

Prognostic alternative splicing regulatory network of splicing events in acute myeloid leukemia patients based on SpliceSeq data from 136 cases

Z. C. XIE¹, L. GAO², G. CHEN², J. MA¹, L. H. YANG¹, R. Q. HE¹, M. W. LI¹, K. T. CAI¹, T. T. LI¹, Z. G. PENG^{1,*}

¹Department of Medical Oncology, First Affiliated Hospital of Guangxi Medical University, Nanning 530021, Guangxi, China; ²Department of Pathology, First Affiliated Hospital of Guangxi Medical University, Nanning 530021, Guangxi, China

*Correspondence: pengzhigang@gxmu.edu.cn

Received September 17, 2019 / Accepted December 8, 2019

This study aimed to create prognostic signatures to predict AML patients' survival using alternative splicing (AS) events. The AS data, RNA sequencing data, and the survival statistics of 136 AML patients were obtained from The Cancer Genome Atlas (TCGA) and TCGA SpliceSeq databases. Total 34,984 AS events generated from 8,656 genes, 2,583 of which were survival-associated AS events, were identified using univariate Cox regression. The prognostic models constructed using independent survival-associated AS events revealed that low-risk splicing better predicted patients' survival. ROC analysis indicated that the predictive efficacy of the alternate terminator model was best in the area under the curve at 0.781. Enrichment analysis revealed several important genes (TP53, BCL2, AURKB, PPP2R1B, FOS, and BIRC5) and pathways, such as the protein processing pathway in the endoplasmic reticulum, RNA transport pathway, and HTLV-I infection pathway. The splicing network of splicing events and factors revealed interesting interactions, such as the positive correlation between HNRNPH3 and CALHM2-13010-AT, which may indicate the potential splicing regulatory mechanism. Taken together, survival-associated splicing events and the prognostic signatures for predicting survival can help provide an overview of splicing in AML patients and facilitate clinical practice. The splicing regulatory network may improve the understanding of spliceosomes in AML.

Key words: alternative splicing, acute myeloid leukemia, The Cancer Genome Atlas (TCGA), prognostic signatures, splicing network

Acute myeloid leukemia (AML), a genetically heterogeneous clonal hematopoietic stem cell malignancy, manifests as an abnormal clonal expansion of myeloid precursors with attenuated differentiation ability [1, 2]. AML is the predominant type of acute leukemia in adults, representing approximately 80% of cases [3]. According to the American Cancer Society, in 2017, there were an estimated 21,380 new cases and 10,590 deaths due to AML [4]. AML patients suffer dismal outcomes, with a five-year overall survival of roughly 40% among younger adults (18–60 years old) and only 10–20% among patients over 60 years old [5]. Multiple prognostic factors, such as chromosomal abnormalities, recurrent genetic abnormalities, and microRNAs abnormalities, may contribute to the outcome of AML patients [6–8]. Current treatments for AML patients mainly involve remission induction and post-remission therapy, which include chemotherapy, targeted therapies, immunotherapy, and allogeneic hematopoietic stem cell transplantation [9–11]. However, drug resistance still poses a daunting challenge

for AML therapy [12]. Over the past few years, significant efforts have been made to identify potential new therapies for AML. For example, the expression of some mRNAs, such as H19, FHL2, iASPP, DNMT3A, U2AF1, and EZH2, has been found to correlate with unfavorable outcomes among AML patients, which may have the potential for novel AML therapies [13–15]. MicroRNAs are also a significant factor and play an important role in the survival of AML patients [16–20]. Long non-coding RNAs were previously found to have a role in the pathogenesis, apoptosis, therapeutic resistance and prognosis of AML patients [21–24]. Other studies have found that mTOR kinase [25], the STAT/TET1 axis [26], and methylation-independent CHFR expression [27] are also potential therapeutic approaches.

Alternative splicing (AS) is a primary driver of the phenotypic polymorphism of *Homo sapiens*, affecting over half of all human genes [28]. Many studies have found that AS functions have a significant role in various biological processes, including cell proliferation, differentiation, and

apoptosis [29, 30]. Imbalance of the splicing process may cause aberrant splicing events, resulting in various diseases and cancers [31]. Recent studies provided more evidence supporting the therapeutic potential of AS for cancer cases [32–34]. AS may function in the cancer process via various molecular mechanisms, enabling replicative immortality, maintaining proliferative signaling, eluding growth suppressors, inducing apoptosis, mediating angiogenesis, activating invasion and metastasis, reducing cellular energetics, escaping immune destruction, and regulating oncogenic signaling pathway [32, 35]. Over the last few years, an increasing number of studies have found that AS is crucial in AML cases. For example, plenty of abnormally spliced mRNA such as NOTCH2, Myc, PTPN6, WT1, FLT3, and Hoxa9 has been discovered as prognostic biomarkers and therapeutic targets [36]. The splicing process is extremely complicated, and plenty of factors could be involved in its regulation. Splicing factors, namely AS regulators, which are a type of RNA-binding protein, have been convincingly implicated in various human diseases, including tumors [37]. Scholars have extensively investigated splicing factor mutations such as SF3B1, U2AF1, SRSF2, and ZRSR2, were vulnerable for disruption of the spliceosome, which may indicate new targeted therapy strategies for cancers [38, 39]. However, few studies have investigated the regulatory network of splicing factors and AS events and posed their prognostic significance in AML patients.

Therefore, we focus on investigating survival-associated AS events and the potential molecular regulatory network via splicing data from the Cancer Genome Atlas (TCGA) concerning 136 AML cohorts. We hope to build splicing prognostic models for AML patients that provide novel insight into therapeutic targets in AML patients, and we try to construct the splicing network to explicate the potential mechanisms of AML.

Materials and methods

Curation and preprocessing of alternative splicing data. TCGA SpliceSeq (<http://bioinformatics.mdanderson.org/TCGASpliceSeq>), a compendium of alternative mRNA splicing data about malignancies, was used in this study. Percent spliced in (PSI) values for splicing events in samples with 33 different types of tumors, including the available adjacent normal samples, from TCGA SpliceSeq were included [40]. Samples with seven kinds of AS – exon skipping (ES), mutually exclusive exons (ME), retained intron (RI), alternate promoter (AP), alternate terminator (AT), alternate donor site (AD), and alternate acceptor site (AA) – were acquired from the TCGA SpliceSeq database. The PSI value multiplied by 100 was used to quantify each type of splicing. In addition, the clinical survival data, including overall survival time and the final status of AML, were downloaded from the TCGA source portal. Only patients with an overall survival time of over 90 days were enrolled in the current

study. The splicing and clinical data were matched by the TCGA labeling number.

Survival-associated alternative splicing events and prognostic signatures in AML. Cox proportional hazard regression is a semi-parametric model for investigating the effect of multiple variables upon time of an event of interest, such as relapse, progression or death, takes to happen [41, 42]. Univariate Cox regression only evaluates the effect of each predictor alone without all the others taken into account, while multivariate Cox regression can compensate for the limitation. Therefore, we adopted both univariate Cox regression and multivariate Cox regression for an integrated analysis. For each type of AS event, univariate Cox regression was first conducted using R version 3.43. AS events with a p-value of <0.05 in the results of univariate Cox regression were selected for their prognostic values. Considering the difference in the number of each type of AS events in the univariate Cox regression, we adopted flexible thresholds to select the optimal number of AS events for multivariate Cox regression. The most significant splicing events in AA (p<0.005), AD (p<0.005), AP (p<0.0001), AT (p<0.001), ES (p<0.0005), ME (p<0.05), and RI (p<0.0001) cases were then selected for multivariate Cox regression using SPSS 25.0 (Chicago, IL, USA) to identify independent prognostic indicators in AML. The splicing events involving each type of splicing that were significant (p<0.05) based on multivariate Cox regression were then used to construct predictive prognostic models. The expression of gene symbols in each model was analyzed using Gene Expression Profiling Interactive Analysis's online tool (<http://gepia.cancer-pku.cn/>), which is based on the TCGA and Genotype-Tissue Expression databases [43]. The risk score was calculated for each model using the following formula: risk score = $\sum_i^n PSI_i * \beta_i$, among which, the β value represents the regression coefficient of each AS event in the multivariate Cox regression [44–46]. The survival information of included patients and corresponding calculated risk scores were visualized as scatter plots and line charts, respectively via GraphPad Prism version 8 software. Patients were then divided into low- or high-risk groups based on the median risk score in each model. The receiver operating characteristic (ROC) curve and Kaplan-Meier (K-M) analysis were performed to evaluate the prognostic value of each prognosis model. ROC curve was achieved via survival ROC package, while K-M curve was generated via survminer package in R version 3.52. A two-sided p value below 0.05 was regarded as statistically significant.

Bioinformatics analysis. Since a single gene may have multiple splicing events, we adopted a novel UpSet plot [47] rather than the traditional Venn diagram to visualize interactive sets of the seven kinds of AS. For the most significant survival-associated splicing gene symbols (p<0.005), a protein interaction network was constructed using the STRING app plugin in Cytoscape version 3.6 [48, 49]. Biological enrichment analysis using gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG)

was conducted to identify significant molecular function terms and signal pathways via the Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.8 (<https://david.ncifcrf.gov/>) [50]. Statistically significant enriched terms ($p < 0.05$) were selected for displaying. Cytoscape version 3.6 was used to visualize the significant pathways and their involved genes.

Correlation analysis and splicing network. Splicing factors are RNA-binding proteins that modulate the spliceosome to catalyze the excision of intronic regions [51]. We wondered whether significant survival-associated AS events were modulated by some splicing factors that correlate with the survival of AML patients. Therefore, we collected known splicing factors from the SpliceAid2 database (www.introni.it/spliceaid.html) [52]. The mRNA-seq data regarding the splicing factors were downloaded from the TCGA portal. Significant survival-associated splicing factors were selected via univariate and multivariate Cox regression analysis. Linear correlation between the expression of the splicing factors and PSI values of splicing events was achieved using Spearman's correlation test. The correlation network was constructed using Cytoscape version 3.6. A p-value lower than 0.05 was deemed statistically significant.

Results

Integrated alternative splicing events in the AML cohort. In total, 136 AML patients from the TCGA database were included in the AS event analysis. A total of 34,984 AS events were generated from 8,656 genes, which suggested that one gene may be associated with different AS events. As shown in Figure 1, 2,646 AAs from 1,938 genes, 2,200 ADs from 1,633 genes, 7,207 APs from 2,917 genes, 7,734 ATs from 3,371 genes, 12,636 ESs from 5,373 genes, 168 MEs from 165 genes, and 2,393 RIs from 1,601 genes were obtained. ES was the most common AS event in AML patients.

Survival-associated alternative splicing events in AML patients. Univariate Cox regression analysis was conducted using the AS profiles in the AML cohort. We identified a total of 2,583 survival-associated AS events derived from 1,716 genes. Interestingly, a single gene might have multiple AS events that were significantly related to the survival of AML patients. As displayed in Figure 2A, we adopted an UpSet plot, which was more legible than a Venn diagram, to display intersecting sets. There were 15 genes with up to three different AS events. For example, the Fos proto-oncogene (FOS) gene, which was previously reported in AML cases, has AA, AD, and RI events. We determined the most significant survival-associated gene symbols ($p < 0.005$) in AS events to construct a protein-protein interaction network using Cytoscape version 3.6. As shown in Figure 2B, several hubs in AML patients, such as TP53, BCL2, AURKB, PPP2R1B, FOS, and BIRC5 (degree ≥ 10), were revealed by the protein network. Moreover, enrichment analysis revealed some significant GO items and KEGG pathways (Figure 3). The

HTLV-I infection pathway, which has been widely reported to occur in leukemogenesis cases, was included in these enrichment pathways.

Construction of a prognostic splicing model of AML patients. In total, five AA, eight AD, three AP, eight AT, eight ES, six ME, and two RI splicing events were selected for prognostic signatures using multivariate Cox regression. The seven groups of significant predictive AS events in multivariate Cox regression are displayed in Table 1. A total of 41 independent predictive AS events were identified. A prognostic model was constructed based on the risk score of each type of AS event. The survival information of the patients included in the models are presented (Figure 4A, Figures S1A–S5A. As shown in Figure 4B (and Figures S1B–S5B), patients with a low risk score for each type of AS event (i.e., AA, AD, AP, AT, ES, and ME) had better outcomes. In Figure 4C (and Figures S1C–S5C), we used the ROC curve to assess the predictive efficacy of each type of AS event. AT showed the best predictive efficacy (AUC=0.781), followed by AA (AUC=0.756), and both had great potential for clinical practice. The expression trend of gene symbols between AML and normal samples in each splicing model is shown (Figure 4D, Figures S1D–S5D). However, the RI prognostic model revealed no significant statistical difference between high-risk and low-risk groups (data not shown).

Regulatory network of survival-associated alternative splicing events. We obtained 66 splicing factors and corresponding expressions from the TCGA database. Univariate Cox regression was performed to identify the survival-associated splicing factors. Six such factors – ESRP1, HNRNPA1, HNRNPH1, HNRNPH3, RBM25, and TRA2B – were selected. Spearman's correlation analysis was performed to investigate the correlation among the six prognostic splicing factors and the 41 survival-associated AS events. As displayed in Figure 5, a splicing regulatory network consisting of 6 splicing factors and 18 AS events was constructed ($p < 0.05$) (Table 2). Four unfavorable AS

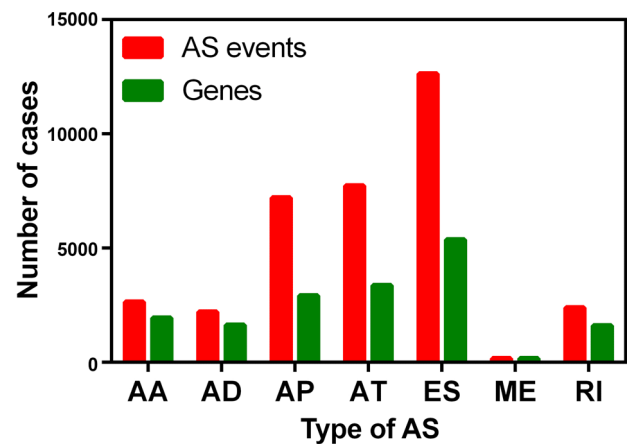


Figure 1. A bar plot presenting the number of alternative splicing (AS) events and the gene symbols for each type of splicing.

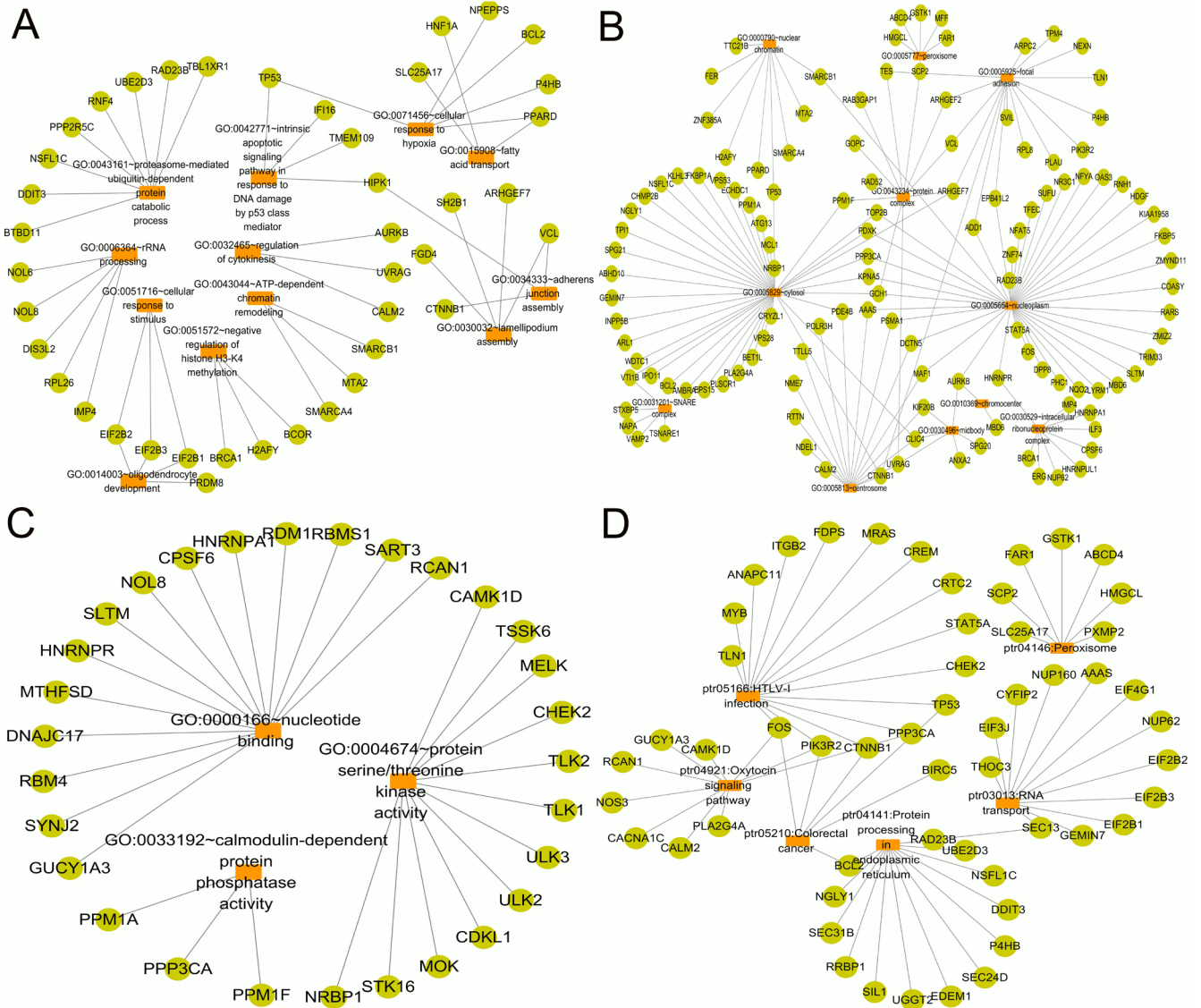


Figure 3. Enrichment analysis of the most significant gene symbols. A) Biological process of gene ontology (GO) analysis. B) Cellular component of GO analysis. C) Molecular function of GO analysis. D) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis.

mutually exclusive exons, alternative 5' acceptor sites, alternative 3' acceptor sites, intron retention, canonical splicing, exon scrambling, and trans-splicing [39]. It has been found that aberrant AS events are closely correlated to cancer cell proliferation, metastasis, apoptosis, metabolism, and drug resistance [54]. For example, a truncated splicing isoform, TEAD4-S, was found to attenuate the Hippo-YAP signaling pathway, inhibiting tumor proliferation [55]. Tien et al. discovered that CDK12 regulates alternative last exon mRNA splicing and promotes breast cancer cell invasion [56]. Moreover, Chen et al. found that snail driving AS of CD44 by ESRP1 is a pivotal regulating factor that enhances the invasion and migration of epithelial ovarian cancer cells [57]. Wan found that upregulated SRSF6 regulates ZO-1

aberrant splicing via direct binding to its motif in exon23, which promotes the progression of colorectal cancer [58]. Regarding cellular apoptosis, Adamopoulos et al. adopted a next-generation sequencing methodology to identify novel AS variants of the BCL2L12 gene, which is involved in the apoptotic mechanism of cancer [59]. Additionally, a previous study demonstrated that a spliced variant of CPEB2, CPEB2B, which acts as a translational activator of TWIST1 and HIF1 α , is a pivotal regulatory factor in core cellular pathways that relate to anoikis resistance and metastasis [60]. Previous studies found that AS affected cancer metabolism through versatile approaches such as regulation of the metabolic mTOR pathway, the c-Myc-SRSF1-mTOR axis, and various metabolic enzymes [61]. Splicing and drug resis-

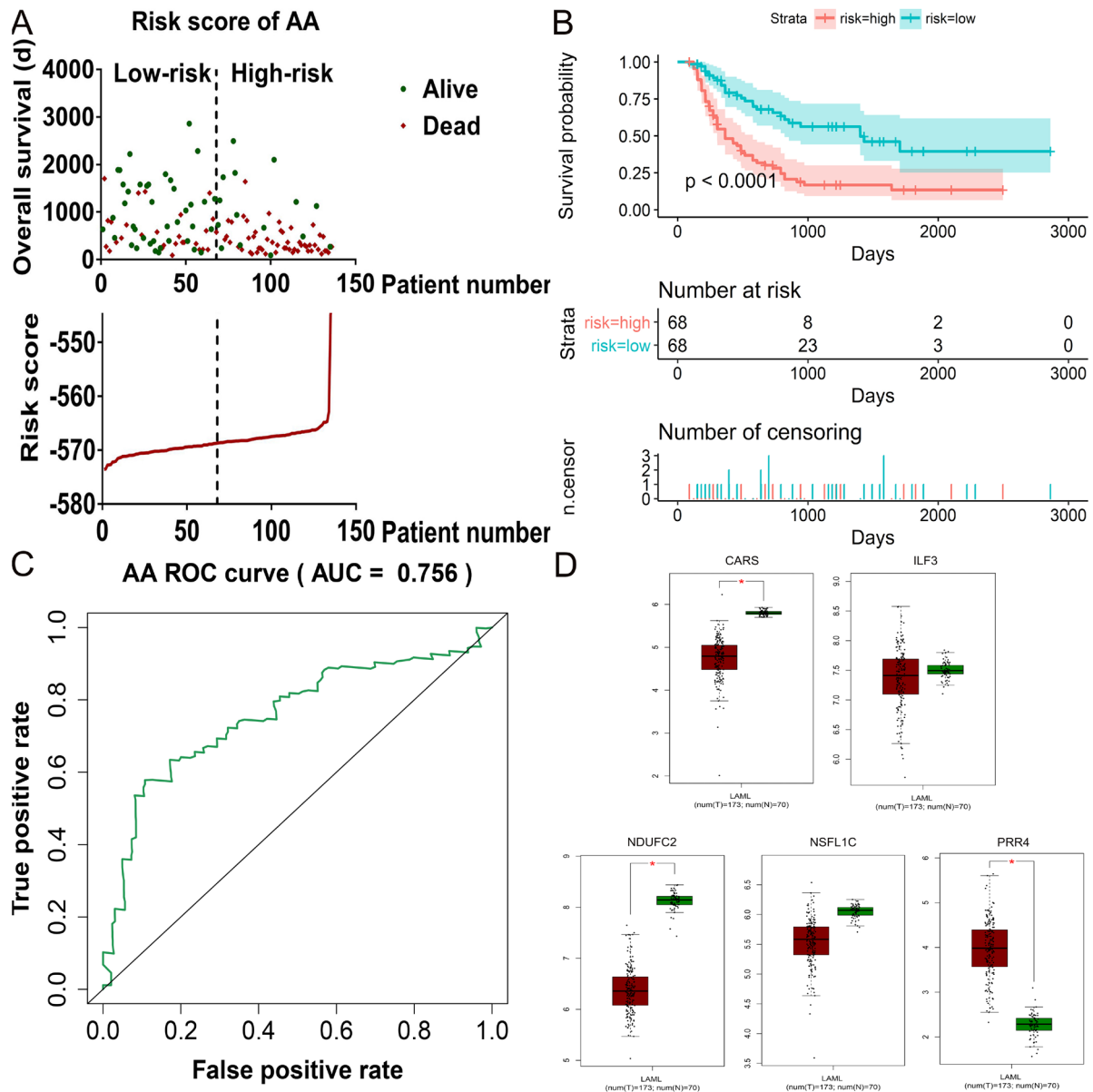


Figure 4. Construction of the prognostic signature of an alternate acceptor site (AA). **A**) Information about patients' survival based on their risk scores. The upper part of the graph displays patients' survival status divided by risk score, and the bottom part displays the linear variation in the risk score. **B**) Kaplan-Meier survival curve for low/high-risk groups based on the risk score. The low-risk group showed better survival than the high-risk group. **C**) Receiver operating characteristic (ROC) curves revealing the predictive capacity of the area under the curve at 0.756. **D**) Expression of each gene symbol and prognostic signatures based on TCGA and GTEx data. The statistically significant difference in gene expression between the tumor and normal samples is marked with an asterisk.

tance were also extensively reported. Additionally, authors have investigated BIM splicing and TKI-resistant CML, foyl-polyglutamate synthetase splicing and antifolate-resistant ALL, CD19 splicing and CART-19 immunotherapy-resistant B-ALL, BRAF splicing and vemurafenib-resistant melanoma, splicing and gemcitabine-resistant pancreatic ductal adenocarcinoma, androgen receptor splicing and drug-resistant castration-resistant prostate cancer, and BRCA1 splicing and PARPi-resistant breast cancer [62]. Splicing events have also

been revealed to be important factors affecting the outcome of cancer patients. For example, Leivonen found that AS could be used to discriminate between molecular subtypes of germinal center B-cell like (GCB) and activated B-cell like (ABC) diffuse large B-cell lymphoma (DLBCL), and it has a prognostic impact on DLBCL [63]. Sakuma found that a spliced variant of Cd44-variable exon 6 (Cd44v6) is modulated by a colorectal cancer metastasis suppressor during the epithelial-mesenchymal transition, which

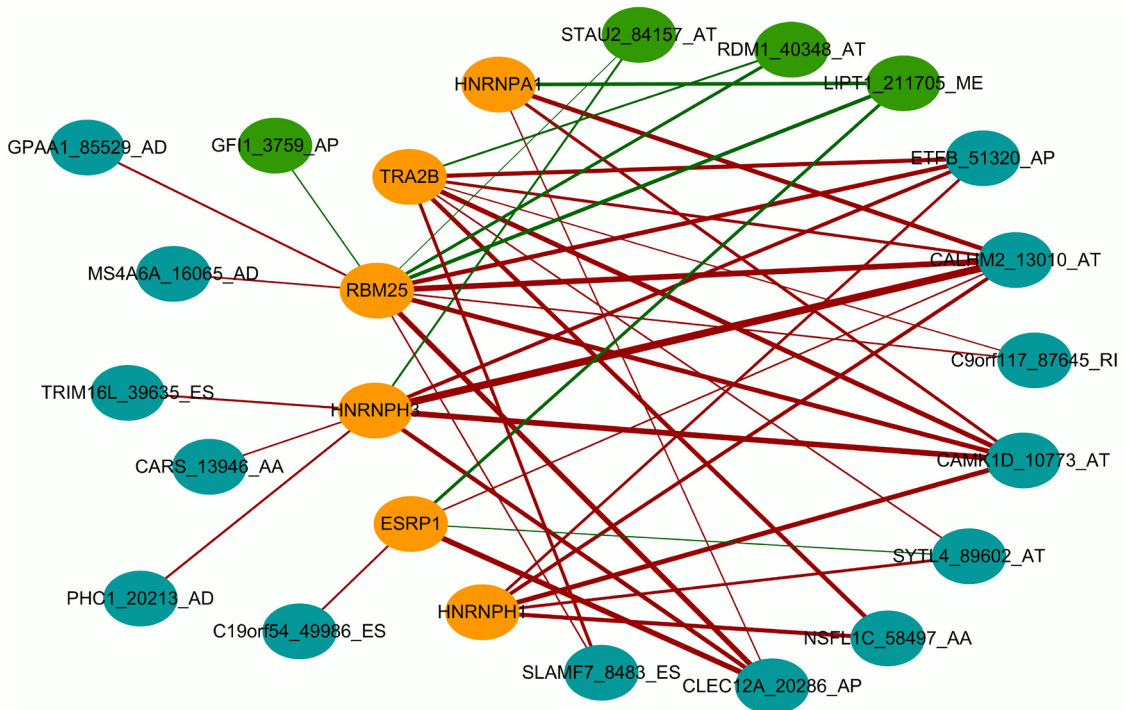


Figure 5. Regulatory splicing network of splicing factors and alternative splicing (AS) events. The orange nodes represent splicing factors. The blue nodes represent favorable AS events (HR<1), while green nodes represent unfavorable AS events (HR>1). The edges represent correlations, and the thicker edges indicate stronger correlations.

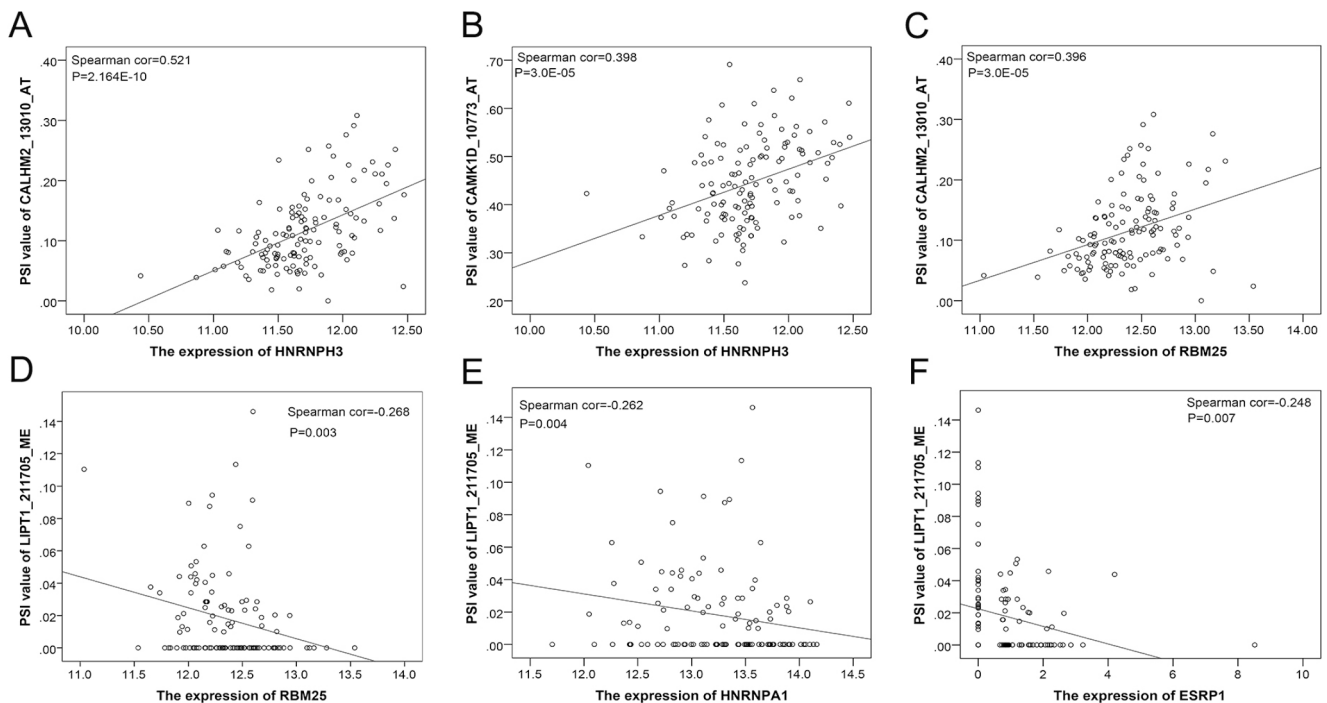


Figure 6. Linear correlation scatterplots for the greatest correlations between splicing factors and splicing events. A) Correlation between HNRNPH3 and CALHM2_13010_AT with a correlation coefficient of 0.521. B) Correlation between HNRNPH3 and CAMK1D_10773_AT with a correlation coefficient of 0.398. C) Correlation between RBM25 and CALHM2_13010_AT with a correlation coefficient of 0.396. D) Correlation between RBM25 and LIPT1_211705_ME with a correlation coefficient of -0.268. E) Correlation between HNRNPA1 and LIPT1_211705_ME with a correlation coefficient of -0.262. F) Correlation between ESRP1 and LIPT1_211705_ME with a correlation coefficient of -0.248.

Table 1. Details of the splicing events for each prognostic model.

Type	ID	β	HR	p-value	Lower	Upper	
AA	CARS-13946-AA	-0.042	0.959	0.015	0.927	0.992	
	ILF3-47575-AA	-1.026	0.359	0.003	0.184	0.699	
	NDUFC2-18006-AA	-4.359	0.013	0.001	0.001	0.166	
	NSFL1C-58497-AA	-0.119	0.888	0	0.831	0.949	
	PRR4-20482-AA	-0.264	0.768	0.046	0.593	0.995	
AD	GPAA1-85529-AD	-0.605	0.546	0.006	0.355	0.84	
	IMP4-55338-AD	-0.801	0.449	0.013	0.238	0.846	
	MS4A6A-16065-AD	-0.311	0.733	0.022	0.561	0.957	
	NBPF11-7331-AD	-0.125	0.882	0	0.829	0.939	
	PHC1-20213-AD	-0.192	0.825	0.005	0.721	0.944	
	PRG2-15831-AD	-0.415	0.661	0.002	0.509	0.857	
	UBL5-47438-AD	-0.293	0.746	0	0.636	0.876	
	HBA1-32835-AD	-0.686	0.504	0.045	0.258	0.984	
	AP	CLEC12A-20286-AP	-0.063	0.938	0.014	0.892	0.987
		ETFB-51320-AP	-0.04	0.961	0.029	0.927	0.996
GFI1-3759-AP		0.051	1.052	0.001	1.022	1.083	
AT	C4BPB-9625-AT	0.146	1.157	0.004	1.047	1.28	
	CALHM2-13010-AT	-0.074	0.929	0.001	0.891	0.969	
	CAMK1D-10773-AT	-0.039	0.962	0.017	0.932	0.993	
	ECE2-67861-AT	0.457	1.58	0.001	1.221	2.044	
	LRR23-19998-AT	-0.025	0.975	0.005	0.958	0.992	
	RDM1-40348-AT	0.039	1.04	0.007	1.011	1.07	
	STAU2-84157-AT	0.048	1.049	0.002	1.017	1.082	
	SYTL4-89602-AT	-0.013	0.987	0.008	0.978	0.997	
	ES	AURKB-39139-ES	-0.902	0.406	0	0.264	0.623
		C19orf54-49986-ES	-0.388	0.678	0.011	0.502	0.916
LYPLA2-1076-ES		-2.027	0.132	0.036	0.02	0.88	
MMAB-24331-ES		-0.132	0.876	0.014	0.788	0.974	
ODF2-87756-ES		-1.602	0.201	0	0.095	0.429	
PPP2R1B-18679-ES		-0.212	0.809	0.001	0.717	0.913	
SLAMF7-8483-ES		-0.089	0.915	0.002	0.865	0.969	
TRIM16L-39635-ES		-0.181	0.834	0.004	0.738	0.942	
ME		ABCB8-234119-ME	0.048	1.049	0.044	1.001	1.099
		C2CD5-251535-ME	-0.049	0.952	0.009	0.917	0.988
	DNM1-87724-ME	0.038	1.039	0.002	1.014	1.065	
	LIPT1-211705-ME	0.198	1.219	0	1.111	1.338	
	MCM8-58660-ME	0.03	1.031	0.005	1.009	1.053	
RI	NEDD1-300849-ME	-0.081	0.922	0.016	0.864	0.985	
	CAST-72851-RI	-2.252	0.105	0	0.046	0.239	
	C9orf117-87645-RI	-0.1	0.905	0	0.866	0.946	

Table 2. The correlations between splicing factors and AS events.

AS factor	AS events	R	p-value	HR
ESRP1	LIPT1_211705_ME	0.248	0.007	1.219
ESRP1	SYTL4_89602_AT	0.189	0.032	0.987
ESRP1	CALHM2_13010_AT	0.201	0.022	0.929
ESRP1	C19orf54_49986_ES	0.213	0.027	0.678
ESRP1	CLEC12A_20286_AP	0.383	0	0.938
HNRNPA1	LIPT1_211705_ME	0.262	0.004	1.219
HNRNPA1	CLEC12A_20286_AP	0.195	0.027	0.938
HNRNPA1	CAMK1D_10773_AT	0.247	0.005	0.962
HNRNPA1	CALHM2_13010_AT	0.32	0	0.929
HNRNPH1	ETFB_51320_AP	0.236	0.007	0.961
HNRNPH1	SYTL4_89602_AT	0.24	0.006	0.987
HNRNPH1	CALHM2_13010_AT	0.264	0.002	0.929
HNRNPH1	NSFL1C_58497_AA	0.289	0.001	0.888
HNRNPH1	CAMK1D_10773_AT	0.343	0	0.962
HNRNPH3	STAU2_84157_AT	0.215	0.014	1.049
HNRNPH3	CARS_13946_AA	0.201	0.032	0.959
HNRNPH3	TRIM16L_39635_ES	0.216	0.018	0.834
HNRNPH3	PHC1_20213_AD	0.219	0.012	0.825
HNRNPH3	ETFB_51320_AP	0.267	0.002	0.961
HNRNPH3	CLEC12A_20286_AP	0.304	0	0.938
HNRNPH3	CAMK1D_10773_AT	0.398	0	0.962
HNRNPH3	CALHM2_13010_AT	0.521	0	0.929
RBM25	LIPT1_211705_ME	0.268	0.003	1.219
RBM25	RDM1_40348_AT	0.247	0.005	1.04
RBM25	GFI1_3759_AP	0.195	0.043	1.052
RBM25	STAU2_84157_AT	0.179	0.042	1.049
RBM25	SLAMF7_8483_ES	0.2	0.037	0.915
RBM25	C9orf117_87645_RI	0.201	0.036	0.905
RBM25	MS4A6A_16065_AD	0.204	0.027	0.733
RBM25	GPAA1_85529_AD	0.215	0.014	0.546
RBM25	ETFB_51320_AP	0.318	0	0.961
RBM25	CAMK1D_10773_AT	0.336	0	0.962
RBM25	CLEC12A_20286_AP	0.37	0	0.938
RBM25	CALHM2_13010_AT	0.396	0	0.929
TRA2B	RDM1_40348_AT	0.214	0.014	1.04
TRA2B	C9orf117_87645_RI	0.192	0.046	0.905
TRA2B	SYTL4_89602_AT	0.202	0.021	0.987
TRA2B	CALHM2_13010_AT	0.241	0.006	0.929
TRA2B	SLAMF7_8483_ES	0.252	0.008	0.915
TRA2B	ETFB_51320_AP	0.295	0.001	0.961
TRA2B	NSFL1C_58497_AA	0.308	0	0.888
TRA2B	CAMK1D_10773_AT	0.344	0	0.962

indicated poor prognosis for colorectal cancer [64]. Taken together, the crucial role of AS in the pathophysiological process and treatment response is beyond doubt. However, the above studies mainly focused on one AS, lacking a general view of AS entirety. Recently, scholars began to investigate the whole AS and their prognostic value in cancers. For example, Li et al. observed many prognostic AS events and created prognostic predictors of non-small cell lung cancer [65], which pioneered the exploration of AS and its prognostic

significance for predicting a patient's survival. Other studies used similar methods to investigate other cancers, such as bladder cancer [66], prostate cancer [46], diffuse large B-cell lymphoma [67], and gastrointestinal pan-adenocarcinomas [68], revealing insights into splice-based therapy in clinical practice. Targeting AS events have been widely studied for application to cancer therapies, such as those targeting the core spliceosome mechanism, targeting splicing regulators, and using oligonucleotide-based modality to modulate

splicing [33]. Therefore, it is important to comprehensively investigate aberrant splicing and its prognostic implications for AML cases to identify novel targeted therapy strategies that can be applied in clinical practice.

In our study, we systematically investigated survival-associated AS events in AML cases via the splicing source, 136 AML patients in the TCGA database. We are the first to comprehensively report AS events and their influence on the prognosis of AML patients. A total of 2,583 AS events were correlated with patient survival in AML cases, and 45 were independent survival-associated AS events. As far as we know, these are the largest number of survival-associated AS events identified in AML so far. The prognostic AS events, specifically the 45 independent survival-associated AS events show promising potential in targeted therapy of AML patients. A gene network was used to identify the most significant gene symbols for AS events, revealing several important prognostic genes, including TP53, BCL2, AURKB, PPP2R1B, FOS, and BIRC5. The prognostic value of these genes for AML patients has been reported in previous studies. For example, AML patients with altered TP53 were found to have 6 times shorter overall survival than those without altered TP53 [69]. Additionally, Zhang et al. discovered that BCL2 was regulated by miR-182-5p, which may act as a potential therapeutic target in AML [70], and Moore found that expression of a splice variant of BIRC5 (survivin) indicated poor outcomes in pediatric AML patients [71]. The previously reported studies provided additional evidence for the prognostic genes and suggested the important research value of our results. The unreported prognostic genes such as PPP2R1B, are worthy of further investigation. Our enrichment analysis uncovered significant pathways in AML patients, such as the protein processing pathway in the endoplasmic reticulum, RNA transport pathway, and HTLV-I infection pathway. In the protein processing pathway in the endoplasmic reticulum, endoplasmic reticulum stress and unfolded protein response are parts of the important process, which have been reported to be associated with AML before [72, 73]. And in the RNA transport pathway, a large macromolecular assembly named nuclear pore complexes (NPCs) is a channel for RNA transportation, which is fundamental for chromatin organization, regulation of gene expression and DNA repair [74–76]. Some component of NPCs such as Nup214, may disturb the regulation of gene expression in AML [77]. However, no study has concretely investigated the relationship between the RNA transport pathway and AS in AML, which is worthy of further studies. As we all know, human T-cell leukemia virus type 1 (HTLV-1) is a pathogenic factor for adult T-cell leukemia/lymphoma. HTLV-1 RNA splicing has been observed before, but the exact mechanism still remains unknown [78]. To sum up, our results provided some clues for the potential role of AS and key signaling pathways in the molecular mechanism of AML. The novel perspectives of AS and key pathways in the mechanism of AML are noteworthy in future research.

Previous studies used a risk score to construct predictive models of AML patients. For instance, Ng used a 17-gene score to predict AML patients' therapy resistance (AUC=0.78), which outperformed cytogenetic risk (AUC=0.70) [79]. Additionally, Wang used a three-lncRNA signature to predict the prognosis of AML patients (AUC=0.710) [80] and Zhu et al. conducted a four-microRNA signature to determine the prognosis in AML patients in a training group (AUC=0.681) and testing group (AUC=0.805) [81]. Previous studies showed feasibility that using combined multi-markers for predicting prognosis of AML patients. In our study, we used a similar method to investigate the prognostic signatures of AS events in AML. The current study is the first to use the risk score of six different types of splicing signatures to predict the survival of AML patients. It provides innovative approaches to apply AS to clinical practice. For all types of splicing, the low-risk group showed significantly better overall survival than the high-risk group. This indicates the potential clinical value of our AS prognostic models in AML patients. We wondered which AS prognostic model will show the highest predictive efficacy for the survival of AML patients. We compared the predictive efficacy of each prognostic model. AT showed the best performance for predicting survival (AUC=0.781), followed by AA (AUC=0.756), ME (AUC=0.733), AP (AUC=0.708), ES (AUC=0.70) and AD (AUC=0.648). Compared with previously reported prognostic models, our splicing prognostic models still showed satisfactory performance in general. It is promising to apply these splicing signatures for predicting the prognosis of AML in clinical practice.

The spliceosome, a complex and highly dynamic ribonucleoprotein, catalyzes pre-mRNA splicing in a step-wise manner to ensure accurate and flexible splicing [82, 83]. The splicing factor is the essential protein in the core spliceosome, which has context-dependent functional flexibility in the regulation of splicing [39, 84]. Some splicing factors, including SF3B1, SRSF2, U2AF1, and ZRSR2, have been widely reported in the context of cancers [85–87]. In our study, we investigated whether survival-associated splicing factors regulate the prognosis-related AS events mentioned above. To reveal the underlying mechanism of splicing in AML patients, we analyzed the prognostic value of 66 splicing factors and constructed a splicing regulatory network. We found that splicing factors are positively correlated with almost all the favorable AS events. In addition, the same splicing factor may have different roles in different AS events, and the same AS events could be regulated by different splicing factors. For example, the splicing factor HNRNPH3 can positively regulate CAMK1D-10773-AT and CALHM2-13010-AT; the splicing factor RBM25 can positively and negatively regulate LIPT1-211705-ME and CALHM2-13010-AT, respectively; and the AS event LIPT1-211705-ME can be negatively regulated by the splicing factors RBM25, HNRNPA1, and ESRP1. These indicated intricate interactions between splicing factors and AS events. The exact

regulatory patterns and functions involved in splicing could be more complicated than revealed here, but the splicing network did provide an overview of the spliceosome in AML patients. This is the first relatively complete splicing regulatory network constructed in AML so far. It provides inspiration in the research of spliceosome in AML. It deserves more future studies paying sustained attention in the underlying mechanism of splicing in AML.

Some limitations of the present study must be addressed. Due to the lack of splicing data for the normal samples in TCGA, we could not compare the differences in splicing between the AML and normal samples. Additionally, only 136 cases were enrolled in the present study. Future studies should employ larger samples. And in the current study, no cross-validation was performed. It is necessary to use another cohort to validate the prognostic models in the future study. In the current prognostic models, we only adopted AS events for the construction of prognostic signatures without taking other factors into account. It still needs a more exhaustive analysis of the prognostic factors in AML patients. Further, regarding the splicing regulatory network, we only investigated regulation among splicing factors and AS events. However, the splicing process is likely more sophisticated, and another protein in the spliceosome could exert regulatory effects. This deserves investigation in the future. Last, the current study was based on bioinformatics methods and thus remains at the bioinformatics level. It is necessary to conduct experimental studies to validate the results.

In conclusion, our study comprehensively investigated the prognostic value of AS data from AML patients. We presented a number of survival-associated AS events and created prognostic signatures to predict the survival of AML patients. Moreover, we constructed a splicing regulatory network, which might help illuminate the molecular mechanisms underlying splicing in AML cases.

Supplementary information is available in the online version of the paper.

Acknowledgments: This study was supported by the Guangxi Natural Science Foundation (2015GXNSFDA139028) and the Students Innovative Project of the First Clinical Medical College of Guangxi Medical University (2018). The funding sources did not affect the study design, data acquisition and analysis, writing of the manuscript, or decision to publish. The authors would like to thank members of our team – Hua-yu Wu, Yi-wu Dang, Zhi-guang Huang, Peng Lin, and Liang Liang – for their help. Additionally, we would like to thank those who created the TCGA database for its valuable data.

References

- [1] DE KOUCHKOVSKY I, ABDUL-HAY M. 'Acute myeloid leukemia: a comprehensive review and 2016 update'. *Blood Cancer J* 2016; 6: e441. <https://doi.org/10.1038/bcj.2016.50>
- [2] PROKOCIMER M, MOLCHADSKY A, ROTTER V. Dysfunctional diversity of p53 proteins in adult acute myeloid leukemia: projections on diagnostic workup and therapy. *Blood* 2017; 130: 699–712. <https://doi.org/10.1182/blood-2017-02-763086>
- [3] LAGUNAS-RANGEL FA, CHAVEZ-VALENCIA V, GOMEZ-GUIJOSA MA, CORTES-PENAGOS C. Acute Myeloid Leukemia-Genetic Alterations and Their Clinical Prognosis. *Int J Hematol Oncol Stem Cell Res* 2017; 11: 328–339.
- [4] SIEGEL RL, MILLER KD, JEMAL A. Cancer Statistics, 2017. *CA Cancer J Clin* 2017; 67: 7–30. <https://doi.org/10.3322/caac.21387>
- [5] KAYSER S, LEVIS MJ. Advances in targeted therapy for acute myeloid leukaemia. *Br J Haematol* 2018; 180: 484–500. <https://doi.org/10.1111/bjh.15032>
- [6] WALLACE JA, O'CONNELL RM. MicroRNAs and acute myeloid leukemia: therapeutic implications and emerging concepts. *Blood* 2017; 130: 1290–1301. <https://doi.org/10.1182/blood-2016-10-697698>
- [7] DOHNER H, ESTEY E, GRIMWADE D, AMADORI S, APPELBAUM FR et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 2017; 129: 424–447. <https://doi.org/10.1182/blood-2016-08-733196>
- [8] TRINO S, LAMORTE D, CAIVANO A, LAURENZANA I, TAGLIAFERRI D et al. MicroRNAs as New Biomarkers for Diagnosis and Prognosis, and as Potential Therapeutic Targets in Acute Myeloid Leukemia. *Int J Mol Sci* 2018; 19. <https://doi.org/10.3390/ijms19020460>
- [9] LIU Y, BEWERSDORF JP, STAHL M, ZEIDAN AM. Immunotherapy in acute myeloid leukemia and myelodysplastic syndromes: The dawn of a new era? *Blood Rev* 2018; 34: 67–83. <https://doi.org/10.1016/j.blre.2018.12.001>
- [10] LUPPI M, FABBIANO F, VISANI G, MARTINELLI G, VENDITTI A. Novel Agents for Acute Myeloid Leukemia. *Cancers (Basel)* 2018; 10: e 429. <https://doi.org/10.3390/cancers10110429>
- [11] KARJALAINEN E, REPASKY GA. Molecular Changes During Acute Myeloid Leukemia (AML) Evolution and Identification of Novel Treatment Strategies Through Molecular Stratification. *Prog Mol Biol Transl Sci* 2016; 144: 383–436. <https://doi.org/10.1016/bs.pmbts.2016.09.005>
- [12] HACKL H, ASTANINA K, WIESER R. Molecular and genetic alterations associated with therapy resistance and relapse of acute myeloid leukemia. *J Hematol Oncol* 2017; 10: 51. <https://doi.org/10.1186/s13045-017-0416-0>
- [13] CHENG Z, DAI Y, PANG Y, JIAO Y, ZHAO H et al. Enhanced expressions of FHL2 and iASPP predict poor prognosis in acute myeloid leukemia. *Cancer Gene Ther* 2018; 26: 17–25. <https://doi.org/10.1038/s41417-018-0027-0>
- [14] ZHANG TJ, ZHOU JD, ZHANG W, LIN J, MA JC et al. H19 overexpression promotes leukemogenesis and predicts unfavorable prognosis in acute myeloid leukemia. *Clin Epigenetics* 2018; 10: 47. <https://doi.org/10.1186/s13148-018-0486-z>
- [15] SAYGIN C, HIRSCH C, PRZYCHODZEN B, SEKERE MA, HAMILTON BK et al. Mutations in DNMT3A, U2AF1, and EZH2 identify intermediate-risk acute myeloid leukemia patients with poor outcome after CR1. *Blood Cancer J* 2018; 8: 4. <https://doi.org/10.1038/s41408-017-0040-9>

- [16] GABRA MM, SALMENA L. microRNAs and Acute Myeloid Leukemia Chemoresistance: A Mechanistic Overview. *Front Oncol* 2017; 7: 255. <https://doi.org/10.3389/fonc.2017.00255>
- [17] LIM EL, TRINH DL, RIES RE, WANG J, GERBING RB et al. MicroRNA Expression-Based Model Indicates Event-Free Survival in Pediatric Acute Myeloid Leukemia. *J Clin Oncol* 2017; 35: 3964–3977. <https://doi.org/10.1200/JCO.2017.74.7451>
- [18] LIN XC, XU Y, SUN GP, WEN JL, LI N et al. Molecular dysfunctions in acute myeloid leukemia revealed by integrated analysis of microRNA and transcription factor. *Int J Oncol* 2016; 48: 2367–2380. <https://doi.org/10.3892/ijo.2016.3489>
- [19] ZHANG TJ, QIAN Z, WEN XM, ZHOU JD, LI XX et al. Lower expression of bone marrow miR-122 is an independent risk factor for overall survival in cytogenetically normal acute myeloid leukemia. *Pathol Res Pract* 2018; 214: 896–901. <https://doi.org/10.1016/j.prp.2018.03.027>
- [20] CHEN X, YANG S, ZENG J and CHEN M. miR12715p inhibits cell proliferation and induces apoptosis in acute myeloid leukemia by targeting ZIC2. *Mol Med Rep* 2018; 19: 508–514.
- [21] LEI L, XIA S, LIU D, LI X, FENG J et al. Genome-wide characterization of lncRNAs in acute myeloid leukemia. *Brief Bioinform* 2018; 19: 627–635. <https://doi.org/10.1093/bib/bbx007>
- [22] PAPAIOANNOU D, NICOLET D, VOLINIA S, MROZEK K, YAN P et al. Prognostic and biologic significance of long non-coding RNA profiling in younger adults with cytogenetically normal acute myeloid leukemia. *Haematologica* 2017; 102: 1391–1400. <https://doi.org/10.3324/haematol.2017.166215>
- [23] HUANG JL, LIU W, TIAN LH, CHAI TT, LIU Y et al. Up-regulation of long non-coding RNA MALAT-1 confers poor prognosis and influences cell proliferation and apoptosis in acute monocytic leukemia. *Oncol Rep* 2017; 38: 1353–1362. <https://doi.org/10.3892/or.2017.5802>
- [24] YIN X, HUANG S, ZHU R, FAN F, SUN C et al. Identification of long non-coding RNA competing interactions and biological pathways associated with prognosis in pediatric and adolescent cytogenetically normal acute myeloid leukemia. *Cancer Cell Int* 2018; 18: 122. <https://doi.org/10.1186/s12935-018-0621-0>
- [25] TABE Y, TAFURI A, SEKIHARA K, YANG H, KONOPLEVA M. Inhibition of mTOR kinase as a therapeutic target for acute myeloid leukemia. *Expert Opin Ther Targets* 2017; 21: 705–714. <https://doi.org/10.1080/14728222.2017.1333600>
- [26] JIANG X, HU C, FERCHEN K, NIE J, CUI X et al. Targeted inhibition of STAT/TET1 axis as a therapeutic strategy for acute myeloid leukemia. *Nat Commun* 2017; 8: 2099. <https://doi.org/10.1038/s41467-017-02290-w>
- [27] ZHOU JD, ZHANG TJ, LI XX, MA JC, GUO H et al. Methylation-independent CHFR expression is a potential biomarker affecting prognosis in acute myeloid leukemia. *J Cell Physiol* 2018; 233: 4707–4714. <https://doi.org/10.1002/jcp.26253>
- [28] BERGSMAN AJ, VAN DER WAL E, BROEDERS M, VAN DER PLOEG AT, PIM PIJNAPPEL WWM. Alternative Splicing in Genetic Diseases: Improved Diagnosis and Novel Treatment Options. *Int Rev Cell Mol Biol* 2018; 335: 85–141. <https://doi.org/10.1016/bs.ircmb.2017.07.008>
- [29] FISZBEIN A, KORNBLIHTT AR. Alternative splicing switches: Important players in cell differentiation. *Bioessays* 2017; 39. <https://doi.org/10.1002/bies.201600157>
- [30] LIN JC, TSAO MF, LIN YJ. Differential Impacts of Alternative Splicing Networks on Apoptosis. *Int J Mol Sci* 2016; 17: e2097. <https://doi.org/10.3390/ijms17122097>
- [31] GALLEGO-PAEZ LM, BORDONE MC, LEOTE AC, SARATIVA-AGOSTINHO N, ASCENSAO-FERREIRA M et al. Alternative splicing: the pledge, the turn, and the prestige: The key role of alternative splicing in human biological systems. *Hum Genet* 2017; 136: 1015–1042. <https://doi.org/10.1007/s00439-017-1790-y>
- [32] MARTINEZ-MONTIEL N, ROSAS-MURRIETA NH, ANAYA RUIZ M, MONJARAZ-GUZMAN E, MARTINEZ-CONTRERAS R. Alternative Splicing as a Target for Cancer Treatment. *Int J Mol Sci* 2018; 19: e545. <https://doi.org/10.3390/ijms19020545>
- [33] LIN JC. Therapeutic Applications of Targeted Alternative Splicing to Cancer Treatment. *Int J Mol Sci* 2017; 19: e75. <https://doi.org/10.3390/ijms19010075>
- [34] SONG X, ZENG Z, WEI H, WANG Z. Alternative splicing in cancers: From aberrant regulation to new therapeutics. *Semin Cell Dev Biol* 2018; 75: 13–22. <https://doi.org/10.1016/j.semcdb.2017.09.018>
- [35] BATES DO, MORRIS JC, OLTEAN S, DONALDSON LF. Pharmacology of Modulators of Alternative Splicing. *Pharmacol Rev* 2017; 69: 63–79. <https://doi.org/10.1124/pr.115.011239>
- [36] ZHOU J, CHNG WJ. Aberrant RNA splicing and mutations in spliceosome complex in acute myeloid leukemia. *Stem Cell Investig* 2017; 4: 6. <https://doi.org/10.21037/sci.2017.01.06>
- [37] CIEPLY B, CARSTENS RP. Functional roles of alternative splicing factors in human disease. *Wiley Interdiscip Rev RNA* 2015; 6: 311–326. <https://doi.org/10.1002/wrna.1276>
- [38] BEJAR R. Splicing Factor Mutations in Cancer. *Adv Exp Med Biol* 2016; 907: 215–228. https://doi.org/10.1007/978-3-319-29073-7_9
- [39] SVEEN A, KILPINEN S, RUUSULEHTO A, LOTHE RA, SKOTHEIM RI. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene* 2016; 35: 2413–2427. <https://doi.org/10.1038/onc.2015.318>
- [40] RYAN M, WONG WC, BROWN R, AKBANI R, SU X et al. TCGASpliceSeq a compendium of alternative mRNA splicing in cancer. *Nucleic Acids Res* 2016; 44: D1018–1022. <https://doi.org/10.1093/nar/gkv1288>
- [41] BURKE K, MACKENZIE G. Multi-parameter regression survival modeling: An alternative to proportional hazards. *Biometrics* 2017; 73: 678–686. <https://doi.org/10.1111/biom.12625>
- [42] MOOLGAVKAR SH, CHANG ET, WATSON HN, LAU EC. An Assessment of the Cox Proportional Hazards Regression Model for Epidemiologic Studies. *Risk Anal* 2018; 38: 777–794. <https://doi.org/10.1111/risa.12865>
- [43] TANG Z, LI C, KANG B, GAO G, LI C et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* 2017; 45: W98–W102. <https://doi.org/10.1093/nar/gkx247>

- [44] YANG X, HUANG WT, HE RQ, MA J, LIN P et al. Determining the prognostic significance of alternative splicing events in soft tissue sarcoma using data from The Cancer Genome Atlas. *J Transl Med* 2019; 17: 283. <https://doi.org/10.1186/s12967-019-2029-6>
- [45] WU HY, PENG ZG, HE RQ, LUO B, MA J et al. Prognostic index of aberrant mRNA splicing profiling acts as a predictive indicator for hepatocellular carcinoma based on TCGA SpliceSeq data. *Int J Oncol* 2019; 55: 425–438. <https://doi.org/10.3892/ijo.2019.4834>
- [46] HUANG ZG, HE RQ, MO ZN. Prognostic value and potential function of splicing events in prostate adenocarcinoma. *Int J Oncol* 2018; 53: 2473–2487. <https://doi.org/10.3892/ijo.2018.4563>
- [47] LEX A, GEHLENBORG N, STROBELT H, VUILLEMOT R, PFISTER H. UpSet: Visualization of Intersecting Sets. *IEEE Trans Vis Comput Graph* 2014; 20: 1983–1992. <https://doi.org/10.1109/TVCG.2014.2346248>
- [48] DONCHEVA NT, MORRIS J, GORODKIN J, JENSEN LJ. Cytoscape stringApp: Network analysis and visualization of proteomics data. *J Proteome Res* 2019; 18: 623–632. <https://doi.org/10.1021/acs.jproteome.8b00702>
- [49] SZKLARCZYK D, MORRIS JH, COOK H, KUHN M, WYDER S et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 2017; 45: D362–D368. <https://doi.org/10.1093/nar/gkw937>
- [50] MA H, GAO G, WEBER GM. Use of DAVID algorithms for clustering custom annotated gene lists in a non-model organism, rainbow trout. *BMC Res Notes* 2018; 11: 63. <https://doi.org/10.1186/s13104-018-3154-7>
- [51] SAEZ B, WALTER MJ, GRAUBERT TA. Splicing factor gene mutations in hematologic malignancies. *Blood* 2017; 129: 1260–1269. <https://doi.org/10.1182/blood-2016-10-692400>
- [52] PIVA F, GIULIETTI M, BURINI AB, PRINCIPATO G. SpliceAid 2: a database of human splicing factors expression data and RNA target motifs. *Hum Mutat* 2012; 33: 81–85. <https://doi.org/10.1002/humu.21609>
- [53] JIN Y, DONG H, SHI Y, BIAN L. Mutually exclusive alternative splicing of pre-mRNAs. *Wiley Interdiscip Rev RNA* 2018; 9: e1468. <https://doi.org/10.1002/wrna.1468>
- [54] CLIMENTE-GONZALEZ H, PORTA-PARDO E, GODZIK A, EYRAS E. The Functional Impact of Alternative Splicing in Cancer. *Cell Rep* 2017; 20: 2215–2226. <https://doi.org/10.1016/j.celrep.2017.08.012>
- [55] QI Y, YU J, HAN W, FAN X, QIAN H et al. A splicing isoform of TEAD4 attenuates the Hippo-YAP signalling to inhibit tumour proliferation. *Nat Commun* 2016; 7: ncomms11840. <https://doi.org/10.1038/ncomms11840>
- [56] TIEN JF, MAZLOOMIAN A, CHENG SG, HUGHES CS, CHOW CCT et al. CDK12 regulates alternative last exon mRNA splicing and promotes breast cancer cell invasion. *Nucleic Acids Res* 2017; 45: 6698–6716. <https://doi.org/10.1093/nar/gkx187>
- [57] CHEN L, YAO Y, SUN L, ZHOU J, MIAO M et al. Snail Driving Alternative Splicing of CD44 by ESRP1 Enhances Invasion and Migration in Epithelial Ovarian Cancer. *Cell Physiol Biochem* 2017; 43: 2489–2504. <https://doi.org/10.1159/000484458>
- [58] WAN L, YU W, SHEN E, SUN W, LIU Y et al. SRSF6-regulated alternative splicing that promotes tumour progression offers a therapy target for colorectal cancer. *Gut* 2017; 68: 118–129. <https://doi.org/10.1136/gutjnl-2017-314983>
- [59] ADAMOPOULOS PG, KONTOS CK, TSIAKANIKAS P, SCORILAS A. Identification of novel alternative splice variants of the BCL2L12 gene in human cancer cells using next-generation sequencing methodology. *Cancer Lett* 2016; 373: 119–129. <https://doi.org/10.1016/j.canlet.2016.01.019>
- [60] DELIGIO JT, LIN G, CHALFANT CE, PARK MA. Splice variants of cytosolic polyadenylation element-binding protein 2 (CPEB2) differentially regulate pathways linked to cancer metastasis. *J Biol Chem* 2017; 292: 17909–17918. <https://doi.org/10.1074/jbc.M117.810127>
- [61] KOZLOVSKI I, SIEGFRIED Z, AMAR-SCHWARTZ A, KARNI R. The role of RNA alternative splicing in regulating cancer metabolism. *Hum Genet* 2017; 136: 1113–1127. <https://doi.org/10.1007/s00439-017-1803-x>
- [62] SIEGFRIED Z, KARNI R. The role of alternative splicing in cancer drug resistance. *Curr Opin Genet Dev* 2018; 48: 16–21. <https://doi.org/10.1016/j.cde.2017.10.001>
- [63] LEIVONEN SK, TASKINEN M, CERVERA A, KARJALAINEN-LINDSBERG ML, DELABIE J et al. Alternative splicing discriminates molecular subtypes and has prognostic impact in diffuse large B-cell lymphoma. *Blood Cancer J* 2017; 7: e596. <https://doi.org/10.1038/bcj.2017.71>
- [64] SAKUMA K, SASAKI E, KIMURA K, KOMORI K, SHIMIZU Y et al. HNRNPLL, a newly identified colorectal cancer metastasis suppressor, modulates alternative splicing of CD44 during epithelial-mesenchymal transition. *Gut* 2018; 67: 1103–1111. <https://doi.org/10.1136/gutjnl-2016-312927>
- [65] LI Y, SUN N, LU Z, SUN S, HUANG J et al. Prognostic alternative mRNA splicing signature in non-small cell lung cancer. *Cancer Lett* 2017; 393: 40–51. <https://doi.org/10.1016/j.canlet.2017.02.016>
- [66] HE RQ, ZHOU XG, YI QY, DENG CW, GAO JM et al. Prognostic Signature of Alternative Splicing Events in Bladder Urothelial Carcinoma Based on Spliceseq Data from 317 Cases. *Cell Physiol Biochem* 2018; 48: 1355–1368. <https://doi.org/10.1159/000492094>
- [67] ZHANG R, LIN P, YANG X, HE RQ, WU HY et al. Survival associated alternative splicing events in diffuse large B-cell lymphoma. *Am J Transl Res* 2018; 10: 2636–2647.
- [68] LIN P, HE RQ, MA FC, LIANG L, HE Y et al. Systematic Analysis of Survival-Associated Alternative Splicing Signatures in Gastrointestinal Pan-Adenocarcinomas. *EBioMedicine* 2018; 34: 46–60. <https://doi.org/10.1016/j.ebiom.2018.07.040>
- [69] STENGEL A, KERN W, HAFERLACH T, MEGGENDORFER M, FASAN A et al. The impact of TP53 mutations and TP53 deletions on survival varies between AML, ALL, MDS and CLL: an analysis of 3307 cases. *Leukemia* 2017; 31: 705–711. <https://doi.org/10.1038/leu.2016.263>
- [70] ZHANG S, ZHANG Q, SHI G, YIN J. MiR-182-5p regulates BCL2L12 and BCL2 expression in acute myeloid leukemia as a potential therapeutic target. *Biomed Pharmacother* 2018; 97: 1189–1194. <https://doi.org/10.1016/j.biopha.2017.11.002>

- [71] MOORE AS, ALONZO TA, GERBING RB, LANGE BJ, HEEREMA NA et al. BIRC5 (survivin) splice variant expression correlates with refractory disease and poor outcome in pediatric acute myeloid leukemia: a report from the Children's Oncology Group. *Pediatr Blood Cancer* 2014; 61: 647–652. <https://doi.org/10.1002/pbc.24822>
- [72] DORON B, ABDELHAMED S, BUTLER JT, HASHMI SK, HORTON TM et al. Transmissible ER stress reconfigures the AML bone marrow compartment. *Leukemia* 2018; 33: 918–930. <https://doi.org/10.1038/s41375-018-0254-2>
- [73] ZHOU C, MARTINEZ E, DI MARCANTONIO D, SOLANKI-PATEL N, AGHAYEV T et al. JUN is a key transcriptional regulator of the unfolded protein response in acute myeloid leukemia. *Leukemia* 2017; 31: 1196–1205. <https://doi.org/10.1038/leu.2016.329>
- [74] BECK M, HURT E. The nuclear pore complex: understanding its function through structural insight. *Nat Rev Mol Cell Biol* 2017; 18: 73–89. <https://doi.org/10.1038/nrm.2016.147>
- [75] D'ANGELO MA. Nuclear pore complexes as hubs for gene regulation. *Nucleus* 2018; 9: 142–148. <https://doi.org/10.1080/19491034.2017.1395542>
- [76] RAICES M, D'ANGELO MA. Nuclear pore complexes and regulation of gene expression. *Curr Opin Cell Biol* 2017; 46: 26–32. <https://doi.org/10.1016/j.ceb.2016.12.006>
- [77] SAITO S, CIGDEM S, OKUWAKI M, NAGATA K. Leukemia-Associated Nup214 Fusion Proteins Disturb the XPO1-Mediated Nuclear-Cytoplasmic Transport Pathway and Thereby the NF-kappaB Signaling Pathway. *Mol Cell Biol* 2016; 36: 1820–1835. <https://doi.org/10.1128/MCB.00158-16>
- [78] AJIRO M, ZHENG ZM. Oncogenes and RNA splicing of human tumor viruses. *Emerg Microbes Infect* 2014; 3: e63. <https://doi.org/10.1038/emi.2014.62>
- [79] NG SW, MITCHELL A, KENNEDY JA, CHEN WC, MCLEOD J et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* 2016; 540: 433–437. <https://doi.org/10.1038/nature20598>
- [80] WANG F, TIAN X, ZHOU J, WANG G, YU W et al. A three-lncRNA signature for prognosis prediction of acute myeloid leukemia in patients. *Mol Med Rep* 2018; 18: 1473–1484. <https://doi.org/10.3892/mmr.2018.9139>
- [81] ZHU R, LIN W, ZHAO W, FAN F, TANG L et al. A 4-microRNA signature for survival prognosis in pediatric and adolescent acute myeloid leukemia. *J Cell Biochem* 2018; 120: 3958–3968. <https://doi.org/10.1002/jcb.27679>
- [82] GALEJ WP. Structural studies of the spliceosome: past, present and future perspectives. *Biochem Soc Trans* 2018; 46: 1407–1422. <https://doi.org/10.1042/BST20170240>
- [83] PLASCHKA C, LIN PC, CHARENTON C, NAGAI K. Pre-spliceosome structure provides insights into spliceosome assembly and regulation. *Nature* 2018; 559: 419–422. <https://doi.org/10.1038/s41586-018-0323-8>
- [84] JENKINS JL, KIELKOPF CL. Splicing Factor Mutations in Myelodysplasias: Insights from Spliceosome Structures. *Trends Genet* 2017; 33: 336–348. <https://doi.org/10.1016/j.tig.2017.03.001>
- [85] ANCZUKOW O, KRAINER AR. Splicing-factor alterations in cancers. *RNA* 2016; 22: 1285–1301. <https://doi.org/10.1261/rna.057919.116>
- [86] URBANSKI LM, LECLAIR N, ANCZUKOW O. Alternative-splicing defects in cancer: Splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics. *Wiley Interdiscip Rev RNA* 2018; 9: e1476. <https://doi.org/10.1002/wrna.1476>
- [87] PHAM D, KOIDE K. Discoveries, target identifications, and biological applications of natural products that inhibit splicing factor 3B subunit 1. *Nat Prod Rep* 2016; 33: 637–647. <https://doi.org/10.1039/c5np00110b>

Prognostic alternative splicing regulatory network of splicing events in acute myeloid leukemia patients based on SpliceSeq data from 136 cases

Z. C. XIE¹, L. GAO², G. CHEN², J. MA¹, L. H. YANG¹, R. Q. HE¹, M. W. LI¹, K. T. CAI¹, T. T. LI¹, Z. G. PENG^{1,*}

Supplementary Information

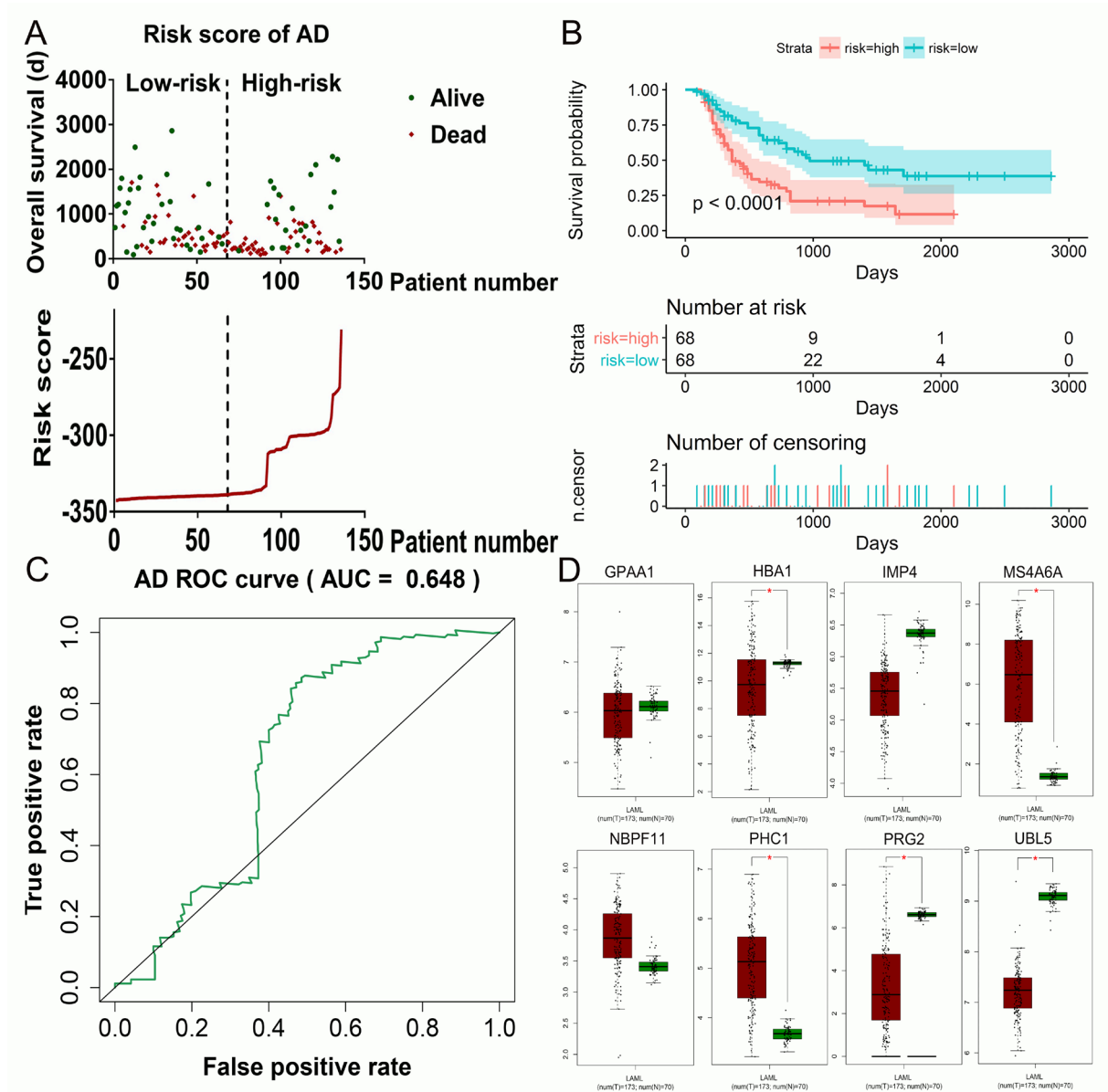


Figure S1. Construction of prognostic signature of the alternate donor site (AD). A) Information about patients' survival based on their risk scores. The upper part of the graph displays patients' survival status divided by risk score, and the bottom part displays the linear variation in the risk score. B) Kaplan-Meier survival curve for low/high-risk groups based on the risk score. The low-risk group shows better survival than the high-risk group. C) Receiver operating characteristic (ROC) curves revealing the predictive capacity of the area under the curve at 0.648. D) Expression of each gene symbol and prognostic signatures based on TCGA and GTEx data. The statistically significant difference in gene expression between the tumor and normal samples is marked with an asterisk.

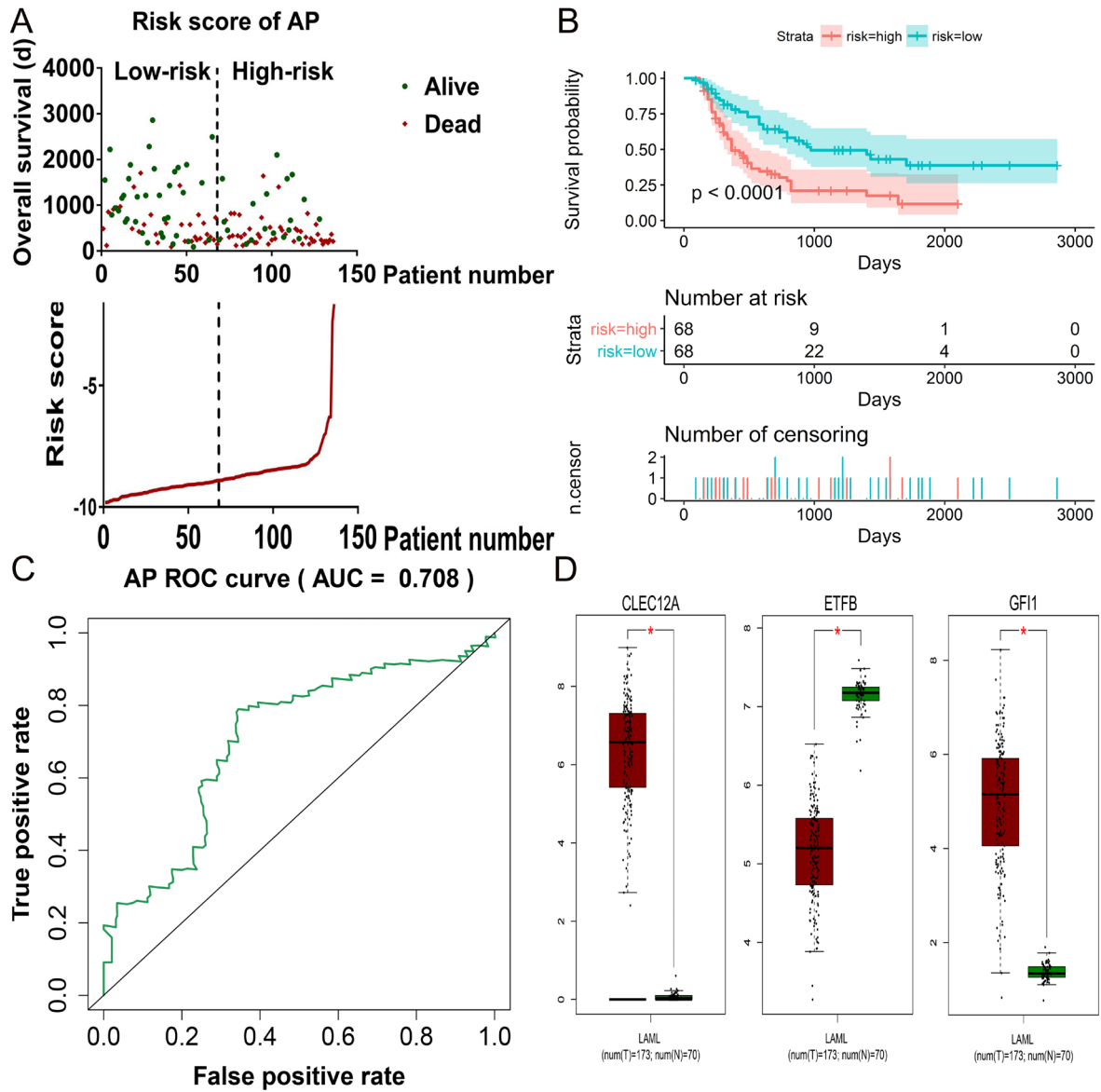


Figure S2. Construction of the prognostic signature of the alternate promoter (AP). A) Information about patients' survival based on the risk score. The upper part of the graph displays patients' survival status divided by the risk score, and the bottom part displays the linear variation in the risk score. B) Kaplan-Meier survival curve for low/high-risk groups based on risk score. The low-risk group shows better survival than the high-risk group. C) Receiver operating characteristic (ROC) curves revealing the predictive capacity of the area under the curve at 0.708. D) Expression of each gene symbol and prognostic signatures based on TCGA and GTEx data. The statistically significant difference in gene expression between the tumor and normal samples is marked with an asterisk.

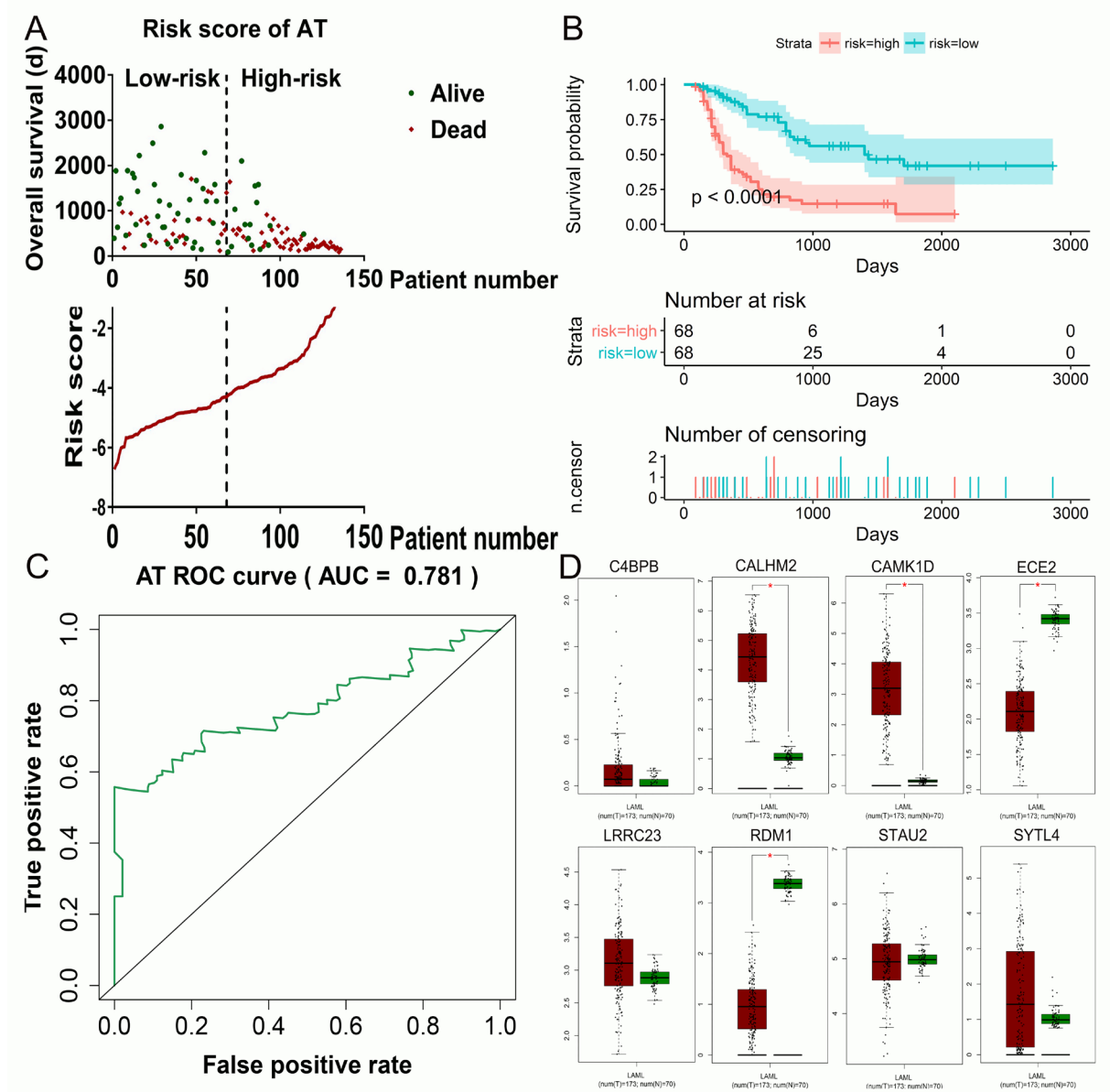


Figure S3. Construction of the prognostic signature of the alternate terminator (AT). A) Information about patients' survival based on the risk score. The upper part of the graph displays patients' survival status divided by the risk score, and the bottom part displays the linear variation in the risk score. B) Kaplan-Meier survival curve for low/high-risk groups based on the risk score. The low-risk group shows better survival than the high-risk group. C) Receiver operating characteristic (ROC) curves revealing the predictive capacity of the area under the curve at 0.781. D) Expression of each gene symbol and prognostic signatures based on TCGA and GTEx data. The statistically significant difference in gene expression between the tumor and normal samples is marked with an asterisk.

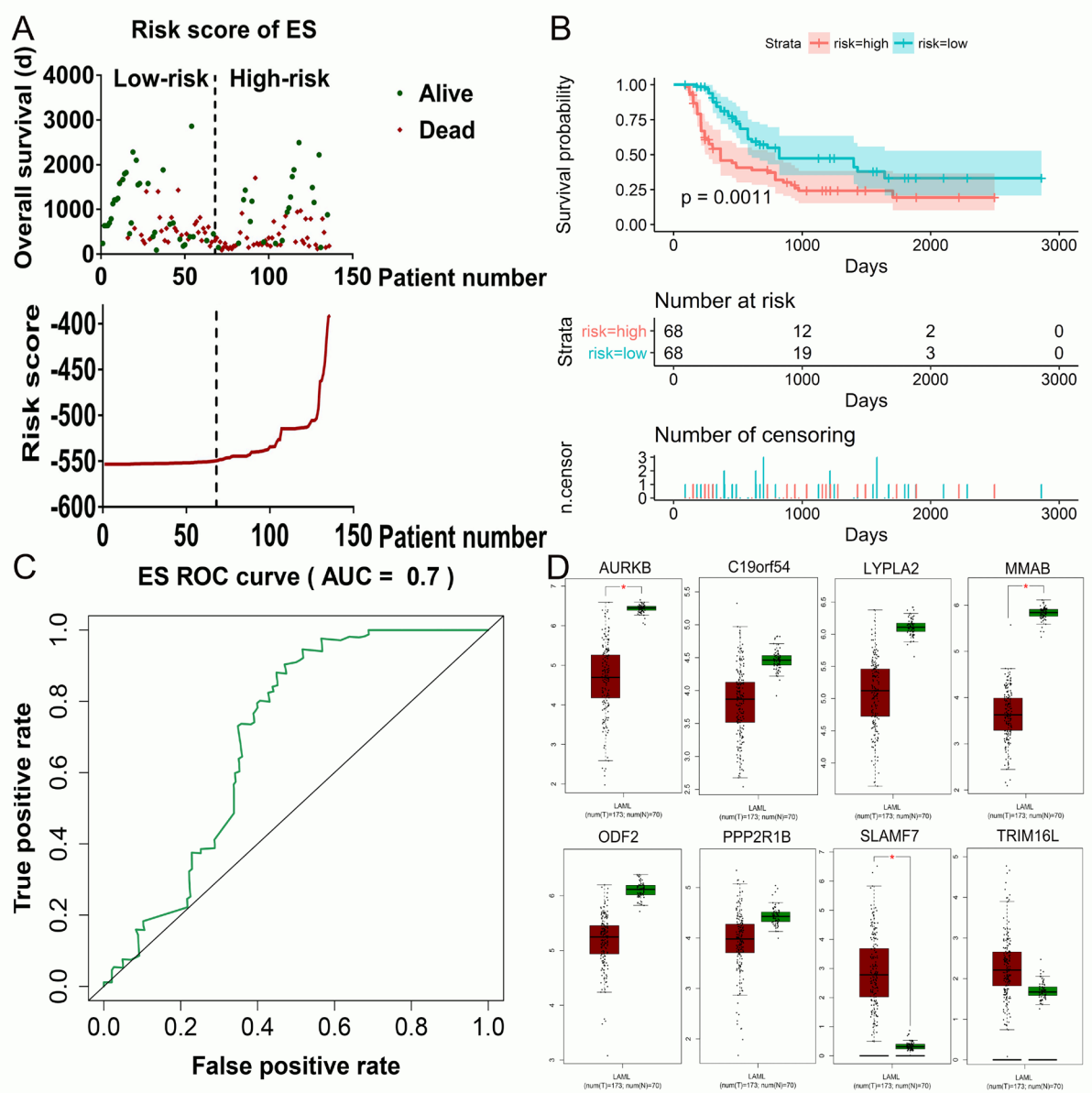


Figure S4. Construction of the prognostic signature of exon skip (ES). A) Information about patients' survival based on the risk score. The upper part of the graph displays patients' survival status divided by the risk score, and the bottom part displays the linear variation in the risk score. B) Kaplan-Meier survival curve for low/high-risk groups based on the risk score. The low-risk group shows better survival than the high-risk group. C) Receiver operating characteristic (ROC) curves revealing the predictive capacity of the area under the curve at 0.7. D) Expression of each gene symbol and prognostic signatures based on TCGA and GTEx data. The statistically significant difference in gene expression between the tumor and normal samples is marked with an asterisk.

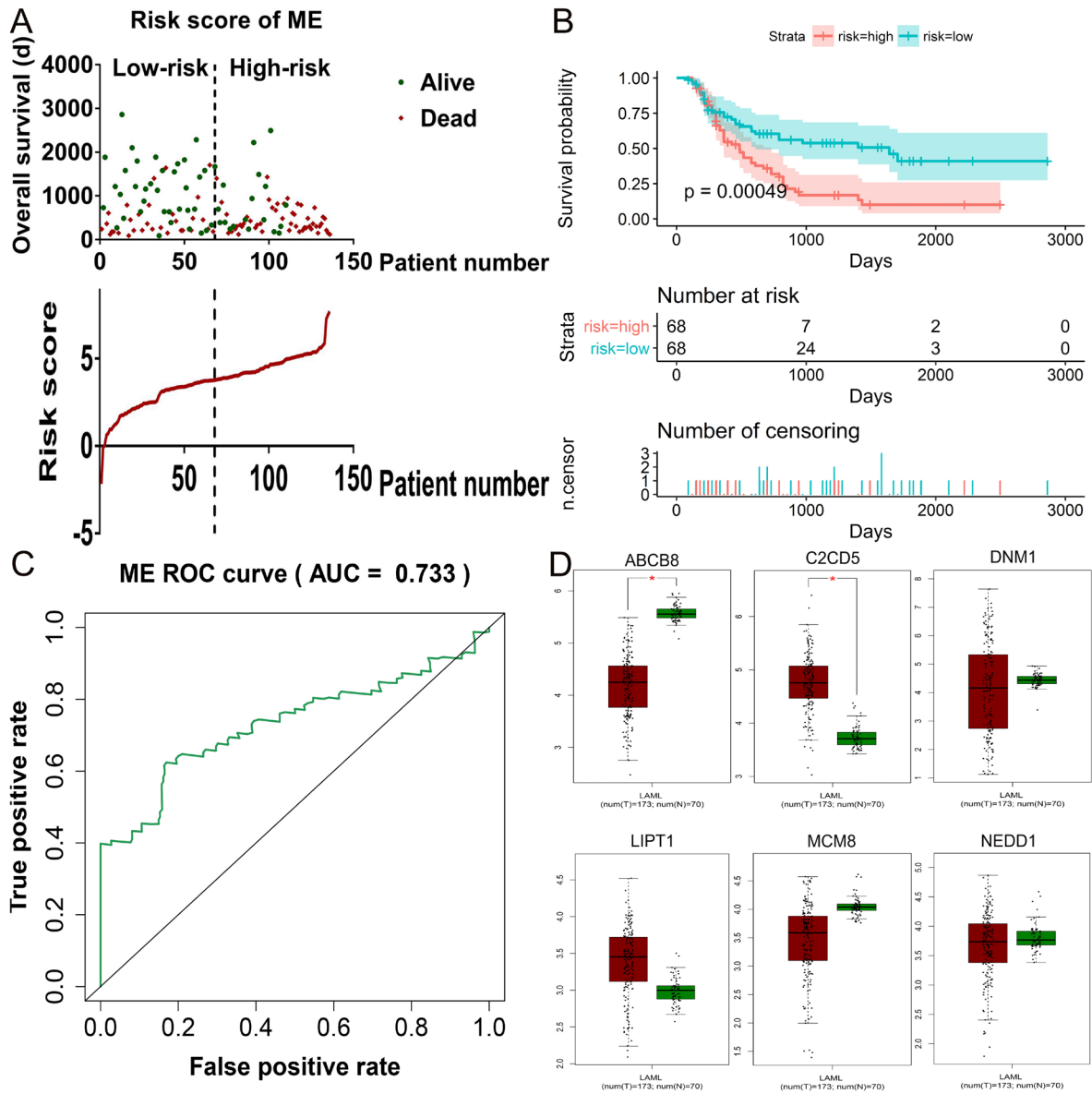


Figure S5. Construction of the prognostic signature of mutually exclusive exons (ME). A) Information about patients' survival based on the risk score. The upper part of the graph displays the survival status divided by the risk score, and the bottom part displays the linear variation of the risk score. B. Kaplan-Meier survival curve for low/high-risk groups based on the risk score. The low-risk group shows better survival than the high-risk group. C) Receiver operating characteristic (ROC) curves revealing the predictive capacity of the area under the curve at 0.733. D) Expression of each gene symbol and prognostic signatures based on TCGA and GTEx data. The statistically significant difference in gene expression between the tumor and normal samples is marked with an asterisk.