Clinical presentation and differential splicing of SRSF2, U2AF1 and SF3B1 mutations in patients with Acute Myeloid Leukaemia

Title:

Clinical presentation and differential splicing of SRSF2, U2AF1 and SF3B1 mutations in patients with Acute Myeloid Leukaemia

Running Head:

Characterization of SF mutations in AML

Stefanos A. Bamopoulos1,2, Aarif M. N. Batcha3,4, Vindi Jurinovic1,3, Maja Rothenberg-Thurley1, Hanna Janke1, Bianka Ksienzyk1, Julia Philippou-Massier5, Alexander Graf5, Stefan Krebs5, Helmut Blum5, Stephanie Schneider1,6, Nikola Konstandin1, Maria Cristina Sauerland7, Dennis Görlich7, Wolfgang E. Berdel8, Bernhard J. Woermann9, Stefan K. Bohlander10, Stefan Canzar11, Ulrich Mansmann3,4,12,13, Wolfgang Hiddemann1,12,13, Jan Braess14, Karsten Spiekermann1,12,13, Klaus H. Metzeler1,12,13 and Tobias Herold1,12,15

1 Laboratory for Leukemia Diagnostics, Department of Medicine III, University Hospital, LMU Munich, Munich, Germany
2 Department of Hematology and Oncology (CBF), Charité University Medicine, Berlin, Germany
3 Institute for Medical Information Processing, Biometry and Epidemiology, LMU Munich, Munich, Germany
4 DIFUTURE (Data integration for Future Medicine (DiFuture, www.difuture.de), LMU Munich, Munich, Germany
5 Laboratory for Functional Genome Analysis (LAFUGA), Gene Center, LMU Munich, Munich, Germany
6 Institute of Human Genetics, University Hospital, LMU Munich, Munich, Germany

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
7 Institute of Biostatistics and Clinical Research, University of Münster, Münster, Germany
8 Department of Medicine, Hematology and Oncology, University of Münster, Münster, Germany
9 German Society of Hematology and Oncology, Berlin, Germany
10 Leukaemia and Blood Cancer Research Unit, Department of Molecular Medicine and Pathology, University of Auckland, Auckland, New Zealand
11 Gene Center, LMU Munich, Munich, Germany
12 German Cancer Consortium (DKTK), Partner Site Munich, Munich, Germany
13 German Cancer Research Center (DKFZ), Heidelberg, Germany
14 Department of Oncology and Hematology, Hospital Barmherzige Brüder, Regensburg, Germany
15 Research Unit Apoptosis in Hematopoietic Stem Cells, Helmholtz Zentrum München, German Center for Environmental Health (HMGU), Munich, Germany

**Corresponding Authors:**

Stefanos A. Bamopoulos; stefanos.bamopoulos@charite.de
and
Tobias Herold, MD; Marchioninistr. 15; 81377 Munich; Germany; Phone: +49 89 4400-0; FAX: +49 89 4400-74242; Email: tobias.herold@med.uni-muenchen.de
Abstract

Previous studies demonstrated that splicing factor mutations are recurrent events in hematopoietic malignancies with both clinical and functional implications. However, their aberrant splicing patterns in acute myeloid leukaemia remain largely unexplored. In this study we characterized mutations in SRSF2, U2AF1 and SF3B1, the most commonly mutated splicing factors. In our clinical analysis of 2678 patients, splicing factor mutations showed inferior relapse-free and overall survival, however, these mutations did not represent independent prognostic markers. RNA-sequencing of 246 and independent validation in 177 patients revealed an isoform expression profile highly characteristic for each individual mutation, with several isoforms showing a strong dysregulation. By establishing a custom differential splice junction usage pipeline we accurately detected aberrant splicing in splicing factor mutated samples. Mutated samples were characterized predominantly by decreased junction usage. A large proportion of differentially used junctions were novel. Targets of splicing dysregulation included several genes with a known role in leukaemia. In SRSF2(P95H) mutants we further explored the possibility of a cascading effect through the dysregulation of the splicing pathway. We conclude that splicing factor mutations do not represent independent prognostic markers. However, they do have genome-wide consequences on gene splicing leading to dysregulated isoform expression of several genes.
Introduction

The discovery of recurring somatic mutations within splicing factor genes in a large spectrum of human malignancies has brought attention to the critical role of splicing and its complex participation in carcinogenesis [1–3]. The spliceosome is a molecular machine assembled from small nuclear RNA (snRNA) and proteins and is responsible for intron removal (splicing) in pre-messenger RNA. In acute myeloid leukaemia (AML), splicing factor mutations occur most frequently in SRSF2, U2AF1 and SF3B1. The splicing factors encoded by these genes are all involved in the recognition of the 3'-splice site during pre-mRNA processing.[4] Splicing factor (SF) mutations are especially common in haematopoietic malignancies, where they occur early on and remain stable throughout the disease evolution of myelodysplastic syndromes (MDS) [1,5–9]. SF mutations are also prevalent in acute myeloid leukaemia (AML), which is often the result of myeloid dysplasia progression, with reported frequencies of 6-10%, 4-8% and 3% for SRSF2, U2AF1 and SF3B1 mutations respectively [2,4,10,11].

SF mutations rarely co-occur within the same patient, implying the lack of a synergistic effect or synthetic lethality [1,2,6]. They are typically heterozygous point mutations, frequently coincide with other recurrent mutations in haematopoietic malignancies and are associated with aberrant splicing in genes recurrently mutated in AML [2,4,8]. Notably, the aberrant splicing patterns are distinct for each SF mutation, suggesting that SF mutations do not share the same mechanism of action and should be recognized as individual alterations [4,9,12–17].

The clinical characteristics and outcome of patients with SF mutations are well defined in MDS [1,3,8,9]. Meanwhile, attempts at determining the role of SF mutations as independent prognostic markers in AML have often been limited to specific subgroups and it remains unclear whether the inferior survival associated with SF mutations is confounded by their association with older age or accompanying mutations [10,18]. Additionally, while evidence of
aberrant splicing due to SF mutations has emerged for many genes relevant in AML, it is yet uncertain whether and how these changes directly influence disease initiation or evolution. The aim of this study was a comprehensive analysis of the prognostic implications of SF mutations in two well-characterized and intensively treated adult AML patient cohorts amounting to a total of 2678 patients. In addition, the core functional consequences of SF mutations were explored using targeted amplicon sequencing in conjunction with RNA-sequencing on two large datasets.
Patients and Methods

Patients

Our primary cohort included a total of 1138 AML patients treated with intensive chemotherapy in two randomized multicenter phase 3 trials of the German AML Cooperative Group (AMLCG). Treatment regimens and inclusion criteria are described elsewhere [2]. A cohort of 1540 AML patients participating in multicenter clinical trials of the German-Austrian AML Study Group (AMLSG), was used for validation [19]. Cohort composition and filtering criteria are outlined in the supplementary.

Molecular Workup

All participants of the AMLCG cohort received cytogenetic analysis, as well as targeted DNA-sequencing as described previously [2]. The subset of the AMLSG cohort included in this study received a corresponding molecular workup, described elsewhere [19].

RNA-Sequencing and data processing

Using the Sense mRNA Seq Library Prep Kit V2 (Lexogen; Vienna, Austria) 246 samples underwent poly(A)-selected, strand-specific, paired-end sequencing on a HiSeq 1500 instrument (Illumina; San Diego, CA, USA). A subset of the Beat AML cohort (n=177) was used for validation [20]. The same bioinformatics analysis was used for both datasets and is described in the supplementary. The samples were aligned to the reference genome (Ensembl GRCh37 release 87) using the STAR [21] aligner with default parameters. Splice junctions from all samples were pooled, filtered and used to create a new genomic index. Multi-sample 2-pass alignments to the re-generated genome index followed, using the STAR recommended parameters for gene-fusion detection. Read counts of transcripts and genes were measured with salmon [22]. Read counts of splice junctions were extracted from the STAR output.
Differential expression analysis and differential splice junction usage (DSJU)

A minimum expression filter was applied prior to each differential analysis. Differentially expressed isoforms were identified with the limma [23] package after TMM-normalization [24] with edgeR [25] and weighting with voom [26,27]. A surrogate variable analysis step using the sva [28] package was included to reduce unwanted technical noise. DSJU was quantified similarly using the diffSplice function of the limma package. Both analyses are described in detail in the supplementary.

Nanopore cDNA sequencing and analysis

Total RNA was transcribed into cDNA using the TeloPrime Full-Length cDNA Amplification Kit (Lexogen) which is highly selective for polyadenylated full-length RNA molecules with 5'-cap structures. Two barcoded samples for multiplexed analysis were sequenced on the Oxford Nanopore Technologies MinION platform. Alternative isoform analysis was performed with FLAIR [29].

Statistics

Statistical analysis was performed using the R-3.4.1 [30] software package. Correlations between variables were performed using the Mann-Whitney U test and the Pearson’s chi-squared test. In case of multiple testing, p-value adjustment was performed as described in the supplementary. Survival analysis was performed and visualized using the Kaplan-Meier method and the log-rank test was utilized to capture differences in relapse free survival (RFS) and overall survival (OS). Patients receiving an allogeneic stem cell transplant were censored at the day of the transplant, for both RFS and OS. Additionally, Cox regressions were performed for all available clinical parameters and recurrent aberrations. Cox multiple regression models were then built separately for RFS and OS, using all variables with an
unadjusted p-value < 0.1 in the single Cox regression models.
Results

Clinical features of AML patients with SF mutations

We characterized SF mutations in two independent patient cohorts (the AMLCG and AMLSG cohorts). Our primary cohort (AMLCG) consisted of 1119 AML patients (Figure S1), 236 (21.1%) of which presented with SF mutations. The three most commonly affected SF genes, SRSF2, U2AF1 and SF3B1 were mutated in 12.1% (n=136), 3.4% (n=38) and 4.1% (n=46) of the patients (Figure 1A). In agreement with previous findings [19], SF mutations were in their majority mutually exclusive, heterozygous hotspot mutations (Figure 1B). The four most common point mutations were SRSF2(P95H) (n=69), SRSF2(P95L) (n=27), U2AF1(S34F) (n=18), and SF3B1(K700E) (n=18) mutations (Figure 1C-E). The clinical characteristics of patients harboring SF mutations are summarized in Tables 1 and S1 (AMLSG cohort), along with a statistical assessment between cohorts (Table S2). We observed a high overall degree of similarity regarding clinical features of SF mutated patients between the AMLCG and AMLSG cohorts, despite their large median age difference. Mutations in SRSF2, U2AF1 and SF3B1 occurred more frequently in secondary AML (44.7% compared to 18.2% in de novo AML) and were all associated with older age. As reported previously [1], SRSF2 and U2AF1 mutated patients were predominantly male (76.7% and 76.3%, respectively). Furthermore, patients harboring SRSF2 mutations presented with a lower white blood cell count (WBC; median 13.3 \(10^9/L\) vs. 22.4 \(10^9/L\)) while U2AF1 mutated patients presented with a reduced blast percentage in their bone marrow when compared to SF wildtype patients (median 60% vs 80%).

Associations of SF mutations and other recurrent alterations in AML

In a second step, we investigated associations between SF mutations and recurrent cytogenetic abnormalities and gene mutations in AML (Figure 2). Notably, SF mutations were
not found in inv(16)/t(16;16) patients (n=124), with the exception of one inv(16)/t(16;16) patient harboring a \textit{U2AF1}(R35Q) mutation. The same held true for t(8;21) patients (n=98), where only one patient had a rare deletion in \textit{SRSF2}. Additionally, all patients in the AMLCG cohort presenting with an isolated trisomy 13 (n=9) also harbored an \textit{SRSF2} mutation (p<0.001), as described previously [31].

Mutations in all SF genes correlated positively with mutations in \textit{BCOR} and \textit{RUNX1} and negatively with mutations in \textit{NPM1}. Expectedly, \textit{SRSF2}(P95H) and \textit{SRSF2}(P95L) mutations shared a similar pattern of co-expression including significant pairwise associations with mutations in \textit{ASXL1}, \textit{IDH2}, \textit{RUNX1} (both p<0.001) and \textit{STAG2} (p<0.001 and p=0.002, respectively). However, apart from \textit{IDH2} mutations where co-occurrence was comparable (OR: 3.4 vs 5.1), mutations in \textit{ASXL1}, \textit{RUNX1} and \textit{STAG2} coincided more frequently with \textit{SRSF2}(P95H) mutations. Despite this, \textit{SRSF2}(P95L) mutations showed a slightly increased co-occurrence with other recurrent AML mutations (median 5 vs 4 mutations, p=0.046).

**Prognostic relevance of SF mutations for relapse-free survival and overall survival**

The prognostic impact of \textit{SRSF2}, \textit{U2AF1} and \textit{SF3B1} mutations was initially assessed using Kaplan-Meier graphs and log-rank testing. All SF mutations presented with both inferior relapse-free survival (RFS) and overall survival (OS) compared to SF wildtype patients (Figures S3.1A-C and S3.2). The effect was most pronounced in \textit{U2AF1} mutated patients with an one-year survival rate of only 29.1\%, followed by \textit{SF3B1} (40.6\%) and \textit{SRSF2} mutated patients (49.2\%). Different point mutations inside the same SF gene did not differ significantly in their effect on OS.

To confirm the observed prognostic impact of SF mutations we performed single Cox regressions on all available clinical and genetic parameters. In agreement with the Kaplan-Meier estimates, patients harboring \textit{SRSF2}(P95H), \textit{SRSF2}(P95L), \textit{U2AF1}(S34F) and
SF3B1(K700E) mutations had significantly reduced RFS and OS (Figures S3.1D and S4.1). To test whether any SF mutation was an independent prognostic marker, multiple Cox regression models (Figure 3) were built by integrating all parameters significantly associated (p < 0.1) with RFS and OS in the single Cox regression models. Along with several known predictors, only U2AF1(S34F) mutations presented with prognostic relevance for both RFS (Hazard Ratio=2.81, p=0.012) and OS (HR=1.90, p=0.034) in the AMLCG cohort, but not in the AMLSG cohort. However, when aggregating mutations at the gene level, mutations in SRSF2 and SF3B1 presented with prognostic relevance for RFS in the AMLSG cohort (HR=1.77, p=0.008; HR=2.15, p=0.014; respectively), while not reaching significance in the AMLCG cohort (Table S4.2). When looking only at de novo AML patients, the prognostic impact of U2AF1(S34F) mutations diminished, yet the prognostic impact observed for SRSF2 and SF3B1 remained significant in the AMLSG cohort (HR=1.84, p=0.009; HR=2.43, p=0.015; respectively) (Tables S4.1-S4.2).

**Differential isoform expression in SF mutated patients**

We next assessed the impact of SF mutations on mRNA expression. To this end, whole-transcriptome RNA-sequencing was performed on 246 AML patients, 29 of which harbored a mutation in the SF genes of interest, while 199 SF wildtype patients were used as a control (Figure S2). The remaining patients either presented with a different SF mutation (n=17) or exhibited more than one SF mutation (n=1) and were excluded. In addition, a subset of the Beat AML cohort (n=177) with matched DNA- and RNA-sequencing data was used for validation [20].

After low-coverage filtering we performed a differential isoform expression analysis for ~90 000 isoforms. Differential expression was restricted to a small fraction of all expressed isoforms (<0.5%; Figure 4A and Table S6.1). Little overlap of differentially expressed (DE)
isoforms was found when different SF mutation groups were compared to the control, consistent with previous observations [32]. However, 10 isoforms were reported as DE in both SRSF2(P95H) and SRSF2(P95L) mutated samples, all with the same fold-change direction (Figure 4B). Out of those, the isoforms in GTF2I, H1F0, INHBC, LAMC1 and one of the isoforms of METTL22 (ENST00000562151) were also significant in the validation cohort for both SRSF2(P95H) and SRSF2(P95L). Additionally, the isoform of H1F0 was also reported as DE for U2AF1(S34F) mutants in both cohorts. For SRSF2(P95H) mutants 107 of all DE isoforms also reached significance in the validation cohort (40.1%), while for the other SF mutation subgroups validation rates ranged from 15.1 to 27.3% increasing with larger mutant sample sizes. Notably, mutated and wildtype samples showed large differences in the expression levels of several isoforms (Figure 4C and Figure S5). The top two overexpressed isoforms in SRSF2(P95H) both corresponded to INTS3, which was recently reported as dysregulated in SRSF2(P95H) mutants co-expressing IDH2 mutations [33]. Several DE isoforms identified in SF mutated patients correspond to cancer-related genes, many of which have a known role in AML. Specifically, genes with DE isoforms included, but were not limited to BRD4 [34], EWSR1 [35] and YBX1 [36] in SRSF2(P95H) mutated samples, CUX1 [37], DEK [15,38] and EZH1 [39] in U2AF1(S34F) mutated samples, as well as PTK2 [40] in SF3B1(K700E) mutated patients (Tables S7.1-S7.2).

Hierarchical clustering using DE isoforms was performed for all samples to assess the expression homogeneity of SF mutations. A tight clustering of samples harboring identical SF point mutations was observed, indicating an isoform expression profile highly characteristic for each individual SF mutation (Figures S6.1-S6.3). When using DE isoforms resulting from the comparison of all SRSF2 mutated samples against SF wildtype samples, the samples did not cluster as well. This stands in agreement with the limited overlap of differentially
expressed isoforms found between the two SRSF2 point mutations examined and suggests at least some heterogeneity among them. The same also held true for U2AF1 mutated samples, however all SF3B1 mutated samples still clustered together when compared as a single group to the control.

Differential splicing in SF mutants

Previous studies have reported differential splicing as causal for isoform dysregulation in SF mutants [41,42]. To detect aberrant splicing in our dataset, we quantified the usage of all unique splice junctions by pooling information from all samples (Figure S7). After filtering out junctions with low expression, 235,730 junctions were assigned to genes. Only junctions within an annotated gene were considered, leading to the exclusion of 11,617 (4.9%) junctions. Junctions present in genes with low-expression or genes with a single junction were excluded, leaving 221,249 unique junctions (19.3% novel) across 15,526 genes (Table S6.1). Applying the same workflow to the Beat AML cohort yielded 194,158 junctions (8.3% novel). Notably, of the 172,518 junctions shared across both datasets, 10,029 (5.8%) were novel. The novel junctions passing our filtering criteria were supported by a high amount of reads and samples with a distribution comparable to that of annotated junctions (Figure 5A). Neither the number of novel junctions nor the number of reads supporting them correlated with the presence of SF mutations, suggesting that novel splicing events are not increased in SF mutants.

In consideration of the high proportion of novel junctions in both datasets, we employed a customized pipeline that can quantify the differential splice junction usage (DSJU) of each individual junction, by harnessing usage information from all junctions inside one gene. Of the several hundred junctions reported as differentially used in our primary cohort (p<0.05, log₂(fold change) > 1), 20.2-45.9% constituted novel junctions (Tables S7.1-S7.3 and Tables
S9.1-S9.2) and were classified according to their relationship with known acceptor and donor sites as described previously (Figure 5B) [15]. Unsurprisingly, validation rates increased with larger mutant sample sizes, ranging from 9.3% (SF3B1(K700E); n = 3) to 74.0% (all SRSF2 mutants; n = 26). Furthermore, validation rates were higher for novel junctions (mean 39.3% vs. 21.5% known junctions), likely due to the stricter initial filtering criteria applied. By performing nanopore sequencing of one SRSF2(P95H) mutant and one SF wildtype sample we were able to confirm the usage of several novel junctions and detect resulting novel isoforms as exemplified for IDH3G in Figure 6A-D. A tendency towards decreased junction usage was observed for all SF point mutations and was most evident in SF3B1(K700E) mutants (1 423/1 927; 73.9% of differentially used junctions). The total number of splicing events, however, was not reduced in SF mutants (mean 9,275,359 events vs. 9 192 697 in wildtype patients), suggesting that decreased splicing is limited to selected junctions rather than being a global effect.

We systematically compared the genes with at least one DE isoform and those reported as differentially spliced in all SF mutation subgroups (Tables S9.3-S9.4). For SRSF2 mutants, genes significant in both analyses included EWSR1, H1F0, INTS3 and YBX1. In general, out of the genes examined in both analyses only 9.8-23.3% (depending on the SF mutation) of genes reported as having a DE isoform were also reported as being differentially spliced. Conversely, 3.3-28.5% of differentially spliced genes were also reported as having a DE isoform. These findings suggest that differential gene splicing does not always lead to altered isoform expression while at the same time differential isoform expression cannot always be attributed to an explicit splicing alteration. Considering the complementary nature of the analyses, we performed gene ontology (GO) analysis by combining the genes with evidence of differential isoform usage or differential splicing (Tables S10.1-S10.7). Interestingly, GO
terms enriched for both SRSF2 mutants included “mRNA splicing, via spliceosome” (p<0.001 and p=0.046, respectively) and “mRNA splice site selection” (p=0.022 and p=0.019, respectively) (Figure 5C).

In an additional step, the splice junction counts reported by Okeyo-Owuor et al. were used to detect DSJU between CD34+ cells with U2AF1(S34) mutations (n=3) and SF wildtype (n=3) via the same pipeline applied to the AMLCG and Beat AML cohorts. While no identical junctions were differentially used in all three datasets, 16 genes were reported as differentially spliced in all, including leukemia or cancer-associated genes (ABI1, DEK, HP1BP3, MCM3 and SET), as well as HNRNPK (a major pre-mRNA binding protein), thereby further refining our list of genes with strong evidence of differential splicing between U2AF1(S34F) mutants and SF wildtype samples (Table S11).

**The effect of SF mutations on the splicing pathway**

It is well established that SF mutations dysregulate splicing by changing RNA binding affinities or altering 3’ splice site recognition of the corresponding splicing factors. However, SRSF2 also plays a splicing-independent role in transcriptional pausing by translocating the positive transcription elongation factor complex (P-TEFb) from the 7SK complex to RNA polymerase II [43]. A recent study reported that mutant SRSF2(P95H) enhances R-loop formation due to impaired transcriptional pause release, thus providing evidence that altered splicing does not account for the entirety of the SRSF2(P95H) mutant phenotype [32]. In another study no difference was found in the total mRNA of 12 Serine/arginine rich splicing factors and 14 out of 16 major heterogeneous nuclear ribonucleoprotein (hnRNP) splicing factors between SRSF2(P95H) mutant and WT CRISPR clones [44].

We cross-referenced our differential expression and differential splicing analysis results with a list of all genes involved in splicing (GO:0000398; mRNA splicing, via spliceosome). Of the
347 splicing-related genes (317 of which were expressed in our dataset) 101 were
dysregulated in at least one SF mutant group. On average 30.5 (range 6-52) splicing-related
genes were dysregulated per SF point mutation. Of note, both SRSF2 point mutations
associated with differential splicing of HNRNPA1 and HNRNPUL1, as well as PCF11 and
TRA2A. A query of the STRING database suggests protein-protein interactions between
SRSF2 and the proteins of the above genes, which are also interconnected (Figure 6E).
Interestingly, one of the differential splicing events reported in both SRSF2 mutants involves
the under-usage of the same novel splice junction in TRA2A (Figure 6F). TRA2A has
previously been shown to be differentially spliced in mouse embryo fibroblasts upon SRSF2
knockout [45]. Furthermore, it has been shown that both HNRNPA1 and SRSF2 interact with
the loop 3 region of 7SK RNA and by favoring the dissociation of SRSF2, HNRNPA1 may
lead to the release of active P-TEFb [46]. Taken together, our results indicate a strong
dysregulation of the splicing pathway in SF mutants including several genes whose gene
products closely interact with SRSF2.
Discussion

The clinical relevance of SF mutations and their aberrant splicing patterns have been explored in myelodysplasia, while comparable data for AML is lacking. In this study we examined two AML patient cohorts, encompassing a total of 2678 patients from randomized prospective trials, to characterize SF mutations clinically. This analysis was complemented by RNA-seq analysis of two large datasets to reveal targets of aberrant splicing in AML.

We show that SF mutations are frequent alterations in AML, identified in 21.4% of our primary patient cohort, especially in elderly patients and in secondary AML. SF mutations are associated with other recurrent mutations in AML, such as BCOR and RUNX1 mutations, however SRSF2(P95L) mutations co-occur less often with those mutations when compared to SRSF2(P95H) mutations, albeit showing a slightly increased mutational load. This suggests a more diverse co-expression profile of SRSF2(P95L).

Previous studies have demonstrated the predictive value of SF mutations in clonal haematopoiesis of indeterminate potential (CHIP) [47], MDS [6,8,48–50] and AML [10,18,19,51]. However, survival analyses in AML were, in their majority, hampered by small sample sizes and limited availability of further risk factors. Therefore, we examined whether SF mutations impact survival while accounting for recently proposed risk parameters included in the ELN 2017 classification [52]. In our analysis, SRSF2 and SF3B1 mutations were no independent prognostic markers for OS in AML. U2AF1(S34F) mutations displayed poor OS in the AMLCG cohort, which we were unable to validate in the AMLSG cohort. The discrepancy in survival of SF mutated patients between the two cohorts lied most likely in the large age difference of the participants (median age difference of 8 years), which also led to a higher percentage of patients receiving allogeneic transplants in the AMLSG cohort (56.5% vs. 30.6% in the AMLCG cohort). In summary, SF mutations are early evolutionary events and
define prognosis and transformation risk in CHIP and MDS patients, yet there is no clear independent prognostic value of SF mutations in AML. Two large RNA-sequencing studies have been performed previously, to detect aberrantly spliced genes in SF mutants, both of which focused on MDS patients [41,42]. In this study we described a distinct differential isoform expression profile for each SF point mutation. Furthermore, we evaluated differential splicing for the four most common SF point mutations via a customized pipeline to determine differential usage of both known and novel splice junctions. Our pipeline enables the differential quantification of individual splice junctions without restricting the analysis to annotated alternative splicing events. We argue that the strength of our analysis lies in the accurate detection of single dysregulated junctions (especially in cases where splice sites are shared by multiple junctions) in an annotation-independent manner achieving validation rates up to 74.0% in our largest mutant sample group (SRSF2, n=19). Limitations of the analysis include the restriction to junctions with both splice sites within the same gene (a restriction shared by most differential splicing algorithms) and genes with at least two junctions. However, the reduced requirements of our analysis could prove valuable in the study of differential splicing in organisms with lacking annotation. All SF point mutations shared a tendency towards decreased splice junction usage, which did not affect the global number of splicing events in SF mutants. Surprisingly, we observed a limited overlap between genes with differentially expressed isoforms and differentially spliced genes. In addition, a recent study by Liang et al. reported that the majority of differential binding events in SRSF2(P95H) mutants do not translate to alternative splicing [53]. Taken together, these findings indicate a “selection” or possibly a compensation of deregulatory events from differential binding through differential splicing to finally differential isoform expression. Furthermore, the enrichment of aberrant splicing in splicing-related genes opens
the possibility of a cascading effect on transcription via the differential alternative splicing of transcriptional components. A congruent hypothesis was stated by Liang et al., where an enrichment of SRSF2(P95H) targets in RNA processing and splicing was shown, further supporting the notion of an indirect effect of mutant SRSF2 facilitated through additional splicing components. Further investigations may provide a mechanistic link between the differential splicing of selected genes and the impairment of transcription and specifically transcriptional pausing observed in SF mutant cells, which contributes to the MDS phenotype [32].

To the best of our knowledge our study represents the most comprehensive analysis of SF mutations in AML to date, both in terms of clinical and functional characterization. This enabled us to study SRSF2(P95H) and SRSF2(P95L) separately, thereby not only outlining their differences but also identifying common and likely core targets of differential splicing in SRSF2 mutants. We conclude that SF mutated patients represent a distinct subgroup of AML patients with poor prognosis that is not attributable solely to the presence of SF mutations. SF mutations induce aberrant splicing throughout the genome including the dysregulation of several genes associated with AML pathogenesis, as well as a number of genes with immediate, functional implications on splicing and transcription. Further studies are required to identify which splicing events are critical in leukaemogenesis and whether they are accessible to new treatments options, such as splicing inhibitors [54] and immunotherapeutic approaches.
Acknowledgments

The authors thank all participants and recruiting centers of the AMLCG, BEAT and AMLSG trials.

Funding

This work is supported by a grant of the Wilhelm-Sander-Stiftung (no. 2013.086.2) and the Physician Scientists Grant (G-509200-004) from the Helmholtz Zentrum München to T.H. and the German Cancer Consortium (Deutsches Konsortium für Translationale Krebsforschung, Heidelberg, Germany). K.H.M., K.S. and T.H. are supported by a grant from Deutsche Forschungsgemeinschaft (DFG SFB 1243, TP A06 and TP A07). S.K.B. is supported by Leukaemia & Blood Cancer New Zealand and the family of Marijanna Kumerich. A.M.N.B. is supported by the BMBF grant 01ZZ1804B (DIFUTURE).

Author Contributions


Additional Information

The authors declare no conflicts of interest.
References


Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms.


sample and observational level variability improves power in RNA-seq analyses.


30. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing;


Table 1: Clinical characteristics of SF mutations in the AMLCG cohort.

<table>
<thead>
<tr>
<th>Variables</th>
<th>SF wildtype</th>
<th>SRSF2</th>
<th>P</th>
<th>U2AF1</th>
<th>P</th>
<th>SF3B1</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of patients</td>
<td>903</td>
<td>133</td>
<td>-</td>
<td>38</td>
<td>-</td>
<td>45</td>
<td>-</td>
</tr>
<tr>
<td>Age, years, median (range)</td>
<td>55 (18-86)</td>
<td>65 (25-80)</td>
<td>&lt;0.001</td>
<td>64 (23-74)</td>
<td>0.007</td>
<td>65 (31-78)</td>
<td>0.001</td>
</tr>
<tr>
<td>Female sex, no. (%)</td>
<td>490 (54.3)</td>
<td>31 (23.3)</td>
<td>&lt;0.001</td>
<td>9 (23.7)</td>
<td>0.003</td>
<td>19 (42.2)</td>
<td>0.387</td>
</tr>
<tr>
<td>Hemoglobin, g/dL, median (range)</td>
<td>9 (3.5-16)</td>
<td>8.9 (3.8-14.7)</td>
<td>0.466</td>
<td>9.2 (6-13.6)</td>
<td>1.000</td>
<td>8.9 (6.8-13.4)</td>
<td>1.000</td>
</tr>
<tr>
<td>WBC count, 10^9/L, median (range)</td>
<td>22.4 (0.1-798.2)</td>
<td>13.3 (0.5-406)</td>
<td>0.008</td>
<td>7 (0.7-666)</td>
<td>0.079</td>
<td>22.4 (0.9-269.5)</td>
<td>1.000</td>
</tr>
<tr>
<td>Platelets, 10^9/L, median (range)</td>
<td>55 (0-1760)</td>
<td>49.5 (0-643)</td>
<td>0.736</td>
<td>47 (11-132)</td>
<td>0.466</td>
<td>67 (5-585)</td>
<td>0.744</td>
</tr>
<tr>
<td>LDH, U/L, median (range)</td>
<td>448 (76-19624)</td>
<td>362 (150-14332)</td>
<td>0.118</td>
<td>346 (128-3085)</td>
<td>0.313</td>
<td>472 (142-7434)</td>
<td>0.950</td>
</tr>
<tr>
<td>BM Blasts, %, median (range)</td>
<td>80 (6-100)</td>
<td>76 (15-100)</td>
<td>0.455</td>
<td>60 (10-95)</td>
<td>0.002</td>
<td>70 (13-95)</td>
<td>0.206</td>
</tr>
<tr>
<td>Performance Status (ECOG) &gt; 1, no. (%)</td>
<td>157 (25.9)</td>
<td>18 (26.1)</td>
<td>1.000</td>
<td>8 (27.6)</td>
<td>1.000</td>
<td>4 (16)</td>
<td>0.941</td>
</tr>
<tr>
<td>primary AML, no. (%)</td>
<td>786 (87)</td>
<td>100 (75.2)</td>
<td>0.016</td>
<td>24 (63.2)</td>
<td>0.003</td>
<td>29 (64.4)</td>
<td>0.002</td>
</tr>
<tr>
<td>secondary AML, no. (%)</td>
<td>69 (7.6)</td>
<td>30 (22.6)</td>
<td>&lt;0.001</td>
<td>13 (34.2)</td>
<td>&lt;0.001</td>
<td>11 (24.4)</td>
<td>0.021</td>
</tr>
<tr>
<td>therapy-related AML, no (%)</td>
<td>48 (5.3)</td>
<td>3 (2.3)</td>
<td>0.507</td>
<td>1 (2.6)</td>
<td>1.000</td>
<td>5 (11.1)</td>
<td>0.380</td>
</tr>
<tr>
<td>Allogeneic transplant, no. (%)</td>
<td>296 (32.8)</td>
<td>27 (20.3)</td>
<td>0.022</td>
<td>7 (18.4)</td>
<td>0.273</td>
<td>12 (26.7)</td>
<td>0.797</td>
</tr>
<tr>
<td>Complete Remission, no. (%)</td>
<td>641 (71)</td>
<td>71 (53.4)</td>
<td>&lt;0.001</td>
<td>18 (47.4)</td>
<td>0.015</td>
<td>19 (42.2)</td>
<td>0.001</td>
</tr>
<tr>
<td>Relapse, no (%)</td>
<td>366 (63.5)</td>
<td>46 (82.1)</td>
<td>0.034</td>
<td>11 (78.6)</td>
<td>0.713</td>
<td>14 (87.5)</td>
<td>0.266</td>
</tr>
<tr>
<td>Deceased, no (%)</td>
<td>575 (63.7)</td>
<td>112 (84.2)</td>
<td>&lt;0.001</td>
<td>34 (89.5)</td>
<td>0.010</td>
<td>41 (91.1)</td>
<td>0.002</td>
</tr>
</tbody>
</table>

WBC: white blood cells, LDH: lactate dehydrogenase, BM: bone marrow, ECOG: Eastern Cooperative Oncology Group performance score
**Figure Legends**

**Figure 1: Frequency and location of SF mutations.** (A) Distribution of SF mutations in the AMLCG cohort. (B) Variant allele frequency of SF mutations in both study cohorts. (C) Mutation plots showing the protein location of all SF mutations in the genes *SRSF2, U2AF1,* and *SF3B1* for all patients of the AMLCG cohort. Number of patients harboring SF mutations are additionally provided for both cohorts.

**Figure 2: Correlations between SF mutations and recurrent abnormalities.** Correlation matrix depicting the co-occurrence of SF mutations and recurrent mutations in AML (A), as well as cytogenetic groups, as defined in the ELN 2017 classification (B). Only variables with a frequency >1% in the AMLCG cohort are shown. FLT3-ITD: *FLT3* internal tandem duplication mutation; FLT3-TKD: *FLT3* tyrosine kinase domain mutation; CN-AML: cytogenetically normal AML

**Figure 3: Multiple Cox regression models for overall survival.** Multiple Cox regression models (for OS) were performed separately for patients in the AMLCG cohort (on the left) and AMLSG cohort (on the right). The models include all variables with *p* < 0.1 in the single Cox regression models of the primary cohort (AMLCG cohort). Significant *p*-values (<0.05) are marked with a star. CN-AML: cytogenetically normal AML; LDH: lactate dehydrogenase; pAML: primary AML; sAML: secondary AML; tAML: therapy-related AML; WBC: white blood cells

**Figure 4: Differential isoform expression analysis in the AMLCG cohort.** (A) Number of differentially expressed isoforms for each SF point mutation. (B) Isoforms reported as differentially expressed in the same direction among patients with different SF mutations vs. SF wildtype patients. HGNC symbols of the genes in which the common isoforms are located are shown. The corresponding Ensembl isoform identifiers are: 1. ENST00000340857, 2.
ENST00000481621, 3. ENST00000309668, 4. ENST00000258341, 5. ENST00000426532, 6. ENST00000562151, 7. ENST00000564133, 8. ENST00000368817, 9. ENST00000566524, 10. ENST00000530211. (C) Volcano plots showing the magnitude of differential isoform expression for \textit{SRSF2}(P95H) and \textit{SRSF2}(P95L). The x-axis corresponds to the log\(_2\)(fold change) of each isoform between mutated and wildtype samples, while the y-axis corresponds to –log\(_{10}\)(FDR), where FDR represents the adjusted p-value for each isoform (False Discovery Rate).

**Figure 5: Differential splice junction usage.** (A) Scatterplot displaying the number of samples as well as the total number of reads supporting each splice junction, separately for known and novel splice junctions in both RNA-Seq datasets. To preserve visibility 20000 random junctions are shown for each group. (B) Barchart showing the annotation status of splice junctions reported as differentially used. Novel splice junctions where classified into 5 groups based on their annotation status as described previously[15] (“DA”: annotated junctions, “NDA”: unknown combination of known donor and acceptor sites, “D”: known donor, but novel acceptor site, “A”: novel donor, but known acceptor site, “N”: previously unknown donor and acceptor site). (C) Venn diagram showing the overlap of GO terms between \textit{SRSF2}(P95H) and \textit{SRSF2}(P95L) mutants for the “biogical process” domain.

**Figure 6: Splicing dysregulation in \textit{SRSF2} mutants.** (A) DSJU of all splice junctions inside \textit{IDH3G} (ENSG00000067829) are shown for \textit{SRSF2}(P95H) mutants compared to SF wildtype patients in the AMLCG and Beat AML cohort. To determine significance the log fold-change (logFC) of each splice junction (SJ) is compared to the logFC of all other junctions inside the same gene. The x-axis denotes individual splice junctions defined by their chromosomal coordinates (note the high number of splice sites shared by multiple splice junctions). (B)-(D) Nanopore sequencing results of one \textit{SRSF2}(P95H) mutated sample and one SF wildtype
sample of the AMLCG cohort. The yellow boxes highlight examples of exon skipping (same exons highlighted in (C) and (D)). (B) FLAIR distribution of transcripts. Only one known isoform is expressed in the samples. Additionally, two novel isoforms were detected which are virtually mutually exclusive in the SRSF2(P95H) mutant and the SF wildtype sample. (C) Exon composition of known and novel isoforms detected (D). Sashimi plots showing the exon sequence coverage as well as the splice junction usage of the SRSF2(P95H) mutated and SF wildtype samples. The black arrow indicates the novel splice junction that is differentially used in SRSF2(P95H) mutated samples compared to SF wildtype samples in both RNA-Seq datasets shown in (A). (E) STRING plot depicting an interaction between SRSF2 and several other splicing-related proteins including TRA2A. (F) Differential splice junction usage of all splice junctions inside TRA2A (ENSG00000164548) are shown for SRSF2(P95H) and SRSF2(P95L) mutants compared to SF wildtype patients. Annotation same as (A).
Figure 1

A

SF wildtype (78.6%) other (1.8%) SF3B1 (4.1%) U2AF1 (3.4%) SRSF2 (12.1%)

B

Variant Allele Frequency

C

SRSF2
AMLCG: n = 136
AMLSG: n = 82

U2AF1
AMLCG: n = 38
AMLSG: n = 33

SF3B1
AMLCG: n = 46
AMLSG: n = 34
Figure 2

A

<table>
<thead>
<tr>
<th>Gene</th>
<th>SF3B1(K700E)</th>
<th>SF3B1(P95H)</th>
<th>U2AF1(S34F)</th>
<th>U2AF1</th>
<th>SF3B1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF3B1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>SF3B1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>SF3B1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>SF3B1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>SF3B1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>Gene</th>
<th>SF3B1(K700E)</th>
<th>SF3B1(P95H)</th>
<th>U2AF1(S34F)</th>
<th>U2AF1</th>
<th>SF3B1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF3B1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>SF3B1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>SF3B1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>SF3B1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>SF3B1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

C

<table>
<thead>
<tr>
<th>Gene</th>
<th>SF3B1(K700E)</th>
<th>SF3B1(P95H)</th>
<th>U2AF1(S34F)</th>
<th>U2AF1</th>
<th>SF3B1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF3B1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>SF3B1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>SF3B1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>SF3B1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>SF3B1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

Significance levels:

- * < 0.05
- ** < 0.01
- *** < 0.001

Validated in the AMLSG cohort
### Figure 3

<table>
<thead>
<tr>
<th>Overall Survival</th>
<th>AMLCG cohort</th>
<th>P-Value</th>
<th>AMLSG cohort</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
<td>Hazard Ratio (CI 95%)</td>
<td></td>
<td>Hazard Ratio (CI 95%)</td>
<td></td>
</tr>
<tr>
<td>ANLHD98B vs ANLHD08A trial</td>
<td>-</td>
<td>-</td>
<td>0.68 (0.56 - 0.81)</td>
<td>0.212</td>
</tr>
<tr>
<td>ANL/SO904 vs ANLHD08A trial</td>
<td>-</td>
<td>-</td>
<td>0.40 (0.31 - 0.52)</td>
<td>-</td>
</tr>
<tr>
<td>AMLCG08 vs AMLCG09 trial</td>
<td>0.59 (0.34 - 0.98)</td>
<td>0.002*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>sAML vs pAML</td>
<td>1.25 (0.6 - 2.63)</td>
<td>-</td>
<td>1.61 (1.1 - 2.35)</td>
<td>0.000*</td>
</tr>
<tr>
<td>tAML vs pAML</td>
<td>1.56 (0.6 - 2.25)</td>
<td>0.018*</td>
<td>1.67 (1.14 - 2.00)</td>
<td>0.034*</td>
</tr>
<tr>
<td>Female sex</td>
<td>0.07 (0.00 - 1.17)</td>
<td>0.74</td>
<td>1.01 (0.80 - 1.27)</td>
<td>0.969</td>
</tr>
<tr>
<td>Age (years), per 5 years</td>
<td>1.43 (1.1 - 1.82)</td>
<td>&lt;0.001*</td>
<td>1.39 (1.11 - 1.73)</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>WBC (10^3/μL), per 50 units</td>
<td>1.16 (1.11 - 1.22)</td>
<td>&lt;0.001*</td>
<td>1.37 (1.31 - 1.43)</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>LDH (U/L), per 500 units</td>
<td>1.01 (0.71 - 1.23)</td>
<td>0.015*</td>
<td>1.13 (1.02 - 1.25)</td>
<td>0.011*</td>
</tr>
<tr>
<td>ASXL1</td>
<td>0.96 (0.91 - 1.27)</td>
<td>0.29</td>
<td>1.15 (1.04 - 1.6)</td>
<td>0.907</td>
</tr>
<tr>
<td>BCROR</td>
<td>1.01 (0.73 - 1.38)</td>
<td>0.953</td>
<td>0.56 (0.44 - 0.66)</td>
<td>0.47</td>
</tr>
<tr>
<td>CEBPA (homozygous)</td>
<td>0.44 (0.22 - 0.88)</td>
<td>0.015*</td>
<td>0.45 (0.21 - 0.98)</td>
<td>0.035*</td>
</tr>
<tr>
<td>JAK2</td>
<td>1.97 (0.95 - 4.1)</td>
<td>0.052</td>
<td>0.69 (0.59 - 1.6)</td>
<td>0.418</td>
</tr>
<tr>
<td>KIT</td>
<td>0.13 (0.33 - 1.21)</td>
<td>0.15</td>
<td>0.66 (0.47 - 1.57)</td>
<td>0.247</td>
</tr>
<tr>
<td>KRAS</td>
<td>0.11 (0.1 - 2.48)</td>
<td>0.004*</td>
<td>1.09 (0.92 - 1.61)</td>
<td>0.755</td>
</tr>
<tr>
<td>NPM1</td>
<td>0.66 (0.51 - 0.85)</td>
<td>0.001*</td>
<td>0.78 (0.56 - 1.1)</td>
<td>0.206</td>
</tr>
<tr>
<td>NRAS</td>
<td>0.89 (0.71 - 1.13)</td>
<td>0.334</td>
<td>0.88 (0.64 - 1.23)</td>
<td>0.043</td>
</tr>
<tr>
<td>PTPN11</td>
<td>0.09 (0.06 - 1.25)</td>
<td>0.334</td>
<td>0.30 (0.21 - 0.56)</td>
<td>0.009</td>
</tr>
<tr>
<td>RUNX1</td>
<td>1.01 (0.73 - 1.38)</td>
<td>0.11</td>
<td>1.27 (1.02 - 1.58)</td>
<td>0.032</td>
</tr>
<tr>
<td>SF3B1 (K70E/E)</td>
<td>0.46 (0.35 - 1.76)</td>
<td>0.015*</td>
<td>0.44 (0.37 - 2.3)</td>
<td>0.506</td>
</tr>
<tr>
<td>SRSF1 (other)</td>
<td>1.26 (0.76 - 2.11)</td>
<td>0.57</td>
<td>0.68 (0.47 - 1.57)</td>
<td>0.97</td>
</tr>
<tr>
<td>SRSF2 (P59L)</td>
<td>0.61 (0.50 - 0.85)</td>
<td>0.28</td>
<td>0.44 (0.31 - 0.89)</td>
<td>0.12</td>
</tr>
<tr>
<td>SRSF2 (P59L)</td>
<td>0.99 (0.82 - 1.22)</td>
<td>0.232</td>
<td>1.28 (0.54 - 2.73)</td>
<td>0.041</td>
</tr>
<tr>
<td>STAT2</td>
<td>1.08 (0.78 - 1.5)</td>
<td>0.26</td>
<td>1.35 (0.87 - 2.0)</td>
<td>0.176</td>
</tr>
<tr>
<td>TEKT</td>
<td>0.57 (0.60 - 1.10)</td>
<td>0.23</td>
<td>0.69 (0.58 - 1.49)</td>
<td>0.967</td>
</tr>
<tr>
<td>TPS3</td>
<td>1.25 (0.81 - 2.29)</td>
<td>0.57</td>
<td>1.05 (1.2 - 3.6)</td>
<td>0.019*</td>
</tr>
<tr>
<td>U2AF1 (25-45)</td>
<td>1.00 (0.50 - 2.1)</td>
<td>0.934*</td>
<td>0.30 (0.25 - 1.5)</td>
<td>0.04</td>
</tr>
<tr>
<td>U2AF1 (other)</td>
<td>1.62 (0.61 - 2.05)</td>
<td>0.36</td>
<td>0.57 (0.26 - 2.1)</td>
<td>0.577</td>
</tr>
<tr>
<td>ZIRG2</td>
<td>1.66 (0.55 - 2.36)</td>
<td>0.004*</td>
<td>2.35 (0.92 - 6.03)</td>
<td>0.075</td>
</tr>
<tr>
<td>Cb-AML</td>
<td>0.82 (0.66 - 0.95)</td>
<td>0.015*</td>
<td>0.89 (0.69 - 1.2)</td>
<td>0.206</td>
</tr>
<tr>
<td>Complex AML</td>
<td>0.73 (0.32 - 2.03)</td>
<td>0.74</td>
<td>0.69 (0.49 - 1.0)</td>
<td>0.056</td>
</tr>
<tr>
<td>ins(3)(3.3)</td>
<td>2.05 (1.11 - 1.49)</td>
<td>0.024*</td>
<td>2.07 (0.86 - 5.0)</td>
<td>0.214</td>
</tr>
<tr>
<td>ins(3)(3.3)</td>
<td>3.3 (0.17 - 0.2)</td>
<td>0.001*</td>
<td>0.31 (0.14 - 0.6)</td>
<td>0.004*</td>
</tr>
<tr>
<td>Ins(10)(10)</td>
<td>0.74 (0.22 - 1.2)</td>
<td>0.957</td>
<td>1.28 (0.70 - 2.29)</td>
<td>0.445</td>
</tr>
<tr>
<td>J7</td>
<td>1.11 (0.73 - 1.68)</td>
<td>0.626</td>
<td>1.71 (1.31 - 2.23)</td>
<td>0.035*</td>
</tr>
<tr>
<td>Ins(17)(17)</td>
<td>0.69 (0.24 - 1.04)</td>
<td>0.088</td>
<td>0.29 (0.24 - 1.44)</td>
<td>0.448</td>
</tr>
</tbody>
</table>

The table above lists selected genetic variations and their impact on overall survival and survival in AMLCG and AMLSG cohorts. The hazard ratio and corresponding p-values are provided for each variation, indicating the significance of their association with survival outcomes.
**Figure 4**

A. Number of DE isoforms

- **SRSF2(P95H)**
  - $n_1 = 16$, $n_2 = 5$
  - 121 upregulated, 46 downregulated, 155 unchanged

- **SRSF2(P95L)**
  - $n_1 = 4$, $n_2 = 3$
  - 121 upregulated, 46 downregulated, 155 unchanged

B. Venn diagram showing the overlap of upregulated and downregulated genes

- **U2AF1(S34F)**
  - $n_1 = 3$, $n_2 = 3$

- **SF3B1(K700E)**
  - $n_1 = 3$, $n_2 = 3$

- **SRSF2(P95L)**
  - $n_1 = 4$, $n_2 = 3$

C. Expression levels

- **SRSF2(P95H)**
  - Expression range from -7 to 7

- **SRSF2(P95L)**
  - Expression range from -7 to 7

Expression categories:
- Below Threshold
- Upregulation
- Downregulation
Figure 5

A

B

C

- mRNA splicing, via spliceosome
- mRNA splicing site selection
- Nuclear-translocated mRNA catabolic process nonsense-mediated decay
- Regulation of antinecrotic factor production
- Stress-enhanced MAPK cascade

SRSF2(P95H)  SRSF2(P95L)

101  13  85
Figure 6

IDH3G (ENSG00000067829)

Virtually exclusive alternative transcript usage

Exon skipping

Validated novel SRSF2(P95H) specific splice site

logFC (this SJ vs rest)

Significance

No

Yes

NA

Mutation

SRSF2(P95H)

SRSF2(P95L)

logFC (this SJ vs rest)

Significance

No

Yes