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. Bamopoulos^{1,2}, Aarif M. N. Batcha^{3,4}, Vindi Jurinovic^{1,3}, Maja Rothenberg-Thurley¹, Hanna Janke¹, Bianka Ksienzyk¹, Julia Philippou-Massier⁵, Alexander Graf⁵, Stefan Krebs⁵, Helmut Blum⁵, Stephanie Schneider^{1,6}, Nikola Konstandin¹, Maria Cristina Sauerland⁷, Dennis Görlich⁷, Wolfgang E. Berdel⁸, Bernhard J. Woermann⁹, Stefan K. Bohlander¹⁰, Stefan Canzar¹¹, Ulrich Mansmann^{3,4,12,13}, Wolfgang Hiddemann^{1,12,13}, Jan Braess¹⁴, Karsten Spiekermann^{1,12,13}, Klaus H. Metzeler^{1,12,13} and Tobias Herold^{1,12,15}

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

- 24 7 Institute of Biostatistics and Clinical Research, University of Münster, Münster, Germany
- 25 8 Department of Medicine, Hematology and Oncology, University of Münster, Münster, Ger-
- 26 many
- 27 9 German Society of Hematology and Oncology, Berlin, Germany
- 28 10 Leukaemia and Blood Cancer Research Unit, Department of Molecular Medicine and Pa-
- 29 thology, University of Auckland, Auckland, New Zealand
- 30 11 Gene Center, LMU Munich, Munich, Germany
- 31 12 German Cancer Consortium (DKTK), Partner Site Munich, Munich, Germany
- 32 13 German Cancer Research Center (DKFZ), Heidelberg, Germany
- 33 14 Department of Oncology and Hematology, Hospital Barmherzige Brüder, Regensburg,
- 34 Germany
- 35 15 Research Unit Apoptosis in Hematopoietic Stem Cells, Helmholtz Zentrum München,
- 36 German Center for Environmental Health (HMGU), Munich, Germany

37 **Corresponding Authors:**

38 Stefanos A. Bamopoulos; stefanos.bamopoulos@charite.de

39 **and**

40 Tobias Herold, MD; Marchioninstr. 15; 81377 Munich; Germany; Phone: +49 89 4400-0; FAX:
41 +49 89 4400-74242; Email: tobias.herold@med.uni-muenchen.de

42

43 **Abstract**

44 Previous studies demonstrated that splicing factor mutations are recurrent events in
 45 hematopoietic malignancies with both clinical and functional implications. However, their
 46 aberrant splicing patterns in acute myeloid leukaemia remain largely unexplored. In this study
 47 we characterized mutations in *SRSF2*, *U2AF1* and *SF3B1*, the most commonly mutated
 48 splicing factors. In our clinical analysis of 2678 patients, splicing factor mutations showed
 49 inferior relapse-free and overall survival, however, these mutations did not represent
 50 independent prognostic markers. RNA-sequencing of 246 and independent validation in 177
 51 patients revealed an isoform expression profile highly characteristic for each individual
 52 mutation, with several isoforms showing a strong dysregulation. By establishing a custom
 53 differential splice junction usage pipeline we accurately detected aberrant splicing in splicing
 54 factor mutated samples. Mutated samples were characterized predominantly by decreased
 55 junction usage. A large proportion of differentially used junctions were novel. Targets of
 56 splicing dysregulation included several genes with a known role in leukaemia. In
 57 *SRSF2*(P95H) mutants we further explored the possibility of a cascading effect through the
 58 dysregulation of the splicing pathway. We conclude that splicing factor mutations do not
 59 represent independent prognostic markers. However, they do have genome-wide
 60 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

85 aberrant splicing due to SF mutations has emerged for many genes relevant in AML, it is yet
 86 uncertain whether and how these changes directly influence disease initiation or evolution.
 87 The aim of this study was a comprehensive analysis of the prognostic implications of SF
 88 mutations in two well-characterized and intensively treated adult AML patient cohorts
 89 amounting to a total of 2678 patients. In addition, the core functional consequences of SF
 90 mutations were explored using targeted amplicon sequencing in conjunction with RNA-
 91 sequencing on two large datasets.

92 **Patients and Methods**

93 **Patients**

94 Our primary cohort included a total of 1138 AML patients treated with intensive chemotherapy
 95 in two randomized multicenter phase 3 trials of the German AML Cooperative Group
 96 (AMLCG). Treatment regimens and inclusion criteria are described elsewhere [2]. A cohort of
 97 1540 AML patients participating in multicenter clinical trials of the German-Austrian AML
 98 Study Group (AMLSG), was used for validation [19]. Cohort composition and filtering criteria
 99 are outlined in the supplementary.

100 **Molecular Workup**

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

104 **RNA-Sequencing and data processing**

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

116

117 **Differential expression analysis and differential splice junction usage (DSJU)**

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

124 **Nanopore cDNA sequencing and analysis**

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

130 **Statistics**

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

140 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 $13.3 \times 10^9/L$ vs. $22.4 \times 10^9/L$) 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 *U2AF1*(R35Q) mutation. The same held true for t(8;21) patients (n=98), where only one patient had a rare deletion in *SRSF2*. Additionally, all patients in the AMLCG cohort presenting with an isolated trisomy 13 (n=9) also harbored an *SRSF2* mutation (p<0.001), as described previously [31].

Mutations in all SF genes correlated positively with mutations in *BCOR* and *RUNX1* and negatively with mutations in *NPM1*. Expectedly, *SRSF2*(P95H) and *SRSF2*(P95L) mutations shared a similar pattern of co-expression including significant pairwise associations with mutations in *ASXL1*, *IDH2*, *RUNX1* (both p<0.001) and *STAG2* (p<0.001 and p=0.002, respectively). However, apart from *IDH2* mutations where co-occurrence was comparable (OR: 3.4 vs 5.1), mutations in *ASXL1*, *RUNX1* and *STAG2* coincided more frequently with *SRSF2*(P95H) mutations. Despite this, *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 *SRSF2*, *U2AF1* and *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 *U2AF1* mutated patients with an one-year survival rate of only 29.1%, followed by *SF3B1* (40.6%) and *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 *SRSF2*(P95H), *SRSF2*(P95L), *U2AF1*(S34F) and

189 *SF3B1*(K700E) mutations had significantly reduced RFS and OS (Figures S3.1D and S4.1).
 190 To test whether any SF mutation was an independent prognostic marker, multiple Cox
 191 regression models (Figure 3) were built by integrating all parameters significantly associated
 192 ($p < 0.1$) with RFS and OS in the single Cox regression models. Along with several known
 193 predictors, only *U2AF1*(S34F) mutations presented with prognostic relevance for both RFS
 194 (Hazard Ratio=2.81, $p=0.012$) and OS (HR=1.90, $p=0.034$) in the AMLCG cohort, but not in
 195 the AMLSG cohort. However, when aggregating mutations at the gene level, mutations in
 196 *SRSF2* and *SF3B1* presented with prognostic relevance for RFS in the AMLSG cohort
 197 (HR=1.77, $p=0.008$; HR=2.15, $p=0.014$; respectively), while not reaching significance in the
 198 AMLCG cohort (Table S4.2). When looking only at de novo AML patients, the prognostic
 199 impact of *U2AF1*(S34F) mutations diminished, yet the prognostic impact observed for *SRSF2*
 200 and *SF3B1* remained significant in the AMLSG cohort (HR=1.84, $p=0.009$; HR=2.43, $p=0.015$;
 201 respectively) (Tables S4.1-S4.2).

202 **Differential isoform expression in SF mutated patients**

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

210 After low-coverage filtering we performed a differential isoform expression analysis for ~90
 211 000 isoforms. Differential expression was restricted to a small fraction of all expressed
 212 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*, *H1FO*, *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 *H1FO* 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, 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 ($p < 0.05$, $\log_2(\text{fold change}) > 1$), 20.2-45.9% constituted novel junctions (Tables S7.1-S7.3 and Tables

261 S9.1-S9.2) and were classified according to their relationship with known acceptor and donor
262 sites as described previously (Figure 5B) [15]. Unsurprisingly, validation rates increased with
263 larger mutant sample sizes, ranging from 9.3% (*SF3B1*(K700E); n = 3) to 74.0% (all *SRSF2*
264 mutants; n = 26). Furthermore, validation rates were higher for novel junctions (mean 39.3%
265 vs. 21.5% known junctions), likely due to the stricter initial filtering criteria applied. By
266 performing nanopore sequencing of one *SRSF2*(P95H) mutant and one SF wildtype sample
267 we were able to confirm the usage of several novel junctions and detect resulting novel
268 isoforms as exemplified for *IDH3G* in Figure 6A-D. A tendency towards decreased junction
269 usage was observed for all SF point mutations and was most evident in *SF3B1*(K700E)
270 mutants (1 423/1 927; 73.9% of differentially used junctions). The total number of splicing
271 events, however, was not reduced in SF mutants (mean 9,275,359 events vs. 9 192 697 in
272 wildtype patients), suggesting that decreased splicing is limited to selected junctions rather
273 than being a global effect.

274 We systematically compared the genes with at least one DE isoform and those reported as
275 differentially spliced in all SF mutation subgroups (Tables S9.3-S9.4). For *SRSF2* mutants,
276 genes significant in both analyses included *EWSR1*, *H1FO*, *INTS3* and *YBX1*. In general, out
277 of the genes examined in both analyses only 9.8-23.3% (depending on the SF mutation) of
278 genes reported as having a DE isoform were also reported as being differentially spliced.
279 Conversely, 3.3-28.5% of differentially spliced genes were also reported as having a DE
280 isoform. These findings suggest that differential gene splicing does not always lead to altered
281 isoform expression while at the same time differential isoform expression cannot always be
282 attributed to an explicit splicing alteration. Considering the complementary nature of the
283 analyses, we performed gene ontology (GO) analysis by combining the genes with evidence
284 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

309 347 splicing-related genes (317 of which were expressed in our dataset) 101 were
310 dysregulated in at least one SF mutant group. On average 30.5 (range 6-52) splicing-related
311 genes were dysregulated per SF point mutation. Of note, both *SRSF2* point mutations
312 associated with differential splicing of *HNRNPA1* and *HNRNPUL1*, as well as *PCF11* and
313 *TRA2A*. A query of the STRING database suggests protein-protein interactions between
314 SRSF2 and the proteins of the above genes, which are also interconnected (Figure 6E).
315 Interestingly, one of the differential splicing events reported in both *SRSF2* mutants involves
316 the under-usage of the same novel splice junction in *TRA2A* (Figure 6F). *TRA2A* has
317 previously been shown to be differentially spliced in mouse embryo fibroblasts upon *SRSF2*
318 knockout [45]. Furthermore, it has been shown that both HNRNPA1 and SRSF2 interact with
319 the loop 3 region of 7SK RNA and by favoring the dissociation of SRSF2, HNRNPA1 may
320 lead to the release of active P-TEFb [46]. Taken together, our results indicate a strong
321 dysregulation of the splicing pathway in SF mutants including several genes whose gene
322 products closely interact with SRSF2.

323
324

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-sequencing 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

349 define prognosis and transformation risk in CHIP and MDS patients, yet there is no clear
 350 independent prognostic value of SF mutations in AML.

351 Two large RNA-sequencing studies have been performed previously, to detect aberrantly
 352 spliced genes in SF mutants, both of which focused on MDS patients [41,42]. In this study we
 353 described a distinct differential isoform expression profile for each SF point mutation.
 354 Furthermore, we evaluated differential splicing for the four most common SF point mutations
 355 via a customized pipeline to determine differential usage of both known and novel splice
 356 junctions. Our pipeline enables the differential quantification of individual splice junctions
 357 without restricting the analysis to annotated alternative splicing events. We argue that the
 358 strength of our analysis lies in the accurate detection of single dysregulated junctions
 359 (especially in cases where splice sites are shared by multiple junctions) in an annotation-
 360 independent manner achieving validation rates up to 74.0% in our largest mutant sample
 361 group (*SRSF2*, n=19). Limitations of the analysis include the restriction to junctions with both
 362 splice sites within the same gene (a restriction shared by most differential splicing algorithms)
 363 and genes with at least two junctions. However, the reduced requirements of our analysis
 364 could prove valuable in the study of differential splicing in organisms with lacking annotation.

365 All SF point mutations shared a tendency towards decreased splice junction usage, which did
 366 not affect the global number of splicing events in SF mutants. Surprisingly, we observed a
 367 limited overlap between genes with differentially expressed isoforms and differentially spliced
 368 genes. In addition, a recent study by *Liang et al.* reported that the majority of differential
 369 binding events in *SRSF2*(P95H) mutants do not translate to alternative splicing [53]. Taken
 370 together, these findings indicate a “selection” or possibly a compensation of deregulatory
 371 events from differential binding through differential splicing to finally differential isoform
 372 expression. Furthermore, the enrichment of aberrant splicing in splicing-related genes opens

373 the possibility of a cascading effect on transcription via the differential alternative splicing of
 374 transcriptional components. A congruent hypothesis was stated by *Liang et al.*, where an
 375 enrichment of *SRSF2*(P95H) targets in RNA processing and splicing was shown, further
 376 supporting the notion of an indirect effect of mutant *SRSF2* facilitated through additional
 377 splicing components. Further investigations may provide a mechanistic link between the
 378 differential splicing of selected genes and the impairment of transcription and specifically
 379 transcriptional pausing observed in SF mutant cells, which contributes to the MDS phenotype
 380 [32].

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

393 **Acknowledgments**

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

396

397 **Funding**

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

405

406 **Author Contributions**

407 S.A.B., A.M.N.B. and T.H. conceived and designed the analysis. S.A.B., A.M.N.B., V.J., M.R.-
408 T., H.J., A.G., S.C., N.K., K.S., K.H.M. and T.H. provided and analyzed data. A.M.N.B., V.J.
409 and U.M. provided bioinformatics support. J.P.-M., S.K. and H.B. managed the Genome
410 Analyzer Ix platform and the RNA-sequencing of the AMLCG samples. M.R.-T., H.J., B.K.,
411 S.S., N.K., S.K.B., K.H.M. and K.S. characterized patient samples; M.C.S., D.G., W.B., B.W.,
412 J.B. and W.H. coordinated the AMLCG clinical trials. S.A.B. and T.H. wrote the manuscript. All
413 authors approved the final manuscript.

414

415 **Additional Information**

416 The authors declare no conflicts of interest.

References

1. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. 2013;122(22):3616–27.
2. Metzeler KH, Herold T, Rothenberg-Thurley M, Amler S, Sauerland MC, Görlich D, et al. Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia. 2016;128(5):686–98.
3. Makishima H, Visconte V, Sakaguchi H, Jankowska AM, Kar SA, Jerez A, et al. Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. 2012;119(14):3203–10.
4. Larsson CA, Cote G, Quintás-Cardama A. The changing mutational landscape of acute myeloid leukemia and myelodysplastic syndrome. 2013;11(8):815–27.
5. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. 2011;478(7367):64–9.
6. Thol F, Kade S, Schlarmann C, Löffeld P, Morgan M, Krauter J, et al. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. 2012;119(15):3578–84.
7. Dolatshad H, Pellagatti A, Fernandez-Mercado M, Yip BH, Malcovati L, Attwood M, et al. Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. 2015;29(5):1092–103.
8. Wu S, Kuo Y, Hou H, Li L, Tseng M, Huang C, et al. The clinical implication of SRSF2 mutation in patients with myelodysplastic syndrome and its stability during disease

- 440 evolution. 2014;120(15):3106–12.
- 441 9. Graubert TA, Shen D, Ding L, Okeyo-owuor T, Cara L, Shao J, et al. Recurrent
442 Mutations in the U2AF1 Splicing Factor In Myelodysplastic Syndromes. 2012;44(1):53–
443 7.
- 444 10. Hou H-A, Liu C-Y, Kuo Y-Y, Chou W-C, Tsai C-H, Lin C-C, et al. Splicing factor
445 mutations predict poor prognosis in patients with de novo acute myeloid leukemia.
446 2016;7(8).
- 447 11. Cho Y-U, Jang S, Seo E-J, Park C-J, Chi H-S, Kim D-Y, et al. Preferential occurrence of
448 spliceosome mutations in acute myeloid leukemia with a preceding myelodysplastic
449 syndrome and/or myelodysplasia morphology. 2014;8194(November):1–25.
- 450 12. Moon H, Cho S, Loh TJ, Jang HN, Liu Y, Choi N, et al. SRSF2 directly inhibits intron
451 splicing to suppresses cassette exon inclusion. 2017;50(8):423–8.
- 452 13. Kim E, Ilagan JO, Liang Y, Daubner GM, Lee SCW, Ramakrishnan A, et al. SRSF2
453 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon
454 Recognition. 2015;27(5):617–30.
- 455 14. Alsafadi S, Houy A, Battistella A, Popova T, Wassef M, Henry E, et al. Cancer-
456 associated SF3B1 mutations affect alternative splicing by promoting alternative
457 branchpoint usage. 2016;7:10615.
- 458 15. Okeyo-Owuor T, White BS, Chatrikhi R, Mohan DR, Kim S, Griffith M, et al. U2AF1
459 mutations alter sequence specificity of pre-mRNA binding and splicing. 2015;29(4):909–
460 17.
- 461 16. Przychodzen B, Jerez A, Guinta K, Sekeres MA, Padgett R, Maciejewski JP, et al.

- 462 Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms.
463 2013;122(6):999–1006.
- 464 17. Shirai CL, Ley JN, White BS, Kim S, Tibbitts J, Shao J, et al. Mutant U2AF1 Expression
465 Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. 2015;27(5):631–43.
- 466 18. Yang J, Yao D, Ma J, Yang L, Guo H, Wen X, et al. The prognostic implication of SRSF2
467 mutations in Chinese patients with acute myeloid leukemia. 2016;37(8):10107–14.
- 468 19. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al.
469 Genomic Classification and Prognosis in Acute Myeloid Leukemia. 2016;374(23):2209–
470 21.
- 471 20. Tyner JW, Tognon CE, Bottomly D, Wilmot B, Kurtz SE, Savage SL, et al. Functional
472 genomic landscape of acute myeloid leukaemia. 2018;562(7728):526–31.
- 473 21. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast
474 universal RNA-seq aligner. 2013;29(1):15–21.
- 475 22. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-
476 aware quantification of transcript expression. 2017;14(4):417–9.
- 477 23. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential
478 expression analyses for RNA-sequencing and microarray studies. 2015;43(7):e47–e47.
- 479 24. Robinson MD, Oshlack A. A scaling normalization method for differential expression
480 analysis of RNA-seq data. 2010;11(3):R25.
- 481 25. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential
482 expression analysis of digital gene expression data. 2009;26(1):139–40.
- 483 26. Liu R, Holik AZ, Su S, Jansz N, Chen K, Leong HS an, et al. Why weight? Modelling

- 484 sample and observational level variability improves power in RNA-seq analyses.
485 2015;43(15):e97.
- 486 27. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model
487 analysis tools for RNA-seq read counts. 2014;15(2):R29.
- 488 28. Leek JT. Svaseq: Removing batch effects and other unwanted noise from sequencing
489 data. 2014;42(21):e161.
- 490 29. Tang AD, Soulette CM, Baren MJ van, Hart K, Hrabeta-Robinson E, Wu CJ, et al. Full-
491 length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia
492 reveals downregulation of retained introns. 2018;410183.
- 493 30. R Core Team. R: A Language and Environment for Statistical Computing [Internet].
494 Vienna, Austria: R Foundation for Statistical Computing;
- 495 31. Herold T, Metzeler KH, Vosberg S, Hartmann L, Ollig C, Olzel FS, et al. Isolated trisomy
496 13 defines a homogeneous AML subgroup with high frequency of mutations in
497 spliceosome genes and poor prognosis. 2014;124(8):1304–11.
- 498 32. Chen L, Chen J-Y, Huang Y-J, Gu Y, Qiu J, Qian H, et al. The Augmented R-Loop Is a
499 Unifying Mechanism for Myelodysplastic Syndromes Induced by High-Risk Splicing
500 Factor Mutations. 2018;69(3):412-425.e6.
- 501 33. Yoshimi A, Lin K-T, Wiseman DH, Rahman MA, Pastore A, Wang B, et al. Coordinated
502 alterations in RNA splicing and epigenetic regulation drive leukaemogenesis.
503 2019;574(7777):273–7.
- 504 34. Roe J-S, Vakoc CR. The Essential Transcriptional Function of BRD4 in Acute Myeloid
505 Leukemia. 2016;81:61–6.

- 506 35. Endo A, Tomizawa D, Aoki Y, Morio T, Mizutani S, Takagi M. EWSR1/ELF5 induces
507 acute myeloid leukemia by inhibiting p53/p21 pathway. 2016;107(12):1745–54.
- 508 36. Perner F, Jayavelu AK, Schnoeder TM, Mashamba N, Mohr J, Hartmann M, et al. The
509 Cold-Shock Protein Ybx1 Is Required for Development and Maintenance of Acute
510 Myeloid Leukemia (AML) in Vitro and In Vivo. 2017;130(Suppl 1).
- 511 37. McNerney ME, Brown CD, Wang X, Bartom ET, Karmakar S, Bandlamudi C, et al.
512 CUX1 is a haploinsufficient tumor suppressor gene on chromosome 7 frequently
513 inactivated in acute myeloid leukemia. 2013;121(6):975–83.
- 514 38. McGarvey T, Rosonina E, McCracken S, Li Q, Arnaout R, Mientjes E, et al. The acute
515 myeloid leukemia-associated protein, DEK, forms a splicing-dependent interaction with
516 exon-product complexes. 2000;150(2):309–20.
- 517 39. Fujita S, Honma D, Adachi N, Araki K, Takamatsu E, Katsumoto T, et al. Dual inhibition
518 of EZH1/2 breaks the quiescence of leukemia stem cells in acute myeloid leukemia.
519 2018;32(4):855–64.
- 520 40. Pallarès V, Hoyos M, Chillón MC, Barragán E, Prieto Conde MI, Llop M, et al. Focal
521 Adhesion Genes Refine the Intermediate-Risk Cytogenetic Classification of Acute
522 Myeloid Leukemia. 2018;10(11).
- 523 41. Shiozawa Y, Malcovati L, Galli A, Sato-Otsubo A, Kataoka K, Sato Y, et al. Aberrant
524 splicing and defective mRNA production induced by somatic spliceosome mutations in
525 myelodysplasia. 2018;9(1):3649.
- 526 42. Pellagatti A, Armstrong RN, Steeples V, Sharma E, Repapi E, Singh S, et al. Impact of
527 spliceosome mutations on RNA splicing in myelodysplasia: Dysregulated
528 genes/pathways and clinical associations. 2018;132(12):1225–40.

- 529 43. Ji X, Zhou Y, Pandit S, Huang J, Li H, Lin CY, et al. SR Proteins Collaborate with 7SK
530 and Promoter-Associated Nascent RNA to Release Paused Polymerase.
531 2013;153(4):855–68.
- 532 44. Zhang J, Lieu YK, Ali AM, Penson A, Reggio KS, Rabadan R, et al. Disease-associated
533 mutation in SRSF2 misregulates splicing by altering RNA-binding affinities.
534 2015;112(34):E4726–34.
- 535 45. Skrdlant L, Stark JM, Lin R-J. Myelodysplasia-associated mutations in serine/arginine-
536 rich splicing factor SRSF2 lead to alternative splicing of CDC25C. 2016;17(1):18.
- 537 46. Lemieux B, Blanchette M, Monette A, Mouland AJ, Wellinger RJ, Chabot B. A Function
538 for the hnRNP A1/A2 Proteins in Transcription Elongation. Caputi M, editor.
539 2015;10(5):e0126654.
- 540 47. Abelson S, Collord G, Ng SWK, Weissbrod O, Mendelson Cohen N, Niemeyer E, et al.
541 Prediction of acute myeloid leukaemia risk in healthy individuals. 2018;559(7714):400–
542 4.
- 543 48. Malcovati L, Papaemmanuil E, Bowen DT, Boulwood J, Della Porta MG, Pascutto C, et
544 al. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and
545 myelodysplastic/myeloproliferative neoplasms. 2011;118(24):6239–46.
- 546 49. Papaemmanuil E, Cazzola M, Boulwood J, Malcovati L, Vyas P, Bowen D, et al.
547 Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. 2011;365(15):1384–
548 95.
- 549 50. Wu L, Song L, Xu L, Chang C, Xu F, Wu D, et al. Genetic landscape of recurrent
550 ASXL1, U2AF1, SF3B1, SRSF2, and EZH2 mutations in 304 Chinese patients with
551 myelodysplastic syndromes. 2016;37(4):4633–40.

- 552 51. Zhang S-J, Rampal R, Manshouri T, Patel J, Mensah N, Kayserian A, et al. Genetic
553 analysis of patients with leukemic transformation of myeloproliferative neoplasms
554 shows recurrent SRSF2 mutations that are associated with adverse outcome.
555 2012;119(19):4480–5.
- 556 52. Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T, et al.
557 Diagnosis and management of AML in adults: 2017 ELN recommendations from an
558 international expert panel [Internet]. Vol. 129, Blood. 2017 [cited 2019 Feb 14]. p. 424–
559 47.
- 560 53. Liang Y, Tebaldi T, Rejeski K, Joshi P, Stefani G, Taylor A, et al. SRSF2 mutations drive
561 oncogenesis by activating a global program of aberrant alternative splicing in
562 hematopoietic cells. 2018;
- 563 54. Lee SCW, Abdel-Wahab O. Therapeutic targeting of splicing in cancer [Internet]. Vol.
564 22, Nature Medicine. NIH Public Access; 2016 [cited 2019 Jan 31]. p. 976–86.

565

566

567

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

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

568

569 WBC: white blood cells, LDH: lactate dehydrogenase, BM: bone marrow, ECOG: Eastern Co-
570 operative Oncology Group performance score

571

572

573

574

575 **Figure Legends**

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

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

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

594 **Figure 4: Differential isoform expression analysis in the AMLCG cohort.** (A) Number of
595 differentially expressed isoforms for each SF point mutation. (B) Isoforms reported as
596 differentially expressed in the same direction among patients with different SF mutations vs.
597 SF wildtype patients. HGNC symbols of the genes in which the common isoforms are located
598 are shown. The corresponding Ensembl isoform identifiers are: 1. ENST00000340857, 2.

599 ENST00000481621, 3. ENST00000309668, 4. ENST00000258341, 5. ENST00000426532, 6.
600 ENST00000562151, 7. ENST00000564133, 8. ENST00000368817, 9. ENST00000566524,
601 10. ENST00000530211. (C) Volcano plots showing the magnitude of differential isoform
602 expression for *SRSF2*(P95H) and *SRSF2*(P95L). The x-axis corresponds to the $\log_2$ (fold
603 change) of each isoform between mutated and wildtype samples, while the y-axis
604 corresponds to $-\log_{10}(\text{FDR})$, where FDR represents the adjusted p-value for each isoform
605 (False Discovery Rate).

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

616 **Figure 6: Splicing dysregulation in *SRSF2* mutants.** (A) DSJU of all splice junctions inside
617 *IDH3G* (ENSG00000067829) are shown for *SRSF2*(P95H) mutants compared to SF wildtype
618 patients in the AMLCG and Beat AML cohort. To determine significance the log fold-change
619 (logFC) of each splice junction (SJ) is compared to the logFC of all other junctions inside the
620 same gene. The x-axis denotes individual splice junctions defined by their chromosomal
621 coordinates (note the high number of splice sites shared by multiple splice junctions). (B)-(D)
622 Nanopore sequencing results of one *SRSF2*(P95H) mutated sample and one SF wildtype

623 sample of the AMLCG cohort. The yellow boxes highlight examples of exon skipping (same
624 exons highlighted in (C) and (D). (B) FLAIR distribution of transcripts. Only one known isoform
625 is expressed in the samples. Additionally, two novel isoforms were detected which are virtually
626 mutually exclusive in the *SRSF2*(P95H) mutant and the SF wildtype sample. (C) Exon
627 composition of known and novel isoforms detected (D). Sashimi plots showing the exon
628 sequence coverage as well as the splice junction usage of the *SRSF2*(P95H) mutated and SF
629 wildtype samples. The black arrow indicates the novel splice junction that is differentially used
630 in *SRSF2*(P95H) mutated samples compared to SF wildtype samples in both RNA-Seq
631 datasets shown in (A). (E) STRING plot depicting an interaction between SRSF2 and several
632 other splicing-related proteins including *TRA2A*. (F) Differential splice junction usage of all
633 splice junctions inside *TRA2A* (ENSG00000164548) are shown for *SRSF2*(P95H) and
634 *SRSF2*(P95L) mutants compared to SF wildtype patients. Annotation same as (A).

635

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6