

Prognostic alternative mRNA splicing signatures and associated splicing factors in acute myeloid leukemia

Peng Jin^a; Yun Tan^{a,b}; Wei Zhang^{a,c}; Junmin Li^a; Kankan Wang^{a,b,*}

^a Shanghai Institute of Hematology, State Key Laboratory of Medical Genomics, National Research Center for Translational Medicine at Shanghai, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China; ^b Sino-French Research Center for Life Sciences and Genomics, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China ^c School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

Abstract

The dysregulation of alternative splicing (AS) has emerged as a mechanism of acute myeloid leukemia (AML). However, the prognostic impact of AS events remains under-explored in AML. Here we report the prognostic value of AS events and associated splicing factors based on three datasets of AML patients. We defined the landscape of AS events in AML and identified 7033 AS events associated with the survival of AML patients. Based on these events, we further developed a composite 15 AS event-based prognostic signature, which was independent of the cytogenetic risk stratification and patient age, and showed a better performance than known gene expression signatures. More importantly, our new signature markedly improved the European LeukemiaNet (ELN) risk classification, indicating a broad applicability in the clinical management of AML. Furthermore, the splicing-regulatory network established the correlations between prognostic AS events and associated splicing factors. The finding was validated by CRISPR-based data, which indicated that the increased expression of RBM39 contributed to the higher exon inclusion of SETD5 and conferred a poor outcome. Together, AS events may serve as a novel assortment of prognosticators for AML and could refine the ELN risk stratification. The splicing regulatory network provides clues regarding the splicing factor-mediated mechanisms of AML.

Neoplasia (2020) 22 447–457

Keywords: Acute myeloid leukemia, Alternative splicing, Prognosis, European LeukemiaNet, Splicing factor

Introduction

Acute myeloid leukemia (AML) is the most common form of acute leukemia in adults. This disease represents a heterogeneous entity characterized by the aggressive proliferation of immature myeloid progenitor cells primarily caused by an interplay of genetic and epigenetic aberrations [1]. Because the 5-year overall survival rate of AML patients remains as low as 30% [1,2], there is an urgent need to improve the risk stratification of

AML so as to prolong survival. Advances in high-throughput sequencing technology make it realistic to develop molecular parameter-based signatures to improve risk stratification of AML patients [3–6]. However, these studies would have missed capturing an important intrinsic biological feature of AML, alternative splicing (AS). The consideration of AS could theoretically further improve the risk stratification and prognosis of AML patients.

As a critical determinant of transcriptome and proteome diversity, AS is a major driver of regulatory complexity and functional versatility in eukaryotes [7]. According to the GENCODE annotation, approximately 95% of multi-exon human genes are alternatively spliced, and around 20,000 human protein-coding genes produce more than 80,000 distinct mRNA variants [8]. Subtle changes in protein-coding genes owing to AS can generate profound effects on the biological characteristics of translated proteins, which may alter protein localization signals and functional protein domains, thereby modifying protein–protein interactions [9,10]. Indeed, AS-related alterations are emerging as important events in the

Abbreviations: AML, Acute myeloid leukemia, AS, Alternative splicing, A3SS, Alternative 3' splice site, A5SS, Alternative 5' splice site, AUC, Area under the curve, CRISPR, Clustered regularly interspaced short palindromic repeats, ELN, European Leukemia Net, HR, Hazard ratio, MXE, Mutually exclusive exons, OS, overall survival, PSI, Percent spliced in, RI, Retained intron, ROC, Receiver operating characteristic, RBSURV, robust likelihood-based survival modeling, RBM39, RNA Binding Motif Protein 39, SE, Skipped exon, SETD5, SET Domain Containing 5, TCGA, The Cancer Genome Atlas

* Corresponding author at: State Key Laboratory of Medical Genomics and Shanghai Institute of Hematology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, 197 Ruijin Er Rd, Shanghai 200025, China. Fax: +86 21 64743206. e-mail address: kankanwang@shsmu.edu.cn (K. Wang).

© 2020 The Author(s). Published by Elsevier Inc. on behalf of SOCIETY. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).
<https://doi.org/10.1016/j.neo.2020.06.004>

development and progress of cancer and, if fully characterized, could be promising biomarkers with prominent prognostic values.

A recent study discovered a distinct AML subgroup, defined by mutations in genes encoding chromatin and/or the spliceosome, that has a poor prognosis [11]. These spliceosomal gene mutations or their deregulated expressions can recurrently generate effects on specific amino acid residues, resulting in altered splice sites and perturbed exon recognition, which finally leads to mis-splicing [10,12,13]. These findings strongly support the hypothesis that aberrant AS is a fundamental aspect of AML pathogenesis. However, a systematic analysis of AS events has not been undertaken nor has their utility as prognostic markers been thoroughly explored for AML.

In the present study, we used an unbiased genome-wide approach to investigate the prognostic values of AS events in AML. We first created a catalog of prognostic AS events, from which we further developed prognostic signatures that had improved performances compared with frequently used signatures based on other genetic events, including mRNA, long non-coding RNA (lncRNA) and microRNA (miRNA). We also proposed a splicing-regulatory network to understand the mechanisms underlying the splicing factor-mediated AS events and used loss-of-function experiment data to validate these correlations.

Materials and methods

Data collection and preprocessing

The RNA-seq raw data and clinical data of three AML cohorts including the TCGA-LAML cohort [2] ($N = 151$, training set), the BeatAML cohort [14] ($N = 430$, validation set) and the TARGET-AML cohort [15] ($N = 179$, validation set) were downloaded from Genomic Data Commons. RNA-seq reads aligned to the human reference genome (hg38) were applied to analyze AS events using rMATS [16], which recognizes five types of AS events, skipped exons (SE), alternative 5' splice sites (A5SS), alternative 3' splice sites (A3SS), mutually exclusive exons (MXE) and retained introns (RI), and unambiguously calculates the Percent-Spliced-In (PSI) values for splicing events. Functional enrichment analysis was performed using metasplice [17]. Genome-wide sgRNA raw counts for 12 human AML cell lines (P31/FUJ, NB4, OCI-AML2, OCI-AML3, OCI-AML5, SKM-1, EOL-1, HEL, Molm-13, MonoMac1, MV4;11 and PL-21) were available and downloaded from <http://sabatinilab.wi.mit.edu/wang/2017/> [18]. The fold-change of each splicing factor gene was calculated according to the below formula:

$$\text{Fold change} = \text{average} \left[\left(\frac{\text{initial sgRNA abundance}}{\text{final sgRNA abundance}} \right) \right],$$

where a gene with a larger fold change demonstrates more essentiality. The expression data of splicing factor genes in AML ($N = 477$) and healthy control subjects ($N = 33$) were downloaded from the Genomic Data Commons (BeatAML project level 3 data). RNA-seq data before and after knocking out RBM39 was obtained from [13] (GSE114558).

Identification of prognostic AS events

AML patients were divided into two groups using the median PSI values of AS events, and the associations between AS events and the overall survival times of patients were subjected to a univariate Cox regression analysis in the Survival R package [19]. Candidate AS events were those with p -value < 0.05 . All candidate AS events were categorized by type. For each type, robust likelihood-based survival (RBSURV) models were built to identify the key AS events influencing the prognosis of AML utilizing the RBSURV package [20]. The detailed procedure was as follows:

(i). We randomly divided samples into the training set and the validation set (sample size, training set: validation set = 2:1). An AS event was first fitted to the training set of samples to generate the estimated parameters for this event. Then we evaluated log-likelihood based on the estimated parameters and the validation set of samples. This evaluation was repeated for each AS event. (ii). We performed the above procedure 10 times, thus obtaining 10 log-likelihoods for each event. The best AS event, $a(1)$, with the largest mean log-likelihood was selected. (iii). We searched the next best event by evaluating every two-event model, $a(1) + a(2)$, and selected an optimal one with the largest mean log-likelihood. (iv). We continued this stepwise forward AS event selection procedure, generating a series of models: $a(1)$, $a(1) + a(2)$, $a(1) + a(2) + a(3)$, ... We computed Akaike information criteria (AICs) for all the candidate models and finally selected an optimal model with the minimal AIC.

To build the composite prognostic predictor containing all types of AS events, five types of prognostic AS events (p -value < 0.05) were further combined and subjected to the RBSURV modeling to jointly leverage the information of AS events across types and search for the optimal number without being trapped in the type-specific marginal optimal events.

Development of prognostic splicing signatures

Risk-based models were established according to the following formula:

$$\text{Risk Score} = \sum_{i=1}^n (\text{PSI}_i \times \text{Coef}_i)$$

where n represents the number of AS events, PSI_i represents the exon-inclusion level of the AS event (PSI value), and Coef_i represents the estimated regression coefficient value.

The prediction capabilities of the splicing signatures were assessed by a time-dependent receiver operating characteristic (ROC) curve with the area under the curve (AUC) value using the survivalROC package [21].

Prediction on the protein function affected by alternative splicing

We first mapped the chromosomal coordinates of each prognostic AS event (original splice junction, e.g., skipped exon, upstream exon and downstream exon) to the respective transcript, which was determined as the canonical isoform. Next, we removed the alternatively spliced exon (e.g. the skipped exon in the case of a skipped exon event) from the canonical transcript to generate the alternative spliced isoform. Then, we retrieved the FASTA sequences of the canonical isoform and the alternative isoform using GffRead (<https://github.com/gperrea/gffread>). The FASTA sequences were then translated, and were used to predict functional structures and important sites using InterProScan [22] (run in nucleotide mode). Finally, we parsed and compared the InterProScan outputs to identify loss/gain of protein structures and functional sites after splicing. To determine whether the alternatively spliced isoform is a previously identified or a novel isoform, we removed the alternatively spliced exon (e.g. the skipped exon in the case of a skipped exon event) from the original splice junction, which was defined as the alternative splice junction. If the alternative splice junction is not mapped to a known transcript, indicating that the alternatively spliced isoform is novel, otherwise it was a previously identified isoform.

Computing previously published AML prognostic scores

To validate the performance of the splicing signature, three recently reported gene expression-based prognostic scores were calculated as previously described: the Gene-24 [4], the Gene-7 [5], and the LSC-17 [6].

Regulatory network combining splicing factors with AS events

Differential expression analysis on genes involved in the mRNA-splicing was performed using DESeq2 [23] (477 AML *versus* 33 healthy control subjects) to find out the dysregulated splicing factor genes in AML, and the genes with significantly altered expression were further screened out using CRISPR-based screening data in a panel of AML cell lines. We determined the essentiality of splicing factor genes as the average fold-change of sgRNA abundance in the 12 AML cell lines and used a stringent cut-off value of average fold-change >1.5 to identify essential splicing factor genes. Spearman's rank correlation analysis was carried out to correlate gene expression of splicing factors with quantifications of AS events. The splicing regulatory network was visualized using Cytoscape [24].

Statistical analysis

All the statistical analysis was performed using the R/Bioconductor statistical environment. The Wilcoxon rank-sum test was used to determine the statistical significance of differences between groups, and *p*-values <0.05 indicated significant differences. Data were presented as the means \pm standard deviations.

Results

AS events are informative for AML prognosis

To systematically identify prognostic AS events in AML, we first defined the landscape of AS events expressed in AML using the RNA-seq data of 151 AML patients from TCGA. Totally, we identified over 300,000 different cassette exons across AML patients, including SE, MXE, A3SS, A5SS and RI (Fig. 1a). We also observed that a single gene could produce multiple types of AS events (Fig. 1b), confirming the importance of AS in diversifying the AML transcriptome. A set of 100,185 high-confidence events were finally generated after filtering for commonality (events detected in $\geq 80\%$ of all samples) and cross-sample variance (range of PSI > 5%; averaged skipping or inclusion level >5%) (Fig. 1a). We focused on the prognostic values of these high-confidence events.

Next, we integrated the clinical data of TCGA-LAML patients and identified 7,033 significant prognostic AS events (*p*-value <0.05) that were derived from 3,861 host genes (Fig. 1c). We found that the expression of the corresponding host genes was much less informative than AS events in prognosis (Fig. 1d), indicating that the identified prognostic AS events were not the result of the expression level alterations of host genes. Further pathway and process enrichment analysis suggested an enrichment of genes in tumor-related functional categories such as DNA repair, cell cycle and histone modification (Fig. 1e). As an example, we performed more detailed analysis on the exon skipping of *HDAC7* in prognosis. We divided the patients into two groups using the median cut of the exon 9 inclusion level (PSI value) of *HDAC7* (Fig. 1f–g). The Kaplan–Meier curve and log–rank test showed that the higher inclusion level of this exon was associated with poorer prognosis (Fig. 1h).

Finally, we observed that four types of prognostic AS events could occur in one single gene (Supplementary Fig. S1a). For example, in the *PILRB* gene, there were A3SS, A5SS, MXE and SE events that were significantly associated with the overall survival of AML patients, but the expression of the gene itself was not associated with survival (Supplementary Fig. S1b–f). Collectively, we identified a rich source of AS events that can act as novel prognostic markers for AML patients.

Establishment of an AS-15 splicing signature having a prognostic value in AML

We next investigated the prognostic impact of each type of identified AS event. We used RBSURV to identify key AS events and prognostic functions for each type. Five signatures were developed, namely for A3SS, A5SS, MXE, RI and SE, consisting of 12, 12, 18, 10 and 13 key prognostic AS events, respectively (Fig. 2a–e, left panel). Using the key prognostic AS events specific to the AS event types, we performed a multivariate Cox regression analysis to comprehensively evaluate their collective prognostic use. Risk scores calculated from type-specific key prognostic AS events were capable of distinguishing the high-risk AML group having a relatively shorter survival time from the low-risk group having a relatively prolonged survival time (Fig. 2a–e, right panel, *p*-value <0.0001).

To reduce the bias caused by the use of a specific AS event type, a robust prognostic predictor should contain as many AS patterns/types as possible. We performed RBSURV modeling that considered all types of significant AS events (*p*-value <0.05) and identified a composite prognostic splicing signature that included 15 AS events (Fig. 2f, left panel). The schematic illustration of the 15 AS events were shown in Supplementary Fig. S2. Using an *in-silico* prediction (detailed in Methods), we found that 73% (11/15) of the 15 AS events altered protein structures or functional sites (Supplementary Table 1), suggesting the potential functional relevance of prognostic AS events. It was noteworthy that several AS events in the composite signature, such as SE of SETD5 and FBRSL1 and MXE of PCBP1-AS1, were not enrolled into the prognostic model for the individual type. Those type-specific top events were redundant and their information had already been captured by other first selected events. The forward gene selection strategy employed in RBSURV ensured that the combined splicing values of these 15 AS events were identified to minimize the model complexity while maintaining the maximum fit of the model to the data.

Finally, to interpret the prognostic value of these 15 AS events, an AS-15 risk score was generated using the PSI values and the coefficients from the multivariable Cox regression model. This score markedly distinguished high-risk and low-risk patients, and a further ROC analysis illustrated that the AS-15 signature resulted in a higher AUC, at 0.931 for 5-year overall survival, than the type-specific AS event-based SE, A3SS, A5SS, RI and MXE signatures (Fig. 2g). AS-15 remained significantly prognostic in the validation set, having AUC value of 0.785, which suggested a promising clinical application of AS events for AML patients (Fig. 2h–i).

The AS-15 signature is an independent prognosticator for AML

We further performed a multivariable Cox regression analysis to determine if AS-15 was an independent prognostic signature for AML patients. After the well-established factors, such as the cytogenetic risk stratification and patient age, were adjusted, the AS-15 score remained significantly associated with overall survival (adjusted *p*-value <0.001, hazard ratio (HR) = 2.505, 2.772 in the training and validation set, respectively, Supplementary Table 2). Within the three cytogenetic risk subgroups, the AS-15 score could substantially distinguish the high-risk patients from the low-risk patients (Supplementary Fig. S3). Similar results were observed when these analyses were performed separately for patients younger or older than 60 years of age in the training set and validation set (Fig. 3a–b).

To broaden the applicability of the AS-15 score in the clinical management of AML and determine whether the AS-15 score was also of clinical interest in pediatric AML, we analyzed the AS profiles of 179 pediatric AML patients and calculated the AS-15 score. As shown in Fig. 3c, the children having lower AS-15 scores had significantly better overall survival

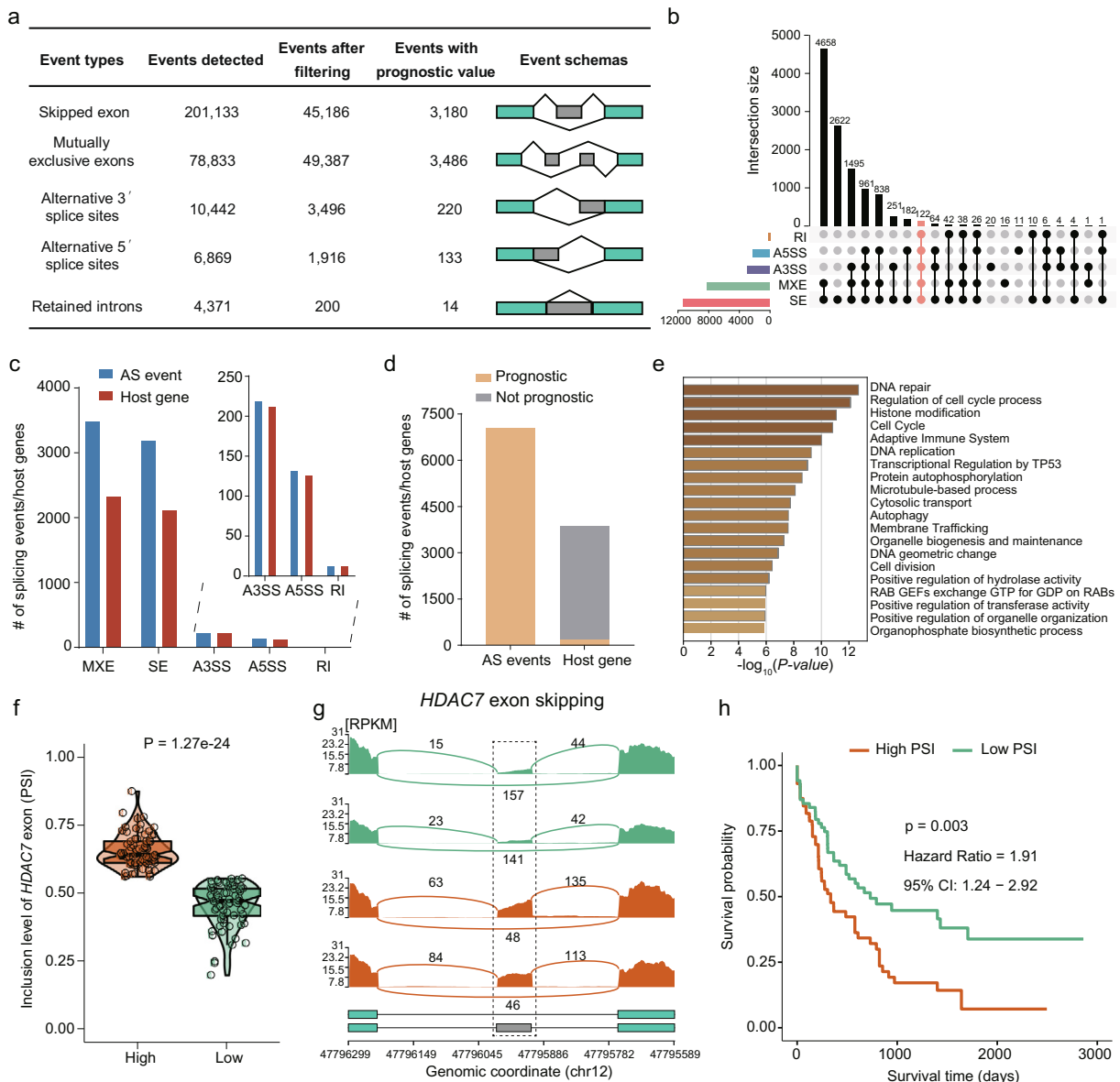


Fig. 1. AS events are informative for AML prognosis. (a) Summary table depicting various exon events detected by rMATS before and after filtering for PSI range and commonality as well as events with prognostic value. The number of defined AS events in the AML transcriptome. (b) UpSet plot of interactions among the five types of AS events in AML. One single gene may occur in all five types of AS events. (c) The distribution of the identified prognostic AS events and their host genes in the five AS event types. (d) The number of survival-associated AS events and their host genes. (e) Functional enrichment of genes associated with the prognostic AS events. (f) Scatterplot showing the two subgroups of AML patients divided by the median inclusion level of *HDCA7* exon 9 (PSI value). (g) Sashimi plots of an alternative exon skipping event in the *HDCA7* gene, with examples from representative patients shown. Lines connecting exons represent splice junctions and numbers on lines indicate the number of supporting sequencing reads. (h) Kaplan–Meier curves showing the survival differences using the PSI value of *HDCA7* splicing events in AML.

outcomes compared with children having higher AS-15 scores. This prognostic relevance was also present in the infant (<3 years) and adolescent (3–24 years) subgroups (Fig. 3d–e), suggesting that the AS-15 score may provide important clinical information in pediatric AML. Collectively, these findings demonstrated that the AS-15 score was a highly reliable independent prognosticator in both adult and pediatric AML patients.

The AS-15 splicing signature outperforms gene expression-based scoring models

Next, we compared the performance of the splicing signature with that of frequently used gene expression-based signatures. First, comparisons were

conducted between the AS-15 score and the expression signatures based on mRNA, lncRNA and miRNA using the expression profiles from the same cohort. Both the Kaplan–Meier model and ROC analysis demonstrated that AS-15 score performed better than those based on other genetic events (Fig. 4a–d). Second, we compared the AS-15 score with other rigorously tested powerful gene expression signatures, including Gene-24 [4], Gene-7 [5] and LSC-17 [6], which were indeed capable of predicting AML survival in a univariate analysis (Supplementary Fig. S4). The ROC curves showed that the AS-15 score outperformed these three gene expression-based signatures both in the training and validation sets (Fig. 4e–f). When combining these three gene-expression-based prognostic scores with our AS-15, as well as, cytogenetic risk stratification and patient age into one multivariate Cox

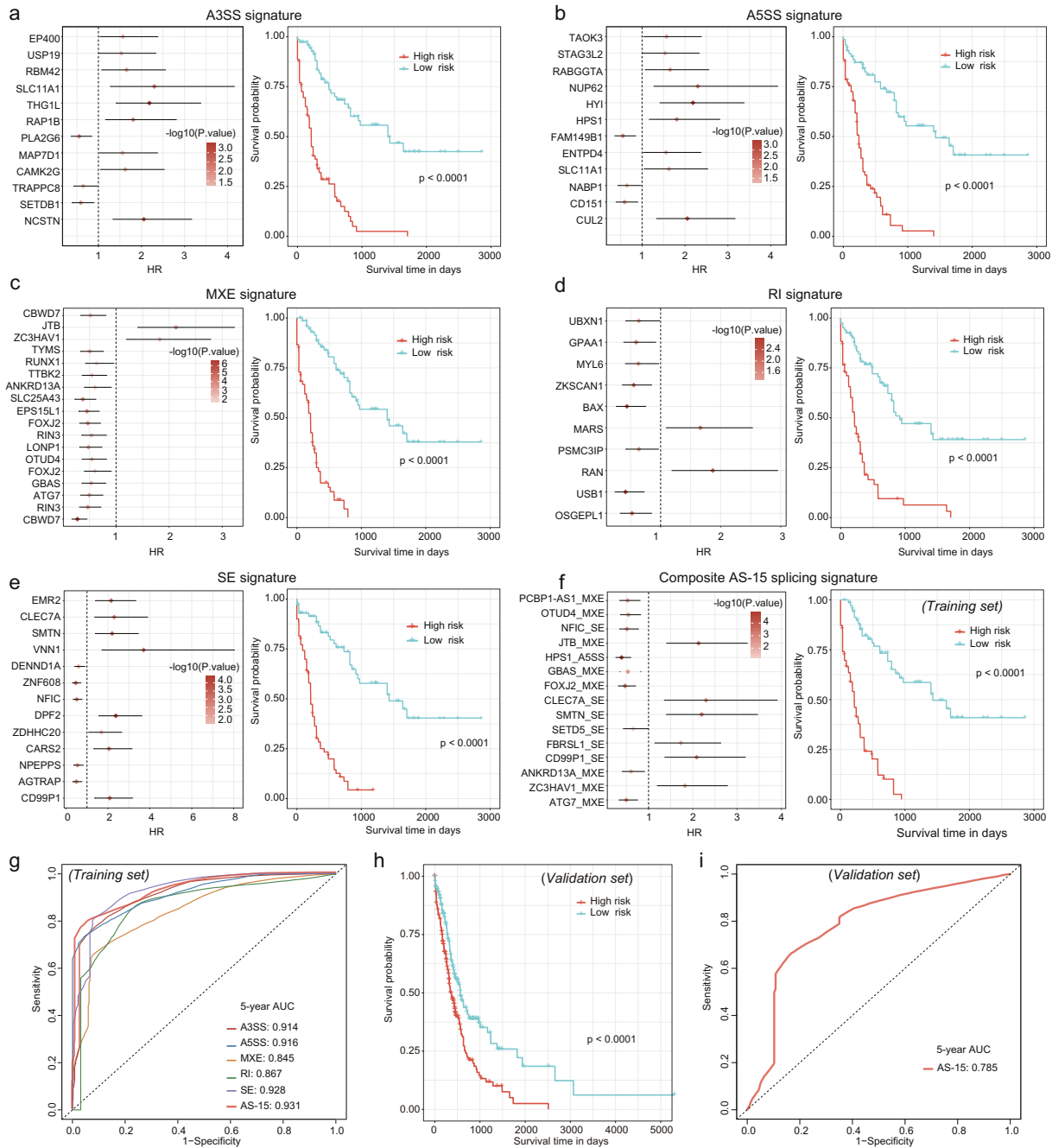


Fig. 2. Development of prognostic splicing signatures for AML. (a–e) AS event-based prognostic signatures in AML. Prognostic impact of each type of AS event: A3SS (a), A5SS (b), MXE (c), RI (d) and SE (e). The left panel shows the hazard ratios of the robust likelihood-based survival modeling regarding the identified prognostic AS events in the indicated AS event type. The right panel shows the Kaplan–Meier curve of the indicated AS type. (f) Prognostic impact of the composite signature consisting of 15 AS events in AML. (g) ROC curve analyses of the splicing signatures based on type-specific and composite AS events. Kaplan–Meier survival analysis of AML patients (h) and ROC curve analysis in the validation set (BeatAML cohort) (i).

regression model, only AS-15 remained independently predictive of the prognosis in the training set and validation set (Supplementary Table 3). Collectively, these findings illustrated that AS events are ideal clinical parameters for the risk stratification of AML and that the AS-15 signature was a highly reliable clinical tool.

The AS-15 signature substantially improves the European LeukemiaNet (ELN) risk classification of AML

The ELN classification considers a combination of cytogenetic and mutational data and is currently regarded as the gold standard for risk

stratification in AML [25]. A multivariable Cox regression analysis revealed that the AS-15 score was independent of the patient age and ELN classification (adjusted p -value < 0.001 , HR = 2.546, 2.705 in the training set and validation set, respectively, Supplementary Table 4). To further enhance the clinical applicability of this splicing signature, we first divided the entire study population having available ELN risk classification information into AS-15^{high} and AS-15^{low} groups using maximally selected rank statistics. Patients in the AS-15^{high} group had significantly shorter overall survival times than patients in the AS-15^{low} group (Supplementary Fig. S5a–b). Next, we investigated whether the AS-15 score could dichotomize survival within the ELN risk groups. Stratifying the patients within

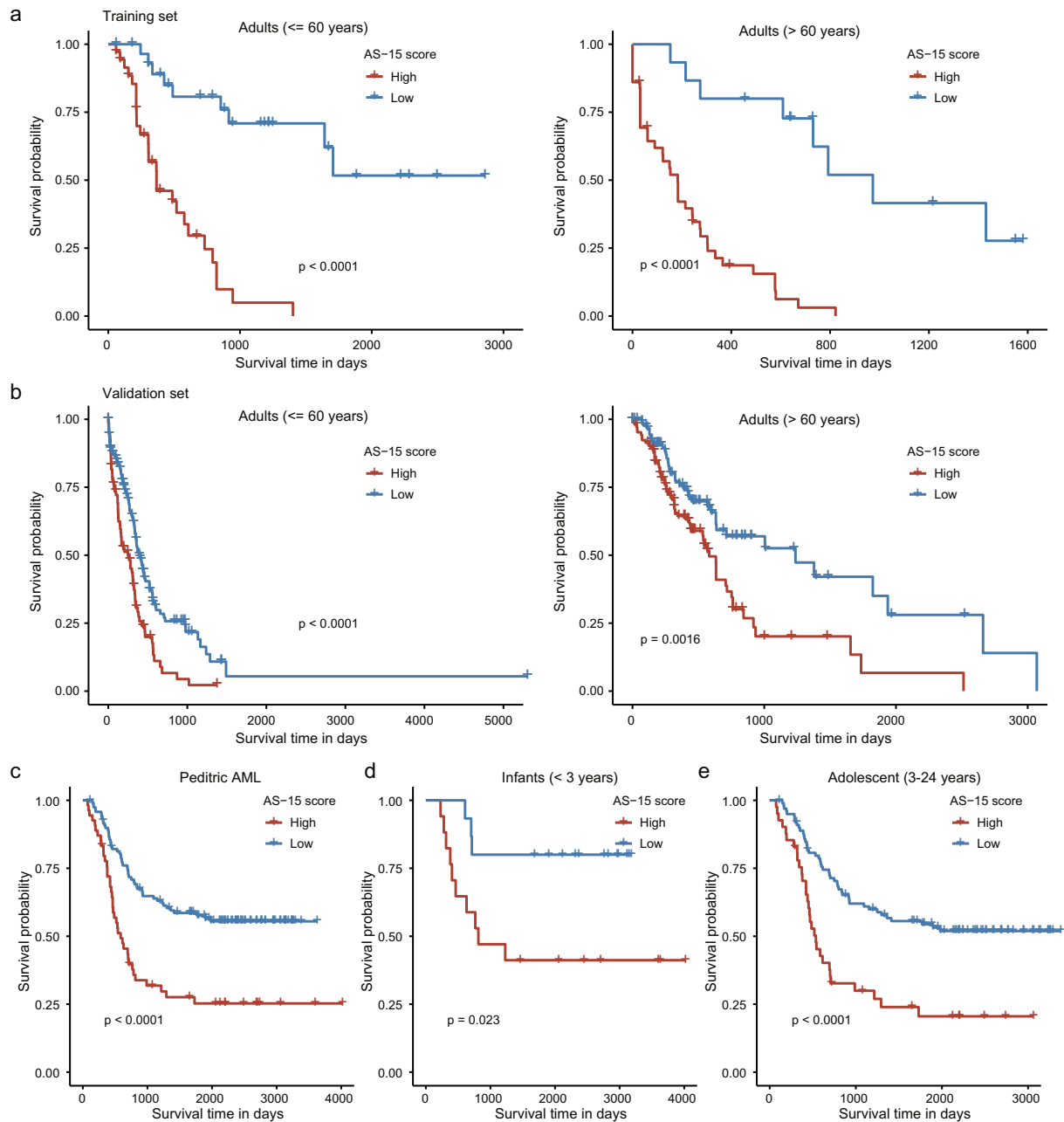


Fig. 3. The independence of the AS-15 score in predicting AML prognosis. Dichotomized survival within AML subtypes using the AS-15 score. Kaplan–Meier curves for overall survival of AML patients classified younger adults (≤ 60 years) and older adults (> 60 years) both in the training set (a) and validation set (b). (c) Kaplan–Meier curve for the overall survival of pediatric AML patients. Kaplan–Meier curves for the overall survival of pediatric AML patients classified as infants (< 3 years) (d) and adolescents (3–24 years) (e).

the ELN favorable, intermediate and adverse groups into AS-15^{high} and AS-15^{low} subgroups resulted in a clear separation of patients having longer and shorter overall survival times, respectively, in the ELN risk groups (Fig. 5a–c, Supplementary Fig. S5c–e).

Based on the above findings, we proposed to improve the ELN classification by including the AS-based signature. We combined the patients from the six groups generated in the previous analysis into three new groups as follows: ELN favorable/AS-15^{high} and ELN adverse/AS-15^{low} patients were re-assigned to the ELN intermediate risk group, and ELN intermediate/AS-15^{high} patients were re-assigned to the ELN adverse risk group (Fig. 5d). Based on the median overall survival, the resulting ELN plus AS-15 score allowed improved risk segregation and successfully

refined the ELN classification (Fig. 5e–f). Similar results were obtained when these analyses were performed independently in the validation set (Fig. 5g–i).

Splicing factors responsible for the prognostic AS events

The AS process is highly organized and regulated by both trans-acting factors and cis-regulatory elements. Splicing factors act as trans-acting factors to influence the exon selection and the splicing site choice by recognizing cis-regulatory elements within pre-mRNAs [26,27], and are extensively dysregulated in AML [13]. We thus hypothesized that

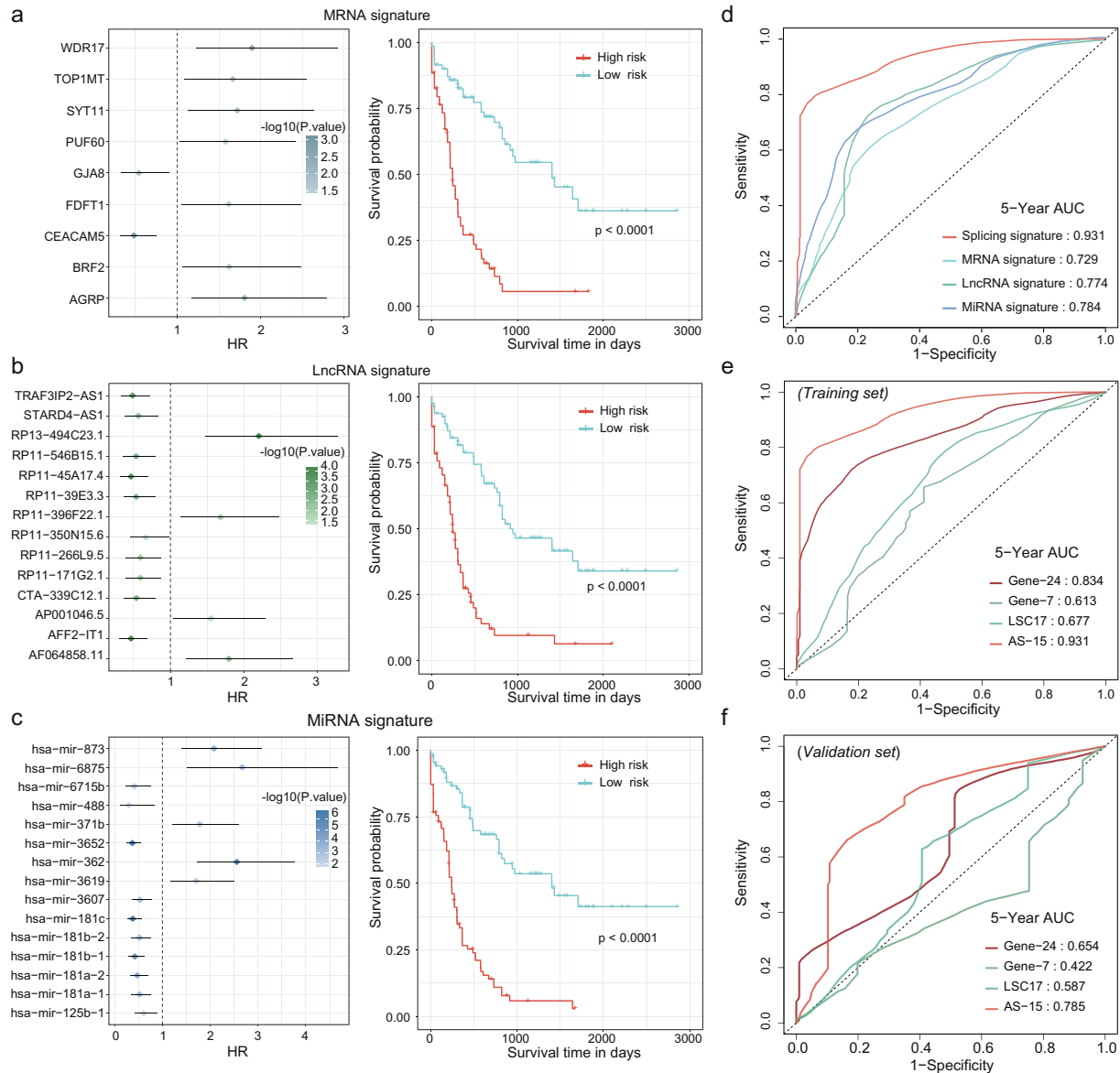


Fig. 4. The AS-15 signature performed better than other expression-based signatures. (a–c) Expression-based signatures in AML. The left panels show the hazard ratios of the robust likelihood-based survival modeling of selected prognostic mRNAs (a), lncRNAs (b) and miRNAs (c). The right panels show the Kaplan–Meier curves for low- and high-risk groups stratified by the median cut algorithm. (d) The ROC curves for the splicing and expression-based signatures, including mRNAs, lncRNAs and miRNAs. The ROC curves for the splicing signature and three other rigorously tested gene expression signatures in the training set (e) and validation set (f).

prognostic AS events in AML are mediated by certain splicing factors, which have potential influences on cancer survival. To do so, we first compared mRNA expression of splicing factor genes in AML patients with normal human bone marrow/peripheral blood samples. We found approximately 61% (359/584) of the genes to be differentially expressed. Among the dysregulated splicing factors in AML, 188 splicing factor genes were significantly upregulated, whereas 171 genes were downregulated (adjusted $p < 0.05$, Fig. 6a, Supplementary Table 5). Next, using the genome-wide CRISPR screening data from 12 AML cell lines and a stringent average 1.5-fold change of sgRNA abundance as the cut-off value, we identified the splicing factor genes that were required for the survival of AML cells. A total of 103 splicing factors were identified, including previously reported RBM39, PCBP1, SRSF2 and RBMX (Fig. 6b, Supplementary Table 6). We then focused on splicing factors that were both differentially expressed in AML and essential for the survival of AML cells

(Supplementary Fig. S6a). Of these splicing factors, we determined those potentially responsible for AS-15 by calculating the correlations between the splicing factors' expression levels and the exon-inclusion levels of these 15 AS events. This correlation analysis defined a splicing network with 132 edges/intercorrelations (p -value < 0.05 , |Spearman's correlation coefficient| > 0.25) between AS-15 and splicing factors (Fig. 6c). On average, one AS event was correlated with ~ 9 splicing factors, suggesting the frequent cross-regulation among splicing factors.

Finally, to validate the roles of splicing factors in AS-15, we retrieved RNA-seq data of Molm13 leukemia cells before and after the knockout of RBM39. The elevated expression of RBM39 in AML contributed to the higher exon inclusion of *SETD5* (Fig. 6d, Supplementary Fig. S6b–d), an essential regulator of histone acetylation during gene transcription [28]. After knocking out RBM39, the exon inclusion level significantly decreased (Fig. 6e). Collectively, the co-expression network analysis

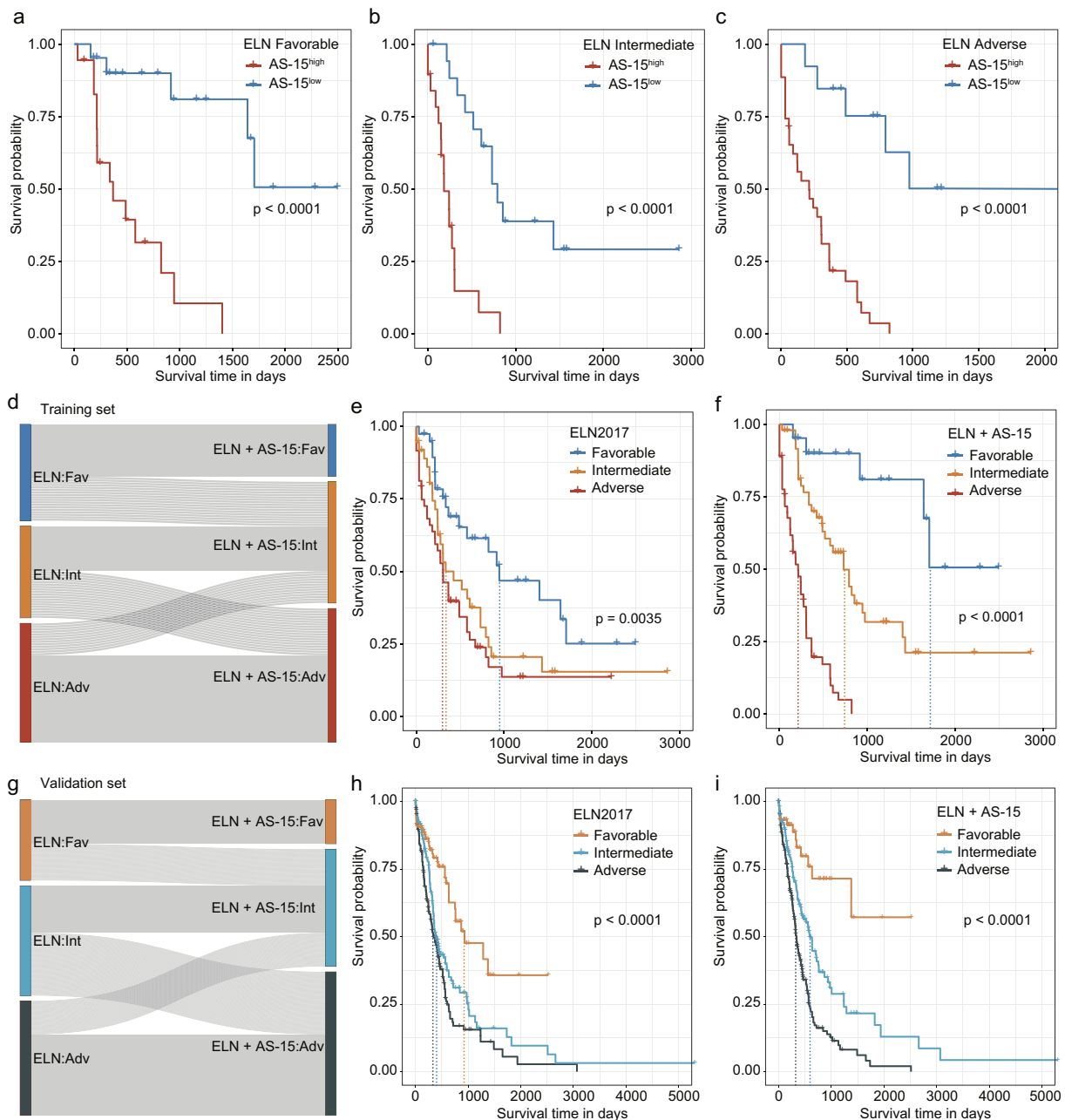


Fig. 5. Correlating the AS-15 score with the ELN classification. (a–c) Kaplan–Meier curves for the overall survival of AML patients classified as ELN favorable (a), intermediate (b) and adverse (c) risk stratified into AS-15^{low} and AS-15^{high} in the training set. (d) Re-stratification of patients in the training set from the three ELN categories (favorable, intermediate and adverse risk) to ELN+AS-15. ELN favorable/AS-15^{high} patients and ELN adverse/AS-15^{low} patients were re-assigned to the intermediate risk group, and ELN intermediate/AS-15^{high} patients were re-assigned to the adverse risk group. (e and f) Kaplan–Meier curves for the overall survival of AML patients in the training set stratified into ELN risk groups (e) and ELN+AS-15 classifications (f). ELN, European LeukemiaNet. AS-15 score, the 15 AS event splicing score. (g) A shift of patients in the validation set from ELN favorable, intermediate and adverse risk groups to ELN+AS-15 favorable, intermediate and adverse risk groups is shown. (h and i) Kaplan–Meier curves for the overall survival of AML patients in the validation set stratified into ELN risk groups (h) and ELN+AS-15 classifications (i).

provided further clues on dysregulated splicing factor-mediated AS mechanisms in AML.

Discussion

Alternative splicing is a highly regulated and coordinated molecular mechanism involved in multiple physiological processes, and its perturba-

tion has gained substantial attention in diverse pathological and disease contexts, including AML [10,29–32]. However, there is limited knowledge regarding their applicability as prognosticators for AML. In this study, through a comprehensive integration analysis of high-throughput RNA-seq datasets, we demonstrated the presence of extensive AS events in AML transcriptome, supporting their role as drivers of regulatory complexity and functional versatility in cells [7]. We also revealed the prognostic value of AS events, providing a rich source of novel prognostic markers

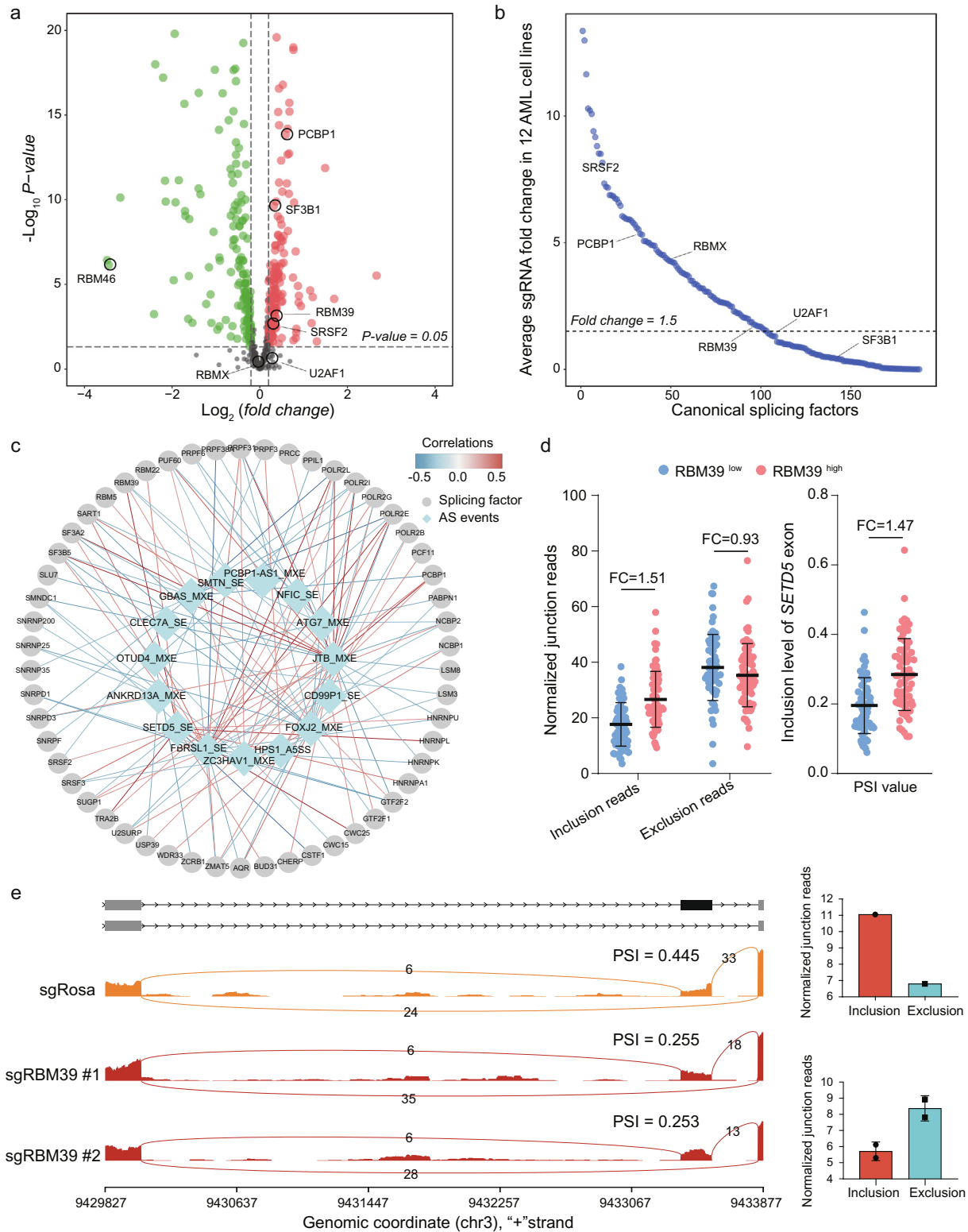


Fig. 6. Splicing-regulatory network correlating splicing factors with prognostic AS events in AML. (a) Volcano plot of differentially expressed splicing factors in AML patient samples ($N = 477$) compared with normal human bone marrow/peripheral blood samples ($N = 33$). (b) Average fold change (initial/final) in sgRNA abundance in a pooled splicing factor-focused negative selection screen of 12 AML cell lines. Each dot represents the average of all the sgRNAs targeting a splicing factor. (c) A splicing co-expression network of 15 prognostic AS events and their associated splicing factor genes. (d) Quantification of the junction reads for the *SETD5* exon are shown for $\text{RBM39}^{\text{low}}$ and $\text{RBM39}^{\text{high}}$ groups based on the RNA-seq data from AML patients. Exon inclusion and exclusion reads were normalized to the total number of reads in each sample. FC, fold change. (e) Sashimi plots showing *SETD5* exon inclusion or exclusion in Molm13 cells before and after the *RBM39* knockout based on RNA-seq junction reads.

for the prognosis of AML patients. Critical genes in AML, such as *RUNX1*, *DNMT3A*, *BAX* and *NOTCH2*, were included in these prognostic events. Previous studies mainly focused on the functional and clinical implications of variations and the dysregulation of these genes [2,11], while AS events that directly orchestrate transcript architecture have been largely overlooked.

While cytogenetic and mutational status have been regarded as the clinical standard for risk stratification and prognosis in AML, the remarkable heterogeneity remains unresolved [1,11,25]. Therefore, assessing additional genetic parameters including gene expression [3–6,33] and DNA methylation [34] has become an efficient strategy for better prognostic risk stratification of patients. As a fundamental aspect of AML pathogenesis [13], AS events captured the splicing programs of leukemic cells and could serve as the prognostic signatures in clinic. Also, a recent study has developed prognostic models for AML patients using a single cohort from the TCGASpliceSeq database [35]. Although there is no overlap of AS events in identified signatures between two studies due to the different feature selection methods, both studies provide evidence for the applicability of AS events in AML prognosis. Furthermore, we found that the AS event-based splicing signature could predict prognosis even better than other well-established signatures based on gene expression [4–6] in two independent datasets. A possible explanation is that AS is not only related to expression of corresponding genes, but also reflects upstream regulations. Collectively, we and others confirm that AS events are an important molecular feature of AML with clinical relevance. Also, the prognostic value of AS events has been uncovered in multiple cancer types, including non-small cell lung cancer [36], ovarian cancer [37], esophageal carcinoma [38], colorectal cancer [39], renal cell carcinoma [40] and pancreatic ductal adenocarcinoma [41]. Notably, a recent study has revealed a pan-cancer view of AS events with consequences for the possible relevance in immunotherapy [42], suggesting that more efforts could be made to uncover the pan-cancer AS signatures, serving as biomarkers or therapeutic targets across cancer types.

In RBSURV, a forward gene selection strategy was employed and the optimal predictive model was selected by using the smallest AIC, an approach to minimize the complexity of predictors while maintaining the maximum fit of the predictor to the data [20]. By subjecting all the five types of AS events to RBSURV, we constructed an optimal composite signature, AS-15, by jointly leveraging the information of AS events across types without being trapped in the type-specific marginal optimal events. By incorporating the splicing information of AS-15 into the ELN classification schema, we improved the accuracy of ELN stratification, which could contribute to treatment decisions in the clinic. These prognostic events may be key to understand the remarkable prognostic heterogeneity of the disease that has hindered prediction based on cytogenetic and mutational analyses only [25]. Furthermore, although pediatric AML represent a genetically distinct disease entity [15], when the AS-15 score was applied to pediatric AML, a highly significant prognostic power was observed in both infants (<3 years) and adolescents (3–24 years). Thus, the splicing signature may have captured the common splicing programs both in adult and pediatric AML patients, representing a common transcriptome feature in AML. However, because of the limited available raw RNA-seq data of AML patients with complete clinical information, a much broader cohort of AML patients is required to confirm the prognostic value of the AS-15 score in future studies.

Because splicing factors are major executors of AS processes [43], we further explored those potentially responsible for the AS-15 splicing process. Our differential analysis revealed that aberrant expression of splicing factor genes occurs ubiquitously in AML, which are consistent with previous findings [13]. Next, CRISPR-based screening system further helped us to focus on splicing factors that were also essential for the survival of AML cells. A splicing regulatory network correlating splicing factors and AS-15 events suggested potential regulatory relationships that could be experi-

mentally validated. Noteworthy, a single splicing factor usually recognizes and regulates splicing of many pre-mRNA targets, and that cross-regulation ubiquitously occur among splicing factors [44]. Our data were consistent with this, where one AS event was correlated with ~9 splicing factors. Finally, based on the prognostic value of the AS of *SETD5*, an essential regulator of histone acetylation during gene transcription, and the significant regulatory correlation with dysregulated RBM39, future work, including RIP-seq/CLIP-seq studies and in-depth functional experiments, is needed to confirm these findings and explore the detailed regulatory mechanisms. Thus, the results of the systematic survival analysis of AS events combined other computational methods led to a hypothetical regulatory mechanism underlying AML and provided further clues regarding dysregulated splicing factor-mediated AS mechanisms in AML.

In conclusion, we performed a comprehensive identification and prognostic interpretation of AS events in AML. We identified a 15 AS event-based splicing signature as a powerful prognostic indicator that has the potential to refine the ELN risk stratification, which is beneficial for patient's treatment decisions. Furthermore, the splicing-regulatory network correlating prognostic AS events and associated dysregulated splicing factors provided clues regarding the splicing factor-mediated mechanisms of AML.

Funding sources

This work was supported in part by the National Key Research and Development Program of China (2019YFA0905900) and National Natural Science Foundation Grants of China (81530003, 81890994, 81770153 and 81911530240).

Author contributions

Peng Jin, Yun Tan, Wei Zhang performed the research; Peng Jin designed the research study; Peng Jin, Yun Tan, Wei Zhang, Junmin Li and Kankan Wang analyzed the data; Peng Jin and Kankan Wang wrote and revised the paper. All authors read and approved the submitted and final versions.

Declaration of competing interests

The authors declare no conflicts of interest.

Acknowledgments

We thank the researchers of TCGA-LAML (phs000178), BeatAML (phs001657) and TARGET-AML (phs000218) projects for making their RNA-seq and clinical data available upon request. We thank Dr. Hai Fang for his critical reading of the manuscript and helpful discussions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neo.2020.06.004>.

References

1. Döhner H, Weisdorf DJ, Bloomfield CD. Acute myeloid leukemia. *N Engl J Med* 2015;**373**(12):1136–52.
2. Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson AG, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013;**368**(22):2059–74.
3. Beck D, Thoms JAL, Palu C, Herold T, Shah A, Olivier J, et al. A four-gene lincRNA expression signature predicts risk in multiple cohorts of acute myeloid leukemia patients. *Leukemia* 2018;**32**(2):263–72.

4. Li Z, Herold T, He C, Valk PJM, Chen P, Jurinovic V, et al. Identification of a 24-gene prognostic signature that improves the European LeukemiaNet risk classification of acute myeloid leukemia: an international collaborative study. *J Clin Oncol* 2013;**31**(9):1172–81.
5. Marcucci G, Yan P, Maharry K, Frankhouser D, Nicolet D, Metzeler KH, et al. Epigenetics meets genetics in acute myeloid leukemia: clinical impact of a novel seven-gene score. *J Clin Oncol* 2014;**32**(6):548–56.
6. Ng SWK, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* 2016;**540**(7633):433–7.
7. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 2008;**40**(12):1413–5.
8. Frankish A, Uszczyńska B, Ritchie GRS, Gonzalez JM, Pervouchine D, Petryszak R, et al. Comparison of GENCODE and RefSeq gene annotation and the impact of reference geneset on variant effect prediction. *BMC Genomics* 2015;**16**(Suppl 8):S2.
9. Gallego-Paez LM, Bordone MC, Leote AC, Saraiva-Agostinho N, Ascensão-Ferreira M, Barbosa-Morais NL. Alternative splicing: the pledge, the turn, and the prestige: the key role of alternative splicing in human biological systems. *Hum Genet* 2017;**136**(9):1015–42.
10. Lee SC-W, Abdel-Wahab O. Therapeutic targeting of splicing in cancer. *Nat Med* 2016;**22**(9):976–86.
11. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med* 2016;**374**(23):2209–21.
12. Zhang J, Manley JL. Misregulation of pre-mRNA alternative splicing in cancer. *Cancer Discov* 2013;**3**(11):1228–37.
13. Wang E, Lu SX, Pastore A, Chen X, Imig J, Chun-Wei Lee S, et al. Targeting an RNA-binding protein network in acute myeloid leukemia. *Cancer Cell* 2019;**35**(3).
14. Tyner JW, Tognon CE, Bottomly D, Wilmot B, Kurtz SE, Savage SL, et al. Functional genomic landscape of acute myeloid leukaemia. *Nature* 2018;**562**(7728):526–31.
15. Bolouri H, Farrar JE, Triche T, Ries RE, Lim EL, Alonzo TA, et al. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat Med* 2018;**24**(1):103–12.
16. Shen S, Park JW, Lu Z-x, Lin L, Henry MD, Wu YN, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci USA* 2014;**111**(51):E5593–601.
17. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019;**10**(1):1523.
18. Wang T, Yu H, Hughes NW, Liu B, Kendirli A, Klein K, et al. Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic Ras. *Cell* 2017;**168**(5).
19. O'Quigley J, Moreau T. Cox's regression model: computing a goodness of fit statistic. *Comput Methods Programs Biomed* 1986;**22**(3):253–6.
20. Cho H, Yu A, Kim S, Kang J, Hong S. Robust likelihood-based survival modeling with microarray data. *J Statistical Software* 2009;**29**(1):1–16.
21. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000;**56**(2):337–44.
22. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* 2019;**47**(D1):D351–60.
23. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**(12):550.
24. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;**13**(11):2498–504.
25. Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Błchner T, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 2017;**129**(4):424–47.
26. Urbanski LM, Leclair N, Anczuk/Ew 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**(4) e1476.
27. Shen S, Wang Y, Wang C, Wu YN, Xing Y. SURVIV for survival analysis of mRNA isoform variation. *Nat Commun* 2016;**7**:11548.
28. Osipovich AB, Gangula R, Vianna PG, Magnuson MA. Setd5 is essential for mammalian development and the co-transcriptional regulation of histone acetylation. *Development* 2016;**143**(24):4595–607.
29. David CJ, Manley JL. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev* 2010;**24**(21):2343–64.
30. Cieply B, Carstens RP. Functional roles of alternative splicing factors in human disease. *Wiley Interdiscip Rev RNA* 2015;**6**(3):311–26.
31. Zhou J, Chng W-J. Aberrant RNA splicing and mutations in spliceosome complex in acute myeloid leukemia. *Stem Cell Investig* 2017;**4**:6.
32. Smith MA, Choudhary GS, Pellagatti A, Choi K, Bolanos LC, Bhagat TD, et al. U2AF1 mutations induce oncogenic IRAK4 isoforms and activate innate immune pathways in myeloid malignancies. *Nat Cell Biol* 2019;**21**(5):640–50.
33. Chuang MK, Chiu YC, Chou WC, Hou HA, Chuang EY, Tien HF. A 3-microRNA scoring system for prognostication in de novo acute myeloid leukemia patients. *Leukemia* 2015;**29**(5):1051–9.
34. Figueroa ME, Lugthart S, Li Y, Erpelinck-Verschueren C, Deng X, Christos PJ, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* 2010;**17**(1):13–27.
35. Xie ZC, Gao L, Chen G, Ma J, Yang LH, He RQ, et al. Prognostic alternative splicing regulatory network of splicing events in acute myeloid leukemia patients based on SpliceSeq data from 136 cases. *Neoplasma* 2020;**67**(3):623–35.
36. Li Y, Sun N, Lu Z, Sun S, Huang J, Chen Z, et al. Prognostic alternative mRNA splicing signature in non-small cell lung cancer. *Cancer Lett* 2017;**393**:40–51.
37. Zhu J, Chen Z, Yong L. Systematic profiling of alternative splicing signature reveals prognostic predictor for ovarian cancer. *Gynecol Oncol* 2018;**148**(2):368–74.
38. Mao S, Li Y, Lu Z, Che Y, Sun S, Huang J, et al. Survival-associated alternative splicing signatures in esophageal carcinoma. *Carcinogenesis* 2019;**40**(1):121–30.
39. Xiong Y, Deng Y, Wang K, Zhou H, Zheng X, Si L, et al. Profiles of alternative splicing in colorectal cancer and their clinical significance: a study based on large-scale sequencing data. *EBioMedicine* 2018;**36**:183–95.
40. Chen T, Zheng W, Chen J, Lin S, Zou Z, Li X, et al. Systematic analysis of survival-associated alternative splicing signatures in clear cell renal cell carcinoma. *J Cell Biochem* 2019.
41. Yang C, Wu Q, Huang K, Wang X, Yu T, Liao X, et al. Genome-wide profiling reveals the landscape of prognostic alternative splicing signatures in pancreatic ductal adenocarcinoma. *Front Oncol* 2019;**9**:511.
42. Kahles A, Lehmann K-V, Toussaint NC, Hfser M, Stark SG, Sachsenberg T, et al. Comprehensive analysis of alternative splicing across tumors from 8,705 patients. *Cancer Cell* 2018;**34**(2).
43. Sebestyán E, Singh B, Mi-ana B, Pagés A, Mateo F, Pujana MA, et al. Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res* 2016;**26**(6):732–44.
44. Sun Y, Bao Y, Han W, Song F, Shen X, Zhao J, et al. Autoregulation of RBM10 and cross-regulation of RBM10/RBM5 via alternative splicing-coupled nonsense-mediated decay. *Nucleic Acids Res* 2017;**45**(14):8524–40.