer patients. A nomogram could quantitatively predict individuals' survival probability by integrating multiple risk factors. Moreover, SEC23A positively regulated by EIF3A may contribute to cervical squamous cell cancer via adherens junction pathway. This study may offer new direction for subsequent experimental research in cervical squamous cell cancer.

Introduction

Cervical squamous cell cancer (CESC) is the fourth most frequently detected malignancy and the fourth leading cause of cancer-associated death in women globally[1]. On the basis of the recent global cancer statistics, there were 570,000 newly diagnosed CESC cases and 311,000 deaths in 2018 worldwide[2]. The incidence and mortality of CESC in developing countries still drastically continue to rise, and CESC tends to occur in younger women[3]. Clinically, surgical resection, chemotherapy and radiotherapy, alone or combined, sometimes may be effective when treating early-stage, focal, sensitive CESC[4]. However, for a considerable proportion of CESC patients with unsatisfying treatment effect, we starve for a novel therapeutic strategy. In addition, clinical prediction for survival time depends heavily on histopathologic type, TNM stage and a fraction of tumor markers, while aforementioned indicators fail to predict survival

time individually[5]. Along with the development of scientific research, we discovered a growing number of biomarker or molecular mechanisms underlying tumorigenesis and progression. These molecular markers gradually guide the nosological classification, precisely evaluate the prognostic outcome, and even become an effective therapeutic target[6–8].

Alternative mRNA splicing (AS) is a ubiquitous biological process that can remarkably increase protein species diversity under the circumstance of a relatively finite number of genes[9]. The exact mechanism is that AS enables precursor mRNA to generate different types of mature mRNA by multiple combinations of introns and exons. There are seven types of alternative splicing pattern, that is, mutually exclusive exons (ME), retained intron (RI), alternate donor site (AD), alternate acceptor site (AA), alternate promotor (AP), alternate terminator (AT) and exon skip (ES)[10]. Under the physiological condition, AS plays a vitally important regulating effect in the process of development, tissue identity, differentiation, cell-to-cell communication, cell senescence and apoptosis[11–13]. AS is accurately regulated by extensive splicing factors which participate in the precise selection of splicing sites and subsequent splicing events[14]. Undoubtedly, aberrant splicing factors or abnormal expression of splicing factors directly leads to changes of splicing events expression, thus causing abnormal cell growth[15]. There is an increasing number of studies that AS also makes a critical difference in the development and progression of cancer[16]. The main mechanism may be the involvement of AS throughout the whole stage of cancer, including tumorigenesis, abnormal proliferation, progression, angiogenesis, metastasis, and immune escape[17–19]. A recent study hinted that there was an indivisible association between AS and cancer prognosis[20]. Therefore, it may be a credible prognostic biomarker and effective therapeutic target. Previous studies demonstrated that AS event-associated signatures may act as prognostic biomarkers in some cancers, including stomach adenocarcinoma, colorectal cancer, prostate adenocarcinoma, sarcoma, bladder urothelial carcinoma and so on[21–25]. However, comprehensive profiles of alternative splicing events are still scarce in cervical squamous cell cancer.

In this study, we constructed an alternative splicing event-associated prognostic signature that could predict overall survival in cervical squamous cell carcinoma, and integrated multiple risk factors, including alternative splicing event-associated prognostic signature and clinical characteristics, to predict the individualized survival probability. Moreover, we investigated the potential interaction network between splicing events and splicing factors, and downstream regulatory network of alternative splicing events in CESC.

Materials And Methods

Data acquirement of alternative splicing events

The data of alternative splicing events for CESC were downloaded from the TCGA SpliceSeq database[26]. Clinical characteristics of 308 CESC patients were obtained by searching the TCGA website. 261 patients whose survival time was more than 90 days met the inclusion criteria. After we

matched the patients with their TCGA SpliceSeq database items according to their sample ID, a total of 216 patients were included in our study.

Identification and construction of the prognostic signature

Univariable cox regression analysis was carried out to identify survival-associated splicing events (SASEs). We used these candidate SASEs to construct prognostic signatures that could predict overall survival time. In order to avoid prognostic signatures overfitting and build an optimal prognostic model, the least absolute shrinkage and selection operator (LASSO) regression analysis was applied to screen out splicing events whose absolute value of coefficients were greater than a predetermined value by using the R package “glmnet”[27]. After excluding SASEs with zero coefficients in the LASSO regression analysis, we calculated risk score of each patient for overall survival prediction by using the formula:

$$\text{risk score} = \sum_i^n PSI_i * \beta_i$$

PSI_i, that is, percent-spliced-in, is a ratio that indicates the efficiency of splicing of sequences of interest into transcripts, and can be used to undertake an intuitive quantitative comparison of splicing events[28]. β_i signifies the coefficients of SASEs in the LASSO regression model. CESC patients were divided into low-risk and high-risk groups by using the median risk score as a cutoff value, respectively.

Assessment of the clinical value of risk scores for prognostic signature

Kaplan-Meier (K-M) survival analysis and receiver operating characteristic (ROC) curve were performed to assess the prognostic value of risk scores, with the area under ROC curve (AUC) indicating the predictive efficacy of prognostic signature construction. Multivariable cox proportional hazards regression analysis was conducted to validate whether risk scores were independent prognostic factors by using the “forestplot” package for R software.

Construction of nomogram

We constructed a nomogram to quantitatively figure out individuals’ survival probability by incorporating multiple risk factors, including age, stage T, N, M and prognostic signatures of SASEs. According to risk contribution to survival probability, each risk factor was set as a specific point. To assess the consistency between actual and predicted survival time, calibration curves were performed for 3-year and 5-year survival. Besides, concordance index, that is C-index, was calculated to evaluate whether the nomogram model had excellent performance for predicting survival time[29].

Construction of the potential correlation network

The data of splicing factors was extracted from the SpliceAid2 database, including 404 splicing factors[30]. Pearson correlation analysis was carried out to discovery the relationship between PSI value of SASEs and expression level of SFs. Correlation coefficient greater than 0.400 and P value less than

0.05 were considered to be statistically significant. Cytoscape 3.6.0 was used to build the potential correlation network[31].

Correlation of SASEs and KEGG pathways

We performed Gene Set Variation Analysis (GSVA) to screen out the differential Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in CESC[32]. Univariable cox regression analysis was utilized to identify the prognostic KEGG pathways related to overall survival. Pearson correlation analysis was carried out to compute the correlation coefficient between SASEs and prognostic KEGG pathways.

Validation of the effect of EIF3A on CESC

CESC cell lines and cell culture

SiHa and Hela cells, human CESC cell lines, were cultured in Dulbecco's Modified Eagle's Medium (DMEM, GIBCO, USA) supplemented with 10% fetal bovine serum (FBS, GIBCO, USA) at 37 °C in a humidified atmosphere with 5% CO_2.

Small interfering RNA transfection

The small interfering RNA (siRNA) sequences were designed and purchased from RIBOBIO. Scramble siRNA control, siRNA-EIF3A-1, siRNA-EIF3A-2 and siRNA-EIF3A-3 were transfected into SiHa and Hela cells using Lipofectamine 2000 Reagent (Invitrogen, Carlsbad, California, USA) according to the manufacturer's instructions. Target sequences for transient silencing were as follows: EIF3A siRNA1: GGATGATGATCGCCTTTCA, EIF3A siRNA2: GAACGCCGATGATGACAGA, EIF3A siRNA3: GTACGGCAATCAATCTTAA. Cells were harvested at 48 hours for subsequent studies.

Cell proliferation assay

24 hours after siRNA (scramble siRNA, or siRNA-EIF3A-1, or siRNA-EIF3A-2, or siRNA-EIF3A-3) transfection, cells were plated in 5 replicates in a 96-well plate (3000 cells/well) and cultured in DMEM supplemented with 10% FBS. Then, cells were incubated with 20 uL CCK8 reagent (Yeasen, China) for 2 hours at 37 °C as instructed by the manufacturer. Viable cells were counted after another 0, 24, 48, or 72 hours by reading the absorbance at 450 nm with a plate reader. The "0 hour" was the time when all seeded cells adhered to the plate.

Wound healing assay

Cells were seeded in a 6-well plate and cultured in a complete medium until reaching 90% confluence. Confluent monolayer cells were linearly scratched using 10 ul pipette tip, and then incubated for 24 hours

in serum-free medium. The scratched region was photographed immediately or 24 hours after scratching using an inversion phase contrast microscope.

Statistical analysis

All statistical analyses were performed using R software version 3.4.1. Univariable and multivariable cox proportional hazards regression method were used to verify the association between alternative mRNA splicing events/prognostic signatures and overall survival. The least absolute shrinkage and selection operator (LASSO) regression analysis was applied to avoid prognostic signature overfitting. The receiver operating characteristic (ROC) curve was applied to assess the sensitivity and specificity of prognostic prediction of risk score. Pearson correlation analysis was performed to calculate the correlation coefficients between SFs/KEGG pathways and SASEs. The scratch width from per experiment was calculated, and data were then obtained and expressed as mean ± standard deviation. A two-tailed $P < 0.05$ was considered to be statistically significant.

Results

Identification of survival-associated alternative mRNA splicing events in CESC

The data of clinical characteristics and AS events of 216 CESC patients were analyzed. The clinical characteristics are listed in Table 1. In total, 41776 AS events in 9961 genes were detected, including 209 MEs in 202 genes, 2723 RIs in 1800 genes, 3017 ADs in 2106 genes, 3424 AAs in 2398 genes, 8066 APs in 3258 genes, 8395 ATs in 3664 genes and 15942 ESs in 6278 genes. The intersection between AS events related genes and alternative splicing event patterns in 216 CESC patients are displayed in Upset plot. AS shown in Fig. 1A, ES was the most common AS pattern and ME was the least common AS pattern in CESC, respectively. Besides, more than half of genes comprised ES events. In order to screen out survival associated alternative splicing events, univariable cox regression analysis was performed. Identified SASEs are illustrated in Upset plot (Fig. 1B). The most significant top 10 or 20 SASEs in each splicing pattern was selected and visualized in each bubble plot (Fig. 2).

Table 1
Baseline Characteristics of 261 CESC patients available from the TCGA database.

Characteristics	Number of patients (%)
Age(year)	
<60	208(79.7)
≥60	53(20.3)
Gender	
Male	0
Female	261(100)
Vital Status	
Death	67(25.7)
Alive	194(74.3)
Overall survival (day)	
≥365	224(85.8)
<365	37(14.2)
Race	
white	191(73.2)
not reported	26(10.0)
black or African American	16 (6.1)
Asian	20 (7.7)
American Indian or Alaska native	8 (3.1)
CESC, Cervical squamous cell cancer.	

Construction of prognostic signature for CESC

In order to construct an optimal prognostic signature, we integrated these most significant candidate SASEs into LASSO cox regression analysis, and calculated regression coefficients. After excluding AS events which could make prognostic signature overfitting, we obtained 106 AS events significantly correlated with survival. Multivariable cox regression analysis was carried out to acquire 70 SASEs that could be an independent prognostic predictor for overall survival in CESC (Table 2). Then, we separately calculated risk score of each patient for overall survival prediction in each AS pattern as well as in all AS pattern. CESC patients were divided into low-risk and high-risk groups by using the median risk score as a

cutoff value. The median risk score was set as 0.951 in AA pattern, 0.850 in AD pattern, 0.842 in AP pattern, 1.026 in AT pattern, 0.944 in ES pattern, 0.957 in ME pattern, 0.937 in RI pattern, and 0.809 in all pattern.

Table 2
Details of survival-associated alternative splicing events (SASEs) used for constructing prognostic signatures.

Type	ID	Coef	HR	95%CI-L	95%CI-H	P value
AA	PRR13|22038|AA	-7.75385	0.000429	3.22E-08	5.71975	0.109579
	MAN2A2|32517|AA	-4.30426	0.013511	0.001023	0.178416	0.001079
	HNRNPA2B1|79035|AA	4.482	88.41136	6.060859	1289.68	0.001047
	RBM7|18824|AA	-4.94462	0.007122	0.000107	0.472926	0.020902
	GATAD2A|48635|AA	-15.4531	1.94E-07	4.40E-11	0.00086	0.000308
	ZFP64|59815|AA	6.476853	649.9222	16.90409	24987.98	0.000504
	ARRB2|38562|AA	-17.5908	2.29E-08	2.60E-16	2.02539	0.059516
	BUB3|13390|AA	3.508728	33.40574	2.050064	544.3459	0.013735
	ENTPD6|58863|AA	-6.06282	0.002328	1.18E-05	0.457421	0.024432
	NCALD|84756|AA	-8.73199	0.000161	1.33E-07	0.195992	0.015966
	PIF1|31133|AA	-8.29642	0.000249	7.70E-08	0.807966	0.044255
	MZF1|52489|AA	-19.7646	2.61E-09	1.64E-16	0.041533	0.019494
	COPS3|39475|AA	-9.22165	9.89E-05	3.56E-08	0.274584	0.02264
AD	FCF1|28425|AD	3.420082	30.57192	1.234163	757.309	0.036758
	CLIP1|24953|AD	-15.765	1.42E-07	1.07E-13	0.188584	0.028386
	RPS15A|34266|AD	12.41799	247209.6	407.4742	1.5E + 08	0.000146
	YDJC|61233|AD	-5.73088	0.003244	5.06E-05	0.207926	0.006937
	HNRNPAB|74850|AD	-10.7503	2.14E-05	3.63E-10	1.267813	0.055157
	NUP62|51129|AD	-13.2582	1.75E-06	1.77E-09	0.001725	0.000164
	NDUFAF2|72172|AD	6.508998	671.1534	5.590973	80566.81	0.007709
	UQCRQ|73319|AD	2.881056	17.83309	2.563357	124.0635	0.003602
	STARD10|17645|AD	-37.8737	3.56E-17	2.59E-27	4.90E-07	0.001475
	CASP1|18527|AD	-10.0739	4.22E-05	7.13E-09	0.249448	0.023008
	SLC38A1|21328|AD	3.401127	29.9979	3.243702	277.4218	0.002728

SASEs, Survival associated splicing events; AA, Alternate acceptor site; AD, Alternate donor site; AP, Alternate promoter; AT, Alternate terminator; ES, Exon skip; ME, Mutually exclusive exons; RI, Retained intron; Coef, Coefficient; HR, Hazard ratio; CI, Confidence interval; P-value less than 0.05 was considered statistically significant.

Type	ID	Coef	HR	95%CI-L	95%CI-H	P value
AP	FOXRED2|62052|AP	2.83723	17.06843	1.264804	230.337	0.032607
	OPA3|50486|AP	-15.1316	2.68E-07	5.31E-10	0.000136	1.90E-06
	C1QTNF1|43985|AP	1.888378	6.608639	0.70822	61.66745	0.097479
	SERPING1|15866|AP	-3.3398	0.035444	0.008369	0.150114	5.76E-06
	SHF|30409|AP	1.21478	3.369553	0.910234	12.47359	0.068893
	ATF3|9733|AP	-19.9397	2.19E-09	1.09E-12	4.41E-06	2.80E-07
	ECE1|963|AP	2.447984	11.56501	0.459903	290.8208	0.136786
AT	ACSS1|58860|AT	-7.40178	0.00061	2.45E-06	0.152089	0.008568
	TNNI1|9376|AT	-11.5468	9.67E-06	4.46E-09	0.020965	0.003218
	POLR2L|13788|AT	5.958918	387.1908	4.990055	30043.11	0.007275
	RNF157|43574|AT	9.730878	16829.33	5.824021	48630704	0.016696
	C1orf86|246|AT	3.870046	47.9446	0.960826	2392.404	0.052388
	IGF1|24051|AT	2.626872	13.83043	2.307043	82.91174	0.004042
	CD59|14907|AT	-33.1201	4.13E-15	7.73E-23	2.21E-07	0.000264
	PTCHD4|76445|AT	3.879451	48.39763	6.389682	366.5801	0.000173
	ZCCHC4|68960|AT	-5.55735	0.003859	1.54E-05	0.9701	0.048755

SASEs, Survival associated splicing events; AA, Alternate acceptor site; AD, Alternate donor site; AP, Alternate promoter; AT, Alternate terminator; ES, Exon skip; ME, Mutually exclusive exons; RI, Retained intron; Coef, Coefficient; HR, Hazard ratio; CI, Confidence interval; P-value less than 0.05 was considered statistically significant.

Table 2
Details of survival-associated alternative splicing events (SASEs) used for constructing prognostic signatures. (To be continued)

Type	ID	Coef	HR	95%CI-L	95%CI-H	P value
ES	NDUFA3|51782|ES	4.482857	88.4871	8.254286	948.5942	0.000212
	HNRNPA1|301521|ES	13.23406	559086.2	3249.039	96206099	4.69E-07
	NME2|42514|ES	-5.18488	0.005601	0.000103	0.30538	0.011042
	EBPL|25914|ES	4.216051	67.76536	8.718308	526.7242	5.59E-05
	FBXO18|10672|ES	10.09561	24236.34	1.415854	4.15E + 08	0.042369
	SSBP2|72672|ES	-5.97425	0.002543	1.55E-06	4.162479	0.11359
	CDK11A|216|ES	10.59941	40111.01	0.061792	2.6E + 10	0.120601
	FANCL|53655|ES	-10.7231	2.20E-05	3.15E-08	0.015403	0.001333
	DMKN|49154|ES	-18.005	1.52E-08	8.23E-15	0.027888	0.014433
	TIMM8B|18730|ES	5.97006	391.5293	7.696127	19918.49	0.002903
	TATDN1|85097|ES	-5.7355	0.003229	8.93E-05	0.116758	0.001729
ME	IL1RN|95654|ME	1.527551	4.606882	1.084439	19.57082	0.038471
	P4HA1|12122|ME	-3.39087	0.033679	0.001065	1.065083	0.054331
	FYN|77273|ME	1.631022	5.109095	1.300908	20.0651	0.019446
	C4orf21|70379|ME	26.43117	3.01E + 11	0.000163	5.55E + 26	0.140537
	UBAP1|86148|ME	-2.40127	0.090603	0.012501	0.656644	0.017493
	PSTPIP1|114371|ME	9.20808	9977.417	0.013465	7.39E + 09	0.18178
	GOLT1B|92984|ME	2.136293	8.467988	1.308857	54.78585	0.024929
	ZFAND6|32173|ME	-21.1046	6.83E-10	4.57E-18	0.102141	0.027984
	RAB6A|17707|ME	2.421756	11.26563	0.559718	226.7469	0.113857
RI	MRPL52|26642|RI	-6.59503	0.001367	9.34E-05	0.020019	1.46E-06
	HNRNPLL|53264|RI	3.698461	40.38508	6.453647	252.7183	7.72E-05
	NMRAL1|33738|RI	3.39307	29.75718	2.834631	312.3827	0.004676

SASEs, Survival associated splicing events; AA, Alternate acceptor site; AD, Alternate donor site; AP, Alternate promoter; AT, Alternate terminator; ES, Exon skip; ME, Mutually exclusive exons; RI, Retained intron; Coef, Coefficient; HR, Hazard ratio; CI, Confidence interval; P-value less than 0.05 was considered statistically significant.

Type	ID	Coef	HR	95%CI-L	95%CI-H	P value
	MAX|27937|RI	-1.89627	0.150128	0.026154	0.861765	0.033435
	CNTNAP3|86469|RI	-2.24474	0.105955	0.028382	0.395548	0.000838
	NEIL2|82632|RI	3.894484	49.13068	6.430176	375.3899	0.000174
	GLYCTK|65200|RI	-4.32843	0.013188	0.001526	0.113975	8.37E-05
	NFE2L2|56132|RI	-3.49627	0.03031	0.001608	0.571402	0.019622
	CCDC74B|55280|RI	2.228842	9.289105	2.773712	31.10902	0.000301
	CYP4F12|48111|RI	2.612908	13.63866	3.843853	48.39232	5.26E-05
All	CLIP1|24953|AD	-32.8198	5.58E-15	3.07E-21	1.01E-08	8.08E-06
	FOXRED2|62052|AP	2.42141	11.26172	0.618434	205.0768	0.101966
	NDUFA3|51782|ES	4.55702	95.29911	6.817313	1332.185	0.000708
	OPA3|50486|AP	-18.0556	1.44E-08	1.81E-11	1.15E-05	1.18E-07
	MAN2A2|32517|AA	-3.36765	0.034471	0.00319	0.372523	0.005553
	RPS15A|34266|AD	8.577861	5312.729	5.732811	4923430	0.013857
	SERPING1|15866|AP	-1.89224	0.150734	0.029899	0.759923	0.021872
	SHF|30409|AP	0.973205	2.646412	0.685181	10.22138	0.158071
	NHLRC3|25701|ES	-4.94767	0.0071	0.000249	0.202215	0.003787

SASEs, Survival associated splicing events; AA, Alternate acceptor site; AD, Alternate donor site; AP, Alternate promoter; AT, Alternate terminator; ES, Exon skip; ME, Mutually exclusive exons; RI, Retained intron; Coef, Coefficient; HR, Hazard ratio; CI, Confidence interval; P-value less than 0.05 was considered statistically significant.

Assessment of the clinical value of prognostic signature

To validate the predictive efficiency of prognostic signature, K-M survival analysis and ROC curve were conducted. K-M survival analyses showed that the difference was statistically significant between low-risk and high-risk groups for overall survival (P<0.01) (Fig. 3A-H). As shown in Fig. 3I-P, prognostic signatures of each AS pattern and all AS pattern had considerable predictive efficacy in differentiating better or worse outcomes for CESC patients. Prognostic signatures for AA (AUC = 0.951) and RI (AUC = 0.895) had optimal performance for predicting overall survival, and the area under ROC curve (AUC) for all AS pattern was 0.932. The correlation between risk score and overall survival was displayed in scatter plot and risk plot for 216 CESC samples. The red and green dots presented high-risk and low-risk group in risk plot. While in the scatterplot, these dots indicated survival status and overall survival of CESC patients, divided by median risk score of each AS pattern or all AS pattern. Expression level of SASEs in each AS pattern was visualized in the heatmap, with the red and green stripe indicating high and low PSI

values, and the red and blue bars presenting low-risk and high-risk groups (Fig. 4A-H). Next, univariable and multivariable cox proportional hazards regression analyses were used to confirm whether prognostic signatures were independent prognostic factors in each AS pattern and all AS pattern. After integrating age, grade and cancer status, multivariate cox regression analyses confirmed that prognostic signatures for risk score were exactly independent prognostic factors for CESC patients (P<0.01). The aforementioned results were shown in forest map (Fig. 5A-H).

Nomogram

As we all known, age and TNM stage are independent prognostic factors in CESC. In order to obtain individuals' survival probability, we performed a comprehensive analysis of multiple risk factors. We constructed a nomogram to quantitatively predict the 3-year and 5-year survival by integrating age, stage T, N, M and prognostic signature of SASEs. According to risk contribution to survival probability, each risk factor was set as a specific point. As shown in Fig. 6A, age less than 50 years old was assigned 68.75 points, stage T1-T2 was assigned 0 points, stage N1 was assigned 7.50 points, stage M0 was assigned 0 points, high-risk score was assigned 71.25 points and so on. Calibration curves showed that there was excellent consistency between actual and predicted survival time (Fig. 6B-C). The concordance index (C-index) was 0.82, indicating that the model had an excellent performance for predicting overall survival in CESC. For example, a 45-year old patient with high-risk score was in stage_T1-T2, stage_Nx, stage_M0 of cervical squamous cell cancer, she would get 147.5 points. Her 3-year survival rate was 65%, and 5-year survival rate was 50%.

SASEs and SFs correlation network in CESC

To explore the upstream regulators of SASEs, we carried out the Pearson correlation analysis to discovery the relationship between SASEs and SFs. The correlation network between 116 SASEs and 15 SFs were illustrated in Fig. 7. The purple arrows signified SFs, red and green circles represented high and low hazard ratio (HR) of SASEs by univariable cox regression analysis, as well as red and green lines indicated positive and negative regulation between SASEs and SFs. Our results showed that EIF3A had a positive regulation of SEC23A-27346-AP with highest correlation coefficient.

Correlation of SASEs and KEGG pathways

To explore the downstream pathways of SASEs, we screened out 185 differential KEGG pathways in CESC by using Gene Set Variation Analysis (GSVA). Univariable cox regression analysis was applied to identify 16 prognostic KEGG pathways related to overall survival. Then, the correlation coefficient of SASEs and prognostic KEGG pathways was calculated by using Pearson correlation analysis. The correlation between SASEs and prognostic KEGG pathways was illustrated in heatmap (Fig. 8). Red and blue dots represented positive and negative correlation between SASEs and prognostic KEGG pathways. The most significant SASEs and prognostic KEGG pathways were SEC23A-27346-AP and KEGG_ADHERENS_JUNCTION (P = 0.02, R = 0.65). Combining the correlation network between SASEs and SFs, EIF3A was the most significant SF positively related to SEC23A-27346-AP (P<0.001, R = 0.69). As a

consequence, the most significant SF, SASEs, and downstream pathway were EIF3A, SEC23A-27346-AP and adherens junction pathway, respectively.

Knockdown of EIF3A decreases CESC cell proliferation and migration.

In order to verify the regulatory relationships between downregulated EIF3A level and CESC cell proliferation and migration, small interfering RNA (siRNA) transfection was performed. After confirming their successful establishment, these siRNAs were applied to a series of functional experiments. Knockdown of EIF3A significantly decrease cell proliferation of Hela viable cells (Fig. 9A). When dropping EIF3A levels, cell proliferation of SiHa viable cells reduced (Fig. 9B). However, Hela cell migration significantly decreased, after EIF3A expression levels was downregulated (Fig. 9C-D). Similarly, knockdown of EIF3A markedly decrease cell migration of SiHa cells (Fig. 9E-F).

Discussion

CESC is one of the frequently detected malignancies and pose a health threat to women all over the world[1–3]. Clinical stage, pathologic type, response to therapy and long-term prognosis can vary considerably among CESC patients. It is important for CESC patients to obtain survival time individually. Research suggested that biomarkers, such as alternative mRNA splicing, might be an effective prognostic predictor[6]. AS is a vital biological process that can expand protein species diversity[9]. There is an increasing number of studies that AS plays a crucial role in the oncogenesis and development of cancer. Previous studies reported that AS may accurately evaluate prognostic outcomes in some cancers, including glioblastoma multiforme, colorectal cancer, prostate adenocarcinoma, bladder urothelial carcinoma, gastric cancer and so on[23, 24]. However, the relevant research of cervical squamous cell cancer is scarce.

In the present study, we constructed AS-associated prognostic signatures. AS-associated prognostic signatures were exactly independent prognostic factors in CESC. And prognostic signatures had high AUC values, proving prognostic signatures had reliable predictive efficacy. The findings highlight the clinical significance of AS-associated prognostic signatures in CESC. After integrating age, stage T, N, M and AS-associated prognostic signatures, we carried out a comprehensive analysis, that is nomogram, to quantitatively predict individuals' survival probability. Compared with histopathological type or TNM stage, it had a more excellent performance in predicting survival time in CESC. These findings suggest that AS-associated prognostic signatures may eventually act as efficient prognostic biomarkers to assess survival time in CESC. This is the first and the most comprehensive analysis to construct a prognostic prediction model, which incorporate age, TNM stage and AS-associated prognostic signatures, in cervical squamous cell cancer.

Alternative splicing events are thought to be accurately regulated by hundreds of splicing factors, which specifically combine with the splicing sites and participate in subsequent splicing events. There is no doubt that aberrant splicing factors or abnormal expression of splicing factors directly leads to changes of splicing events pattern and splicing events expression[14, 15]. Some studies reported that there was correlation between splicing factors and poor prognosis in chronic lymphocytic leukemia, lung adenocarcinoma, pancreatic cancer, breast cancer and melanoma[33–35]. In our study, splicing factors were collected from the SpliceAid2 database, and correlation network between SASEs and SFs was explored. The most significant SASEs and SFs were SEC23A-27346-AP and EIF3A with the maximum correlation coefficient. EIF3A had a strongly positive regulation with SEC23A-27346-AP. Our findings revealed that EIF3A might be the upstream regulator of SEC23A-27346-AP. The in-depth correlation analysis of regulation network between SESAs and SFs provides a new insight into the mechanism of AS involved in the unfavorable prognosis of cervical squamous cell cancer.

Alternative splicing events not only can be regulated by splicing factors, but also can respond to multiple signaling pathways. As we all know, weakening or lacking tight cell-cell adherens junction has been considered to be a fatal hallmark of cancer[36, 37]. In normal physiological conditions, cell-cell adherens junction is the foundation of tissues architecture[38]. However, tumor cells tend to loss of adherens junctions, and display local dissemination and distant metastasis especially in advanced cancers. Cadherins and catenins are the core proteins of cell-cell adherens junction. Low or loss of expression of cadherins and catenins inevitably lead to tumor progression and unfavorable prognosis[39]. In the current study, we selected prognostic KEGG pathways in CESC by Gene Set Variation Analysis (GSVA). The correlation analysis between SASEs and prognostic KEGG pathways was carried out. The most significant SASEs and prognostic KEGG pathways were SEC23A-27346-AP and adherens junction pathway with the maximum correlation coefficient. Therefore, we supposed that adherens junction pathway might be one of the downstream pathways of SEC23A-27346-AP in CESC.

SEC23A, as a core component of coat protein complex Ⅱ vesicles, can not only transport proteins but also cause human diseases. Aberrant SEC23A or abnormal expression of SEC23A was reported to be relevant to the development and progression of human cancer[40]. A previous study implicated that RNA binding motif 5 (RBM5) participate in regulating mRNA splicing of SEC23A[41]. However, our results indicated that SEC23A-27346-AP was associated with poor prognosis in CESC patients, and splicing factor EIF3A positively regulated SEC23A-27346-AP. Above all, SEC23A regulated by EIF3A may facilitate the development of cervical squamous cell cancer via adherens junction pathway.

EIF3A is the largest subunit of eukaryotic initiation factor 3 (EIF3), which plays a crucial role in all steps of translation initiation. Numbers of studies have shown that EIF3A exerts an important influence on the regulation of cell cycle, cell growth and differentiation [42]. Recent studies have found that EIF3a is correlated with carcinogenesis, development and chemotherapy efficacy[43, 44]. However, it is unknown whether EIF3A is a proto-oncogene in CESC. In the study, our results indicated that EIF3A knockdown decrease cell proliferation and cell migration of CESC. Therefore, EIF3A can promote proliferation and migration of CESC.

There are some limitations in this study. Firstly, the number of CESC patients in TCGA database was relatively small. Clinical information of 308 CESC patients was extracted from the TCGA website. After excluding the patients who didn't accord with the inclusive criteria, a total of 216 CESC patients were enrolled in our study. Secondly, we performed a nomogram to more precisely obtain individuals' survival probability by integrating multiple risk factors. However, we failed to take into consideration other risk factors, such as pathological grade, therapeutic strategies and so on. Moreover, the mechanism of alternative splicing events in the development of CESC was unclear. We carried out correlation analysis to explore upstream regulators and downstream pathways of alternative splicing events. In the meanwhile, we merely validated the relationship between EIF3A and CESC proliferation and migration. In the follow-up study, the molecular mechanisms of SEC23A and adherens junction pathway in the occurrence and development of cervical squamous cell cancer need further investigations to verify.

Conclusion

We constructed a splicing event-associated prognostic signature that could predict overall survival in cervical squamous cell carcinoma, and integrated multiple risk factors, including splicing event-associated prognostic signature and clinical characteristics, to predict the individualized survival probability. Moreover, the potential interaction network between splicing events and splicing factors as well as downstream regulatory mechanisms in CESC were provided. SEC23A positively regulated by EIF3A may contribute to cervical squamous cell cancer via adherens junction pathway. This study may offer new direction for subsequent experimental research in cervical squamous cell cancer.

Abbreviations

AA: Alternate acceptor site; AD: Alternate donor site; AP, Alternate promotor; AS: Alternative mRNA splicing; AT: Alternate terminator; AUC: The area under ROC curve; CESC: Cervical squamous cell cancer; EIF3: Eukaryotic initiation factor 3; ES: Exon skip; GSVA: Gene Set Variation Analysis; HR: Hazard ratio; KEGG: Kyoto Encyclopedia of Genes and Genomes; LASSO: The least absolute shrinkage and selection operator; ME: Mutually exclusive exons; PSI: Percent-spliced-in; RI: Retained intron; ROC: Receiver operating characteristic (ROC) curve; SASEs: Survival-associated splicing events; SF: Splicing factors.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The study was approved by uploading the raw data onto the research data deposit (RDD) public platform. You will find more detailed information in the RDD public platform (www.researchdata.org.cn).

Competing Interests

The authors have declared that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (grant # 81870878) and Guang Dong Natural Science Foundation (2018A030313092).

Author Contributions

HX and JDX designed the study. CL and FY conducted the study and wrote the original draft. TW and HFG helped data analysis and interpretation of the results. GC and QL helped data collection. All authors have read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Fader AN: **Surgery in Cervical Cancer**. *The New England journal of medicine* 2018, **379**(20):1955-1957.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries**. *CA: a cancer journal for clinicians* 2018, **68**(6):394-424.
3. Vu M, Yu J, Awolude OA, Chuang L: **Cervical cancer worldwide**. *Current problems in cancer* 2018, **42**(5):457-465.
4. Naga Ch P, Gurram L, Chopra S, Mahantshetty U: **The management of locally advanced cervical cancer**. *Current opinion in oncology* 2018, **30**(5):323-329.
5. Park KJ, Braschi-Amirfarzan M, DiPiro PJ, Giardino AA, Jagannathan JP, Howard SA, Shinagare AB, Krajewski KM: **Multimodality imaging of locally recurrent and metastatic cervical cancer: emphasis on histology, prognosis, and management**. *Abdominal radiology (New York)* 2016, **41**(12):2496-2508.

6. Gao C, Zhou C, Zhuang J, Liu L, Liu C, Li H, Liu G, Wei J, Sun C: **MicroRNA expression in cervical cancer: Novel diagnostic and prognostic biomarkers**. *Journal of cellular biochemistry* 2018, **119**(8):7080-7090.
7. Aalijahan H, Ghorbian S: **Long non-coding RNAs and cervical cancer**. *Experimental and molecular pathology* 2019, **106**:7-16.
8. Liu S, Meng F, Ding J, Ji H, Lin M, Zhu J, Ma R: **High TRIM44 expression as a valuable biomarker for diagnosis and prognosis in cervical cancer**. *Bioscience reports* 2019, **39**(3).
9. Ule J, Blencowe BJ: **Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution**. *Molecular cell* 2019, **76**(2):329-345.
10. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB *et al*: **Predicting Splicing from Primary Sequence with Deep Learning**. *Cell* 2019, **176**(3):535-548.e524.
11. Baralle FE, Giudice J: **Alternative splicing as a regulator of development and tissue identity**. *Nature reviews Molecular cell biology* 2017, **18**(7):437-451.
12. Furlanis E, Scheiffele P: **Regulation of Neuronal Differentiation, Function, and Plasticity by Alternative Splicing**. *Annual review of cell and developmental biology* 2018, **34**:451-469.
13. Lin JC, Tsao MF, Lin YJ: **Differential Impacts of Alternative Splicing Networks on Apoptosis**. *International journal of molecular sciences* 2016, **17**(12).
14. Lee Y, Rio DC: **Mechanisms and Regulation of Alternative Pre-mRNA Splicing**. *Annual review of biochemistry* 2015, **84**:291-323.
15. Yoshida K, Ogawa S: **Splicing factor mutations and cancer**. *Wiley interdisciplinary reviews RNA* 2014, **5**(4):445-459.
16. Climente-Gonzalez H, Porta-Pardo E, Godzik A, Eyras E: **The Functional Impact of Alternative Splicing in Cancer**. *Cell reports* 2017, **20**(9):2215-2226.
17. Bowler E, Oltean S: **Alternative Splicing in Angiogenesis**. *International journal of molecular sciences* 2019, **20**(9).
18. Marzese DM, Manughian-Peter AO, Orozco JIJ, Hoon DSB: **Alternative splicing and cancer metastasis: prognostic and therapeutic applications**. *Clinical & experimental metastasis* 2018, **35**(5-6):393-402.
19. Oltean S, Bates DO: **Hallmarks of alternative splicing in cancer**. *Oncogene* 2014, **33**(46):5311-5318.
20. Xiping Z, Qingshan W, Shuai Z, Hongjian Y, Xiaowen D: **A summary of relationships between alternative splicing and breast cancer**. *Oncotarget* 2017, **8**(31):51986-51993.
21. Shi Y, Chen Z, Gao J, Wu S, Gao H, Feng G: **Transcriptome-wide analysis of alternative mRNA splicing signature in the diagnosis and prognosis of stomach adenocarcinoma**. *Oncology reports* 2018, **40**(4):2014-2022.
22. Liu J, Li H, Shen S, Sun L, Yuan Y, Xing C: **Alternative splicing events implicated in carcinogenesis and prognosis of colorectal cancer**. *Journal of Cancer* 2018, **9**(10):1754-1764.

23. Huang ZG, He RQ, Mo ZN: **Prognostic value and potential function of splicing events in prostate adenocarcinoma**. *International journal of oncology* 2018, **53**(6):2473-2487.
24. Hong W, Zhang W, Guan R, Liang Y, Hu S, Ji Y, Liu M, Lu H, Yu M, Ma L: **Genome-wide profiling of prognosis-related alternative splicing signatures in sarcoma**. *Annals of translational medicine* 2019, **7**(20):557.
25. He RQ, Zhou XG, Yi QY, Deng CW, Gao JM, Chen G, Wang QY: **Prognostic Signature of Alternative Splicing Events in Bladder Urothelial Carcinoma Based on Spliceseq Data from 317 Cases**. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* 2018, **48**(3):1355-1368.
26. Ryan M, Wong WC, Brown R, Akbani R, Su X, Broom B, Melott J, Weinstein J: **TCGASpliceSeq a compendium of alternative mRNA splicing in cancer**. *Nucleic acids research* 2016, **44**(D1):D1018-1022.
27. Shen Y, Zhu H, Yang Y, Shen Y: **A sparse loudspeaker array for surround sound reproduction using the least absolute shrinkage and selection operator algorithm**. *The Journal of the Acoustical Society of America* 2019, **145**(5):El430.
28. Schafer S, Miao K, Benson CC, Heinig M, Cook SA, Hubner N: **Alternative Splicing Signatures in RNA-seq Data: Percent Spliced in (PSI)**. *Current protocols in human genetics* 2015, **87**:11.16.11-11.16.14.
29. Park SY: **Nomogram: An analogue tool to deliver digital knowledge**. *The Journal of thoracic and cardiovascular surgery* 2018, **155**(4):1793.
30. Piva F, Giulietti M, Burini AB, Principato G: **SpliceAid 2: a database of human splicing factors expression data and RNA target motifs**. *Human mutation* 2012, **33**(1):81-85.
31. Su G, Morris JH, Demchak B, Bader GD: **Biological network exploration with Cytoscape 3**. *Current protocols in bioinformatics* 2014, **47**:8.13.11-24.
32. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes**. *Nucleic acids research* 2000, **28**(1):27-30.
33. Liao W, Jordaan G, Nham P, Phan RT, Pelegrini M, Sharma S: **Gene expression and splicing alterations analyzed by high throughput RNA sequencing of chronic lymphocytic leukemia specimens**. *BMC cancer* 2015, **15**:714.
34. Jia R, Ajiro M, Yu L, McCoy P, Jr., Zheng ZM: **Oncogenic splicing factor SRSF3 regulates ILF3 alternative splicing to promote cancer cell proliferation and transformation**. *RNA (New York, NY)* 2019, **25**(5):630-644.
35. Seiler M, Peng S, Agrawal AA, Palacino J, Teng T, Zhu P, Smith PG, Buonamici S, Yu L: **Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types**. *Cell reports* 2018, **23**(1):282-296.e284.
36. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation**. *Cell* 2011, **144**(5):646-674.
37. Gloushankova NA, Rubtsova SN, Zhitnyak IY: **Cadherin-mediated cell-cell interactions in normal and cancer cells**. *Tissue barriers* 2017, **5**(3):e1356900.

38. Mege RM, Ishiyama N: **Integration of Cadherin Adhesion and Cytoskeleton at Adherens Junctions**. *Cold Spring Harbor perspectives in biology* 2017, **9**(5).
39. Wong SHM, Fang CM, Chuah LH, Leong CO, Ngai SC: **E-cadherin: Its dysregulation in carcinogenesis and clinical implications**. *Critical reviews in oncology/hematology* 2018, **121**:11-22.
40. Jing J, Wang B, Liu P: **The Functional Role of SEC23 in Vesicle Transportation, Autophagy and Cancer**. *International journal of biological sciences* 2019, **15**(11):2419-2426.
41. Jackson TC, Kotermanski SE, Kochanek PM: **Whole-transcriptome microarray analysis reveals regulation of Rab4 by RBM5 in neurons**. *Neuroscience* 2017, **361**:93-107.
42. Dong Z, Liu Z, Cui P, Pincheira R, Yang Y, Liu J, Zhang JT: **Role of eIF3a in regulating cell cycle progression**. *Experimental cell research* 2009, **315**(11):1889-1894.
43. Yin Y, Long J, Sun Y, Li H, Jiang E, Zeng C, Zhu W: **The function and clinical significance of eIF3 in cancer**. *Gene* 2018, **673**:130-133.
44. Yin JY, Zhang JT, Zhang W, Zhou HH, Liu ZQ: **eIF3a: A new anticancer drug target in the eIF family**. *Cancer letters* 2018, **412**:81-87.

Figures

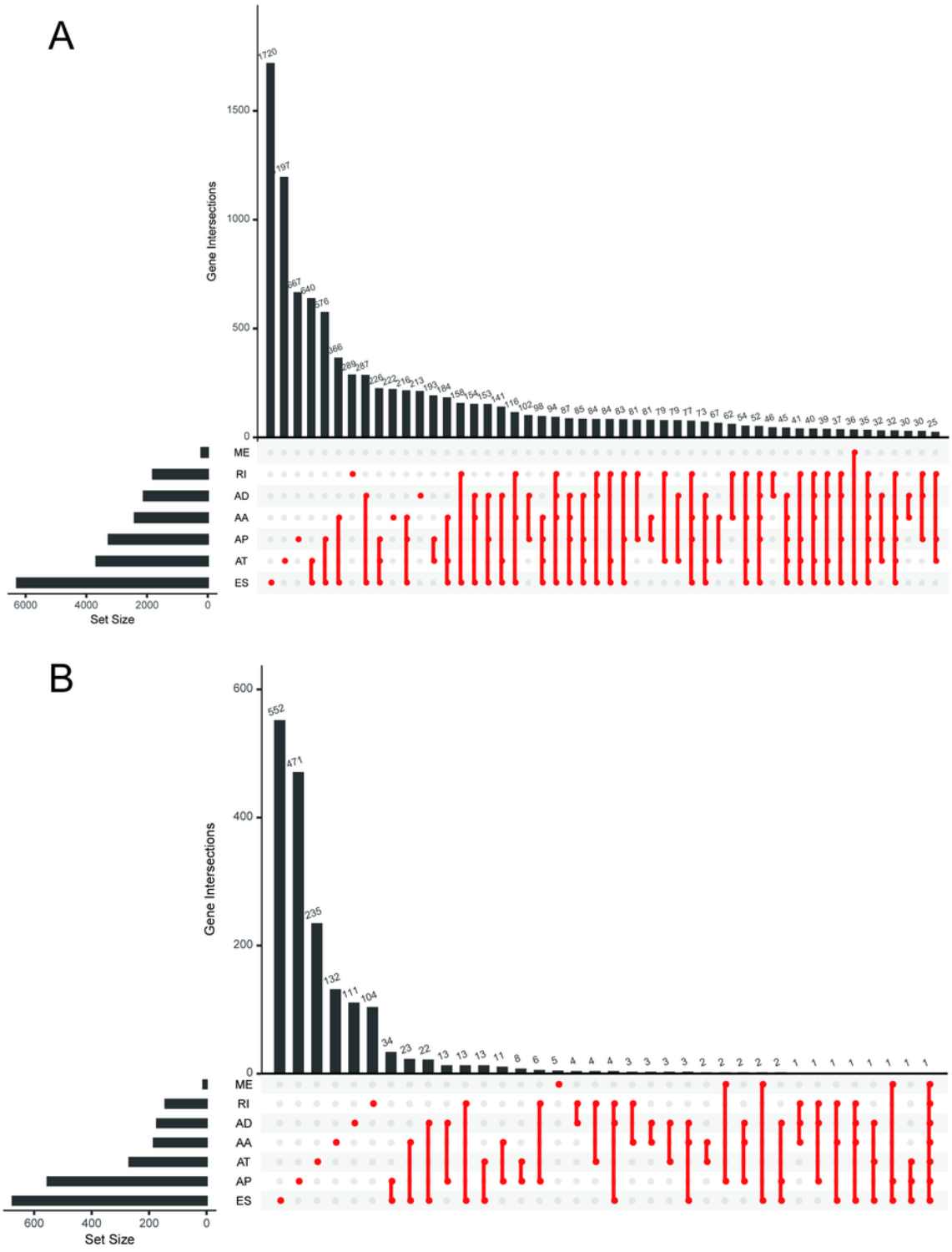

Figure 1

UpSet plots. UpSet plots of AS event patterns and genes in CESC (A) and SASEs patterns and genes in CESC (B). Horizontal axis represents AS patterns. Vertical axis represents the number of genes corresponding to certain AS event patterns. AS, Alternative splicing; SASEs, Survival associated splicing events; ME, Mutually exclusive exons; RI, Retained intron; AD, Alternate donor site; AA, Alternate acceptor site; AT, Alternate terminator; AP, Alternate promoter; ES, Exon skip.

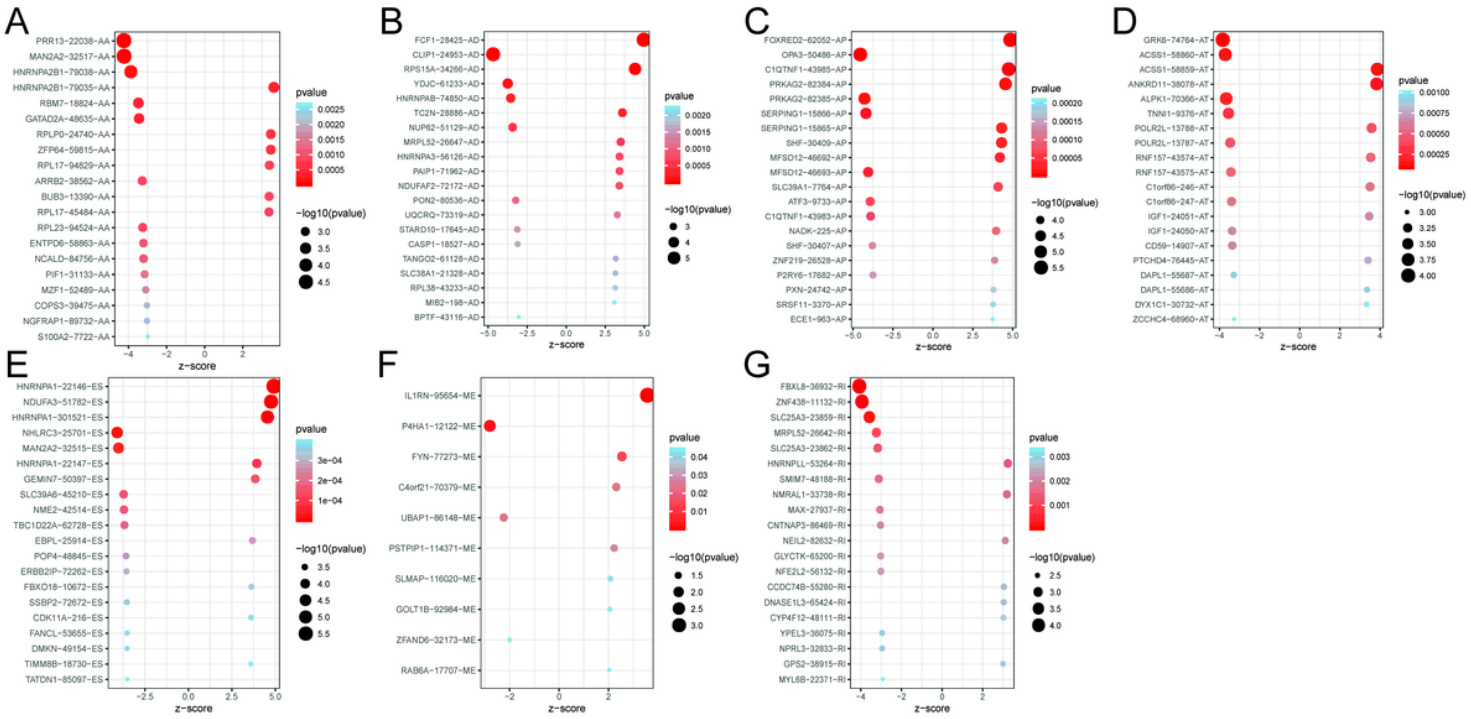

Figure 2

Bubble plots. Bubble plots of the most significant top 10 or 20 SASEs in each splicing pattern (A–G). SASEs, Survival associated splicing events; AA, Alternate acceptor site; AD, Alternate donor site; AP, Alternate promoter; AT, Alternate terminator; ES, Exon skip; ME, Mutually exclusive exons; RI, Retained intron.

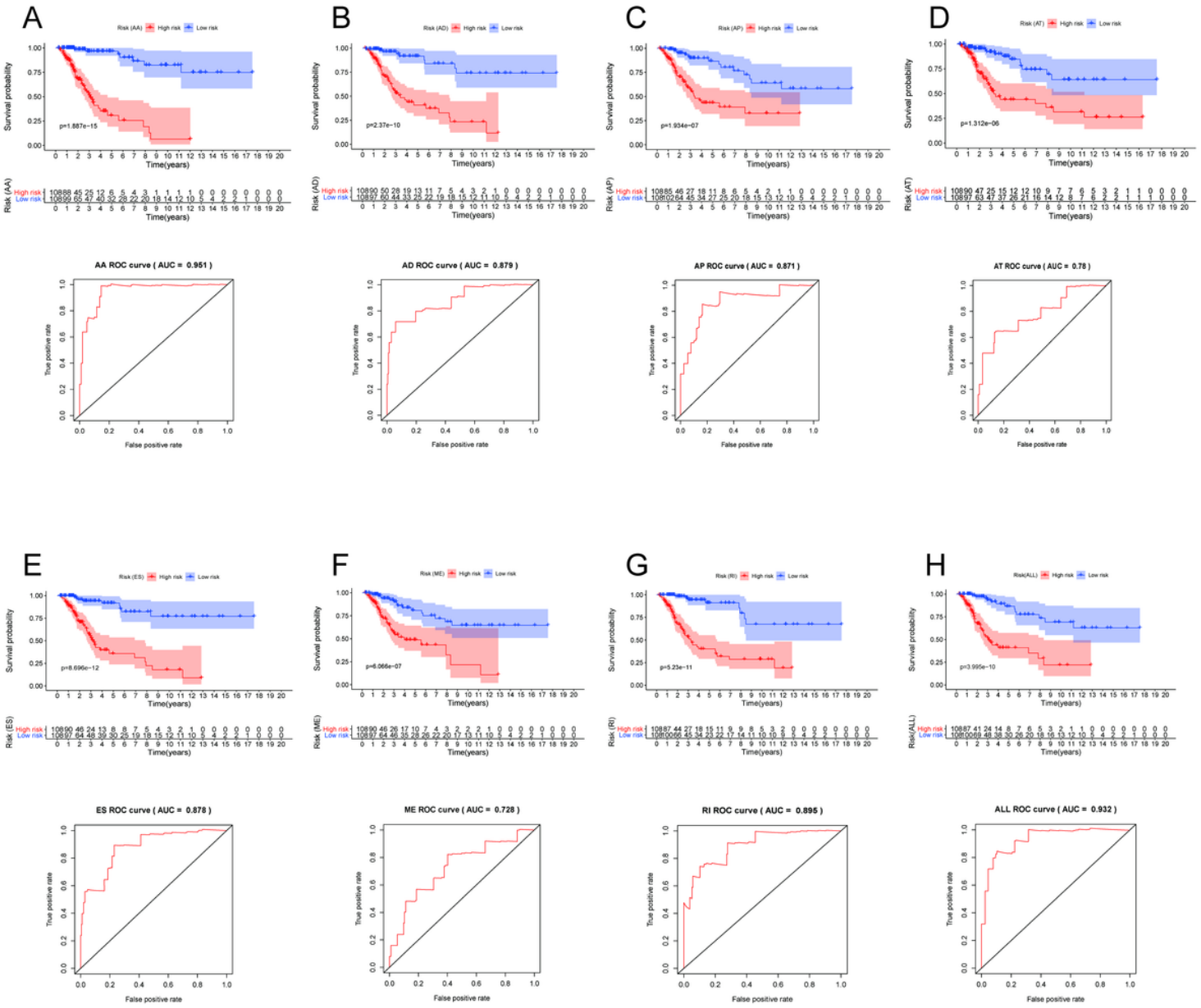

Figure 3

Kaplan-Meier survival curve. Kaplan-Meier survival curve of each AS pattern and all AS pattern in CESC patients, divided into low-risk and high-risk groups by using the median risk score as a cutoff value (A-H). ROC curve of each AS pattern and all AS pattern in CESC patients (I-P). AS, Alternative splicing; AA, Alternate acceptor site; AD, Alternate donor site; AP, Alternate promoter; AT, Alternate terminator; ES, Exon skip; ME, Mutually exclusive exons; RI, Retained intron.

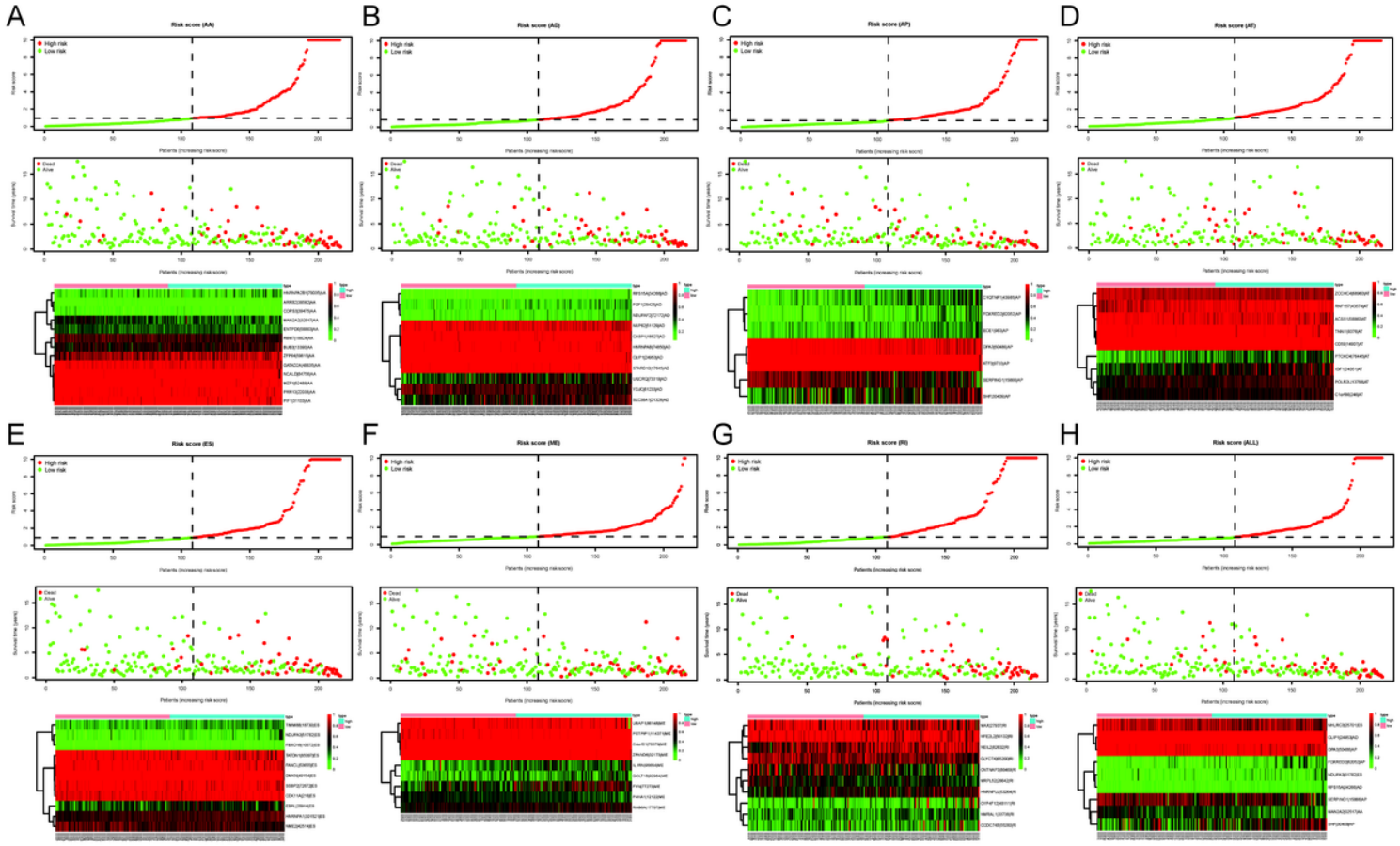

Figure 4

Risk plot and scatterplot. Risk plot and scatterplot of low-risk and high-risk groups for each CESC sample in AA, AD, AP, AT, ES, ME, RI and all AS pattern (A-H). The green and red dots in risk plot present low-risk and high-risk groups divided by median risk score. These dots in scatter plot indicate survival status and overall survival. Heatmap of survival associated splicing events expression level in each AS pattern and all AS pattern (A-H). Red and green stripe indicate high and low PSI values; Red and blue bars present low-risk and high-risk groups. AS, Alternative splicing; AA, Alternate acceptor site; AD, Alternate donor site; AP, Alternate promoter; AT, Alternate terminator; ES, Exon skip; ME, Mutually exclusive exons; RI, Retained intron.

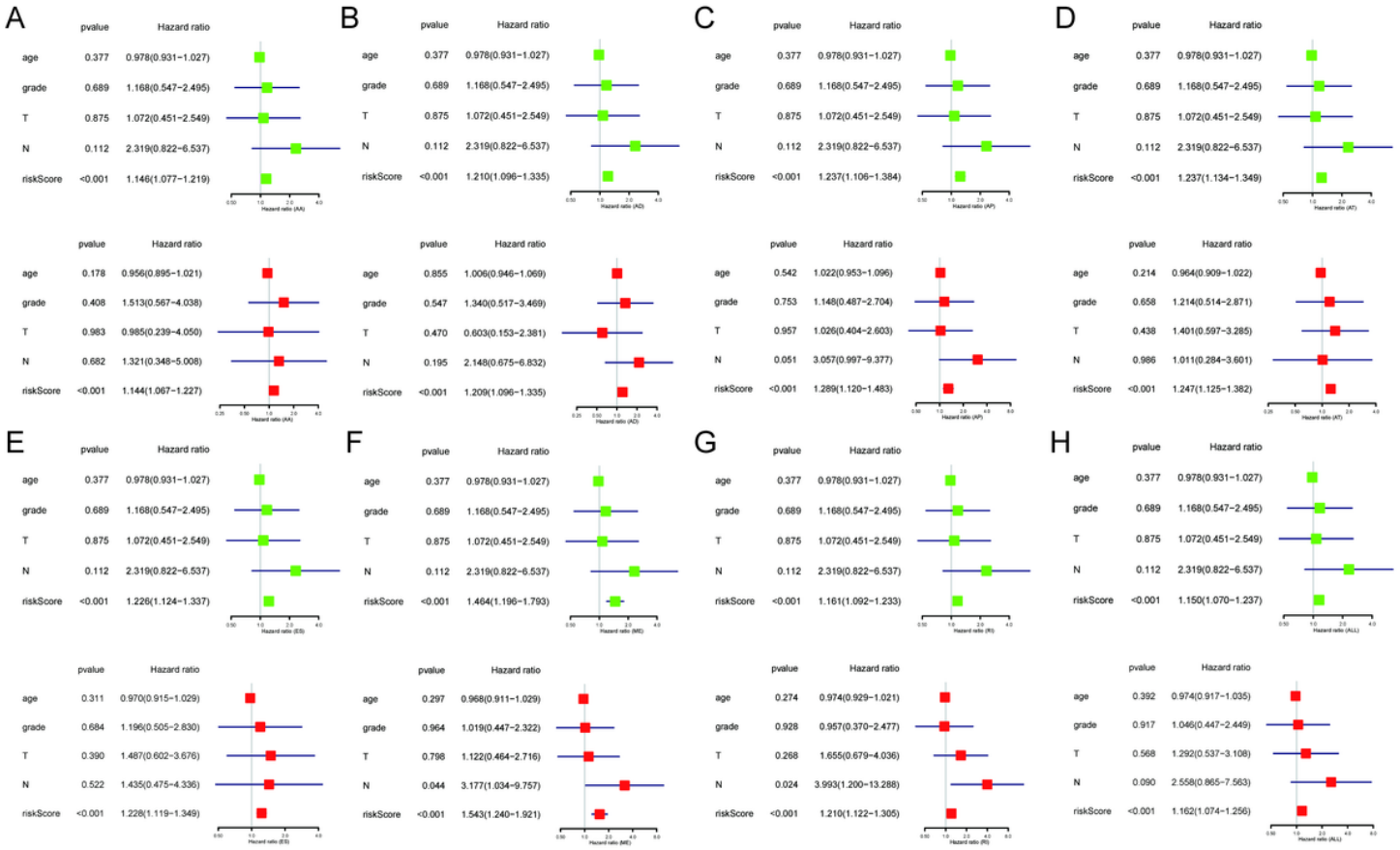

Figure 5

Forest plots. Forest plots of the HRs of age, grade, TNM stage and AS event associated prognostic signatures in AA, AD, AP, AT, ES, ME, RI and all AS pattern (A-H). Green represents univariate cox regression analysis, and red represents multivariate cox regression analysis. Dots represent the HR, and horizontal lines represent the 95% CIs. AS, Alternative splicing; AA, Alternate acceptor site; AD, Alternate donor site; AP, Alternate promoter; AT, Alternate terminator; ES, Exon skip; ME, Mutually exclusive exons; RI, Retained intron; HR, Hazard ratio; CI, Confidence interval.

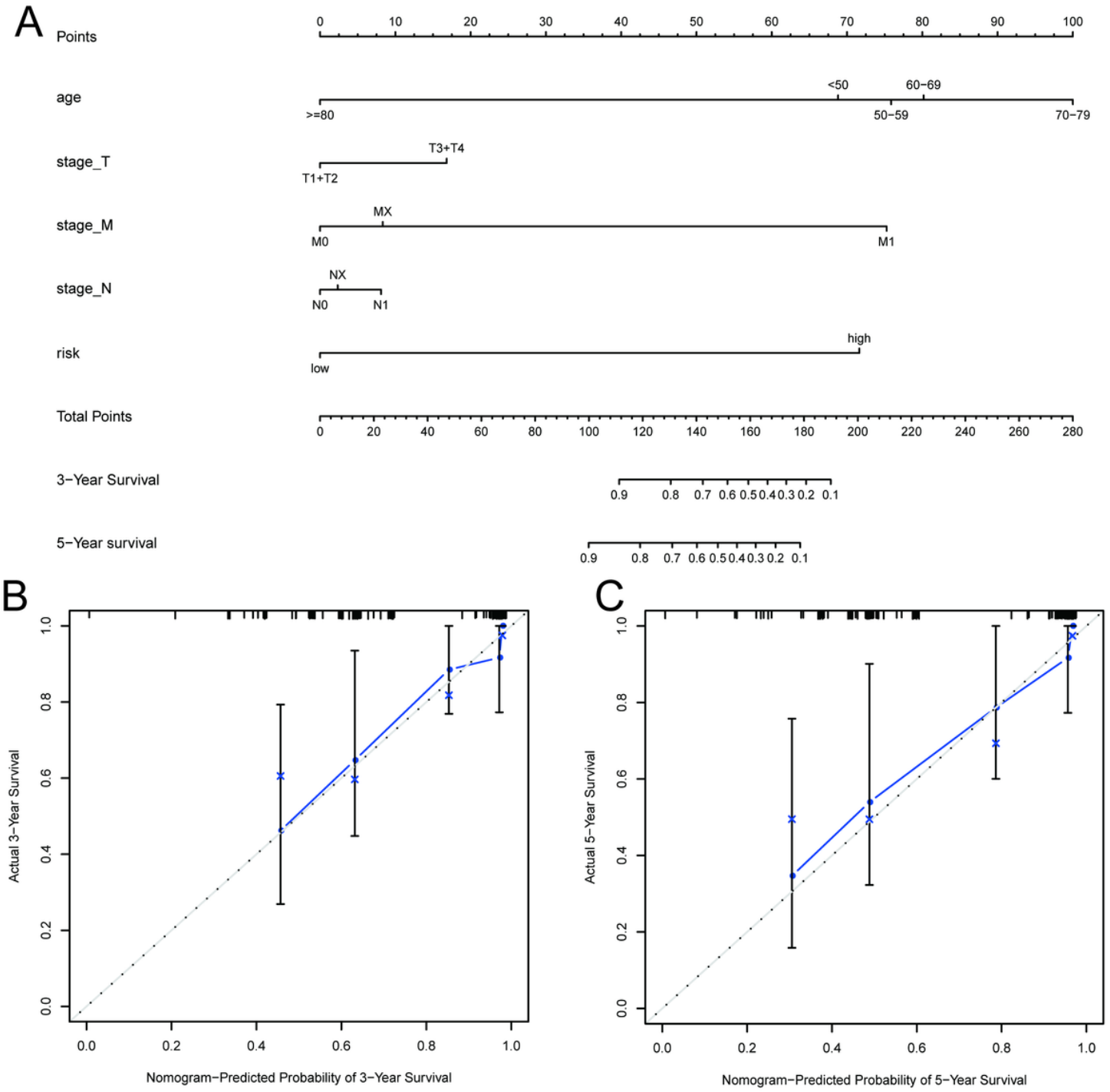

Figure 6

Nomogram. Nomogram for predicting individuals' survival probability in CESC (A). Calibration curves for predicting 3-year (B) and 5-year (C) survival. The x axis represents nomogram predicted survival, and y axis represents actual survival. The blue line represents the predicted survival probability, and vertical lines represent the 95% CIs. CI, Confidence interval.

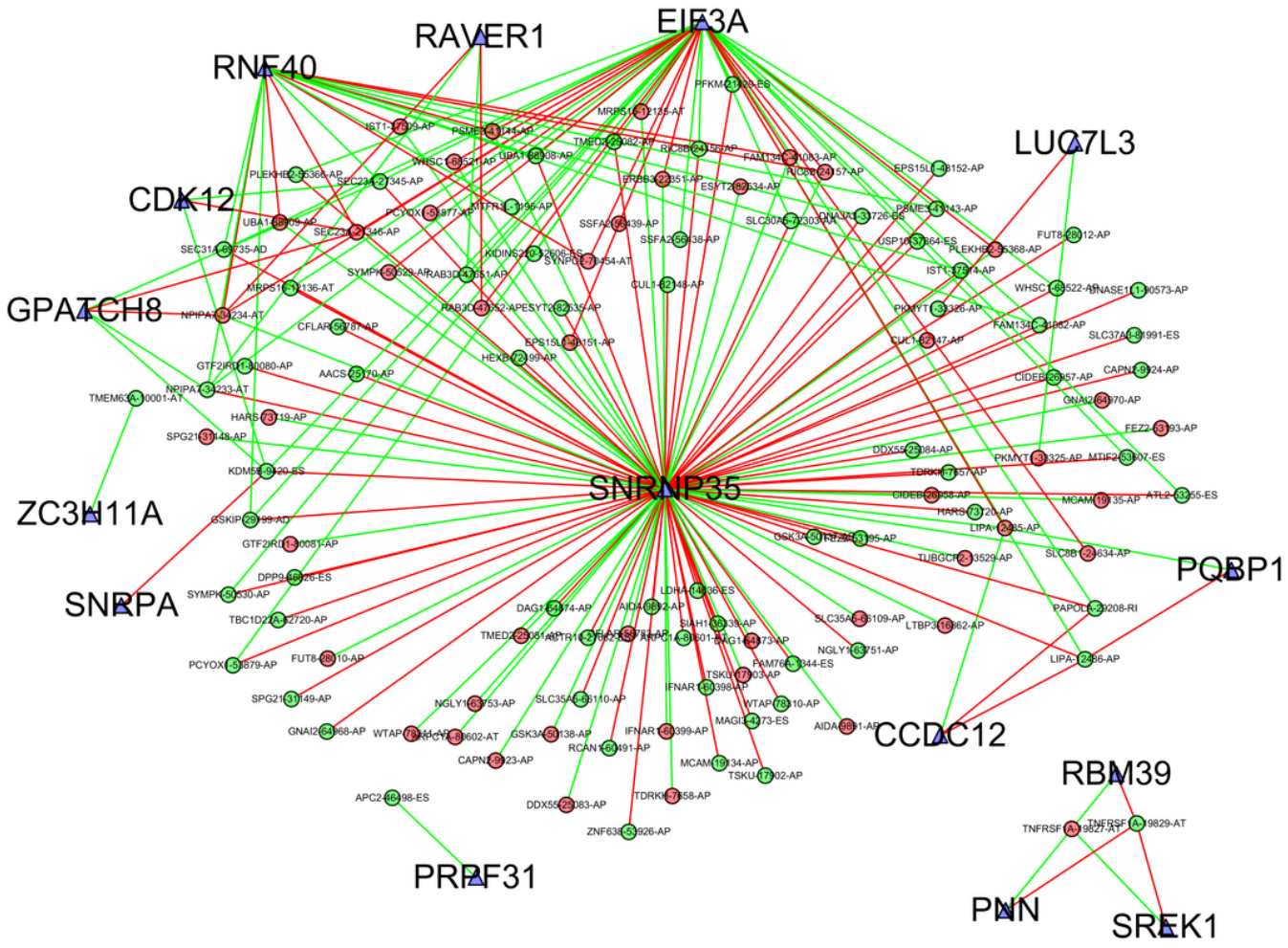

Figure 7

Correlation network between SASEs and SFs. Purple arrows represent SFs; Red and green circles represent high and low HR; Red and green lines represent positive and negative regulation. SASEs, Survival associated splicing events; SFs, Splicing factors; HR, Hazard ratio.

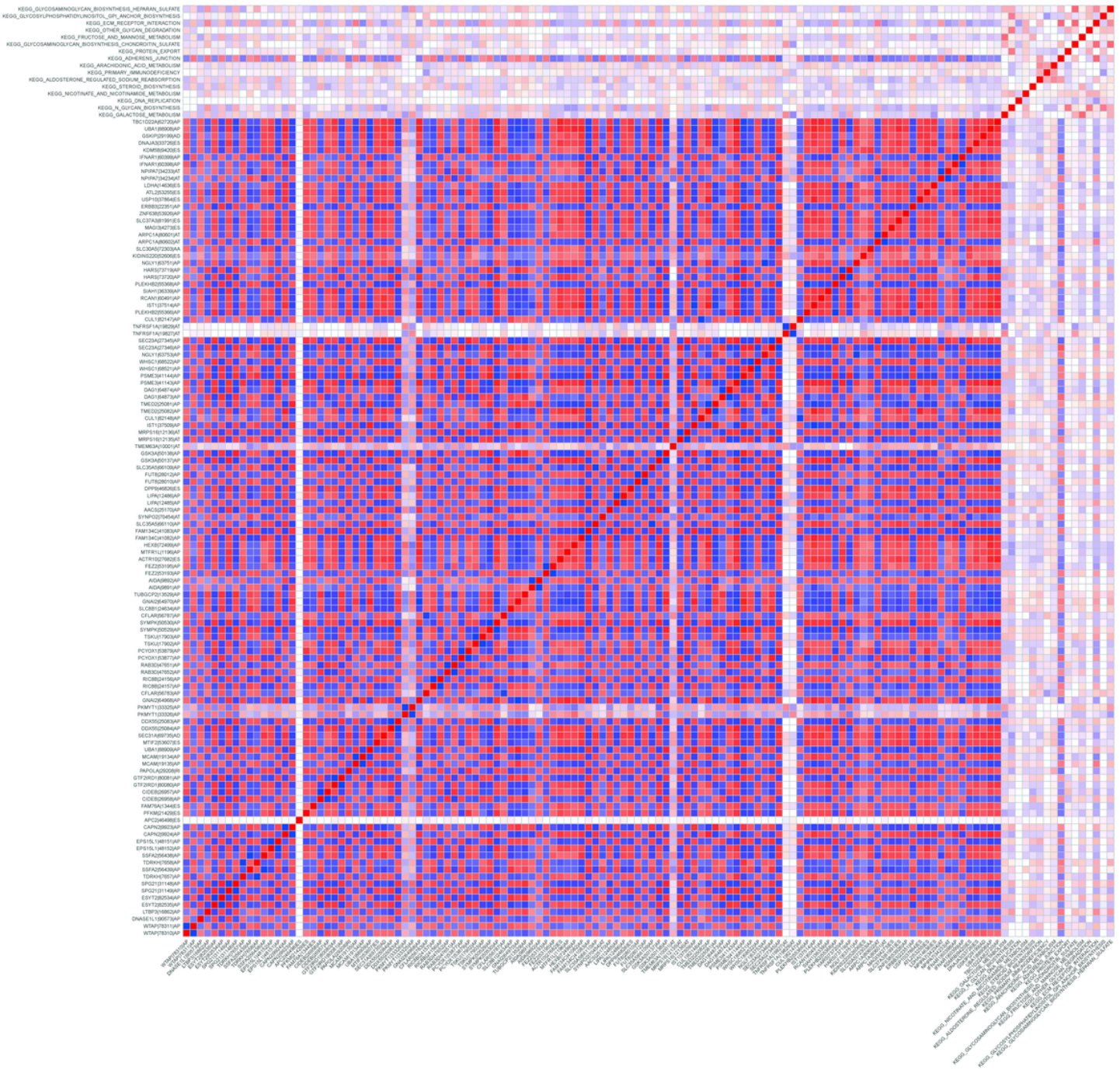

Figure 8

Heatmap of co-expression SASEs and KEGG pathways. Red and blue dots represent positive and negative correlation between SASEs and KEGG pathways. SASEs, Survival associated splicing events; KEGG, Kyoto Encyclopedia of Genes and Genomes.

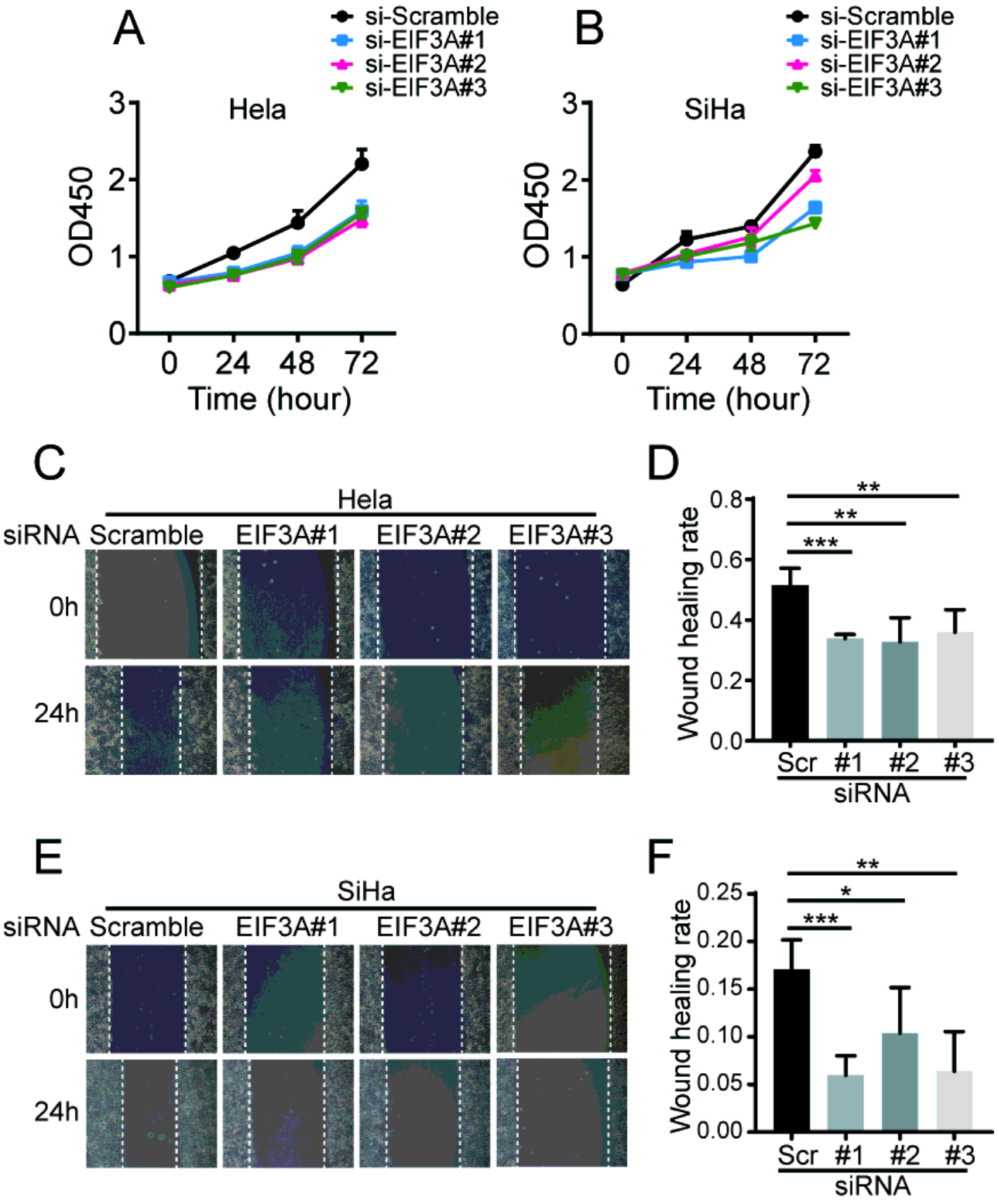

Figure 9

Knockdown of EIF3A decreases CESC cell proliferation and migration. Cell proliferation for the number of viable Hela and SiHa cells at 24h, 48h, and 72h using CCK-8 detection assays (A-B). Wound-healing assay for Hela cells after seeding. Representative photos are shown on the left (C-D). Wound-healing assay for SiHa cells after seeding (E-F). Column bars in (D) and (F) represent the wound healing rate from 3 independent experiments. Error bars represent ± SD. * P<0.05, ** P<0.01, *** P<0.001.