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-Thurley[1], Hanna Janke[1], Bianka Ksienzyk[1], Julia Philippou-Massier[5], Alexander Graf[5], Stefan Krebs[5], Helmut Blum[5], Stephanie Schneider[1,6], Nikola Konstandin[1], Maria Cristina Sauerland[7], Dennis Görlich[7], Wolfgang E. Berdel[8], Bernhard J. Woermann[9], Stefan K. Bohlander[10], Stefan Canzar[11], Ulrich Mansmann[3,4,12,13], Wolfgang Hiddemann[1,12,13], Jan Braess[14], 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

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@protonmail.com

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 have demonstrated that splicing factor mutations are recurrent events in hematopoietic malignancies. Their clinical characteristics and aberrant splicing patterns have been explored in myelodysplasia, however, their functional consequences in acute myeloid leukaemia are largely unknown. The aim of this study was the comprehensive clinical and functional analysis of mutations in the three most commonly afflicted splicing factor genes: *SRSF2*, *U2AF1* and *SF3B1*. To this end, we examined the prognostic role of splicing factor mutations in two large independent cohorts, encompassing a total of 2678 acute myeloid leukaemia patients treated with intensive chemotherapy. The clinical analysis was complemented by RNA-sequencing of 246 patients to identify targets of splicing dysregulation. Results were validated in an additional RNA-sequencing dataset of 177 patients. Patients with splicing factor mutations show inferior relapse-free and overall survival, however, these mutations do not represent independent prognostic markers. Differential isoform expression analysis revealed a characteristic expression profile for each splicing factor mutation with a strong dysregulation of several isoforms. Furthermore, by establishing a custom differential splice junction usage pipeline we accurately detected aberrant splicing in splicing factor mutated samples. Mutated samples were characterized by predominantly decreased splice junction utilization of a large number of genes. A large proportion of differentially used spliced junctions were novel. Targets of splicing dysregulation included several genes with a known role in acute myeloid leukaemia. In *SRSF2*(P95H) mutants we further explored the possibility of a cascading effect through the dysregulation of the splicing pathway.

Taken together, our findings suggest that splicing factor mutations does 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.

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

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

129 **Nanopore cDNA sequencing and analysis**

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

135 **Statistics**

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

146 **Results**

147 **Clinical features of AML patients with SF mutations**

148 We characterized SF mutations in two independent patient cohorts (the AMLCG and AMLSG
149 cohorts). Our primary cohort (AMLCG) consisted of 1119 AML patients (Figure S1), 236 (21.1%)
150 of which presented with SF mutations. The three most commonly affected SF genes, *SRSF2*,
151 *U2AF1* and *SF3B1* were mutated in 12.1% (n=136), 3.4% (n=38) and 4.1% (n=46) of the
152 patients (Figure 1A). In agreement with previous findings [19], SF mutations were in their
153 majority mutually exclusive, heterozygous hotspot mutations (Figure 1B). The four most
154 common point mutations were *SRSF2*(P95H) (n=69), *SRSF2*(P95L) (n=27), *U2AF1*(S34F)
155 (n=18), and *SF3B1*(K700E) (n=18) mutations (Figure 1C-E). The clinical characteristics of
156 patients harboring SF mutations are summarized in Tables 1 and S1 (AMLSG cohort), along
157 with a statistical assessment between cohorts (Table S2). We observed a high overall degree
158 of similarity regarding clinical features of SF mutated patients between the AMLCG and AMLSG
159 cohorts, despite their large median age difference. Mutations in *SRSF2*, *U2AF1* and *SF3B1*
160 occurred more frequently in secondary AML (44.7% compared to 18.2% in *de novo* AML) and
161 were all associated with older age. As reported previously [1], *SRSF2* and *U2AF1* mutated
162 patients were predominantly male (76.7% and 76.3%, respectively). Furthermore, patients
163 harboring *SRSF2* mutations presented with a lower white blood cell count (WBC; median 13.3
164 10^9/L vs. 22.4 10^9/L) while *U2AF1* mutated patients presented with a reduced blast percentage
165 in their bone marrow when compared to SF wildtype patients (median 60% vs 80%).

166 **Associations of SF mutations and other recurrent alterations in AML**

167 In a second step, we investigated associations between SF mutations and recurrent cytogenetic
168 abnormalities and gene mutations in AML (Figure 2). Notably, SF mutations were not found in
169 inv(16)/t(16;16) patients (n=124), with the exception of one inv(16)/t(16;16) patient harboring a

170 *U2AF1*(R35Q) mutation. The same held true for t(8;21) patients (n=98), where only one patient
171 had a rare deletion in *SRSF2*. Additionally, all patients in the AMLCG cohort presenting with an
172 isolated trisomy 13 (n=9) also harbored an *SRSF2* mutation (p<0.001), as described previously
173 [31].
174 Mutations in all SF genes correlated positively with mutations in *BCOR* and *RUNX1* and
175 negatively with mutations in *NPM1*. Expectedly, *SRSF2*(P95H) and *SRSF2*(P95L) mutations
176 shared a similar pattern of co-expression including significant pairwise associations with
177 mutations in *ASXL1*, *IDH2*, *RUNX1* (both p<0.001) and *STAG2* (p<0.001 and p=0.002,
178 respectively). However, apart from *IDH2* mutations where co-occurrence was comparable (OR:
179 3.4 vs 5.1), mutations in *ASXL1*, *RUNX1* and *STAG2* coincided more frequently with
180 *SRSF2*(P95H) mutations. Despite this, *SRSF2*(P95L) mutations showed a slightly increased
181 co-occurrence with other recurrent AML mutations (median 5 vs 4 mutations, p=0.046).
182 **Prognostic relevance of SF mutations for relapse-free survival and overall survival**
183 The prognostic impact of *SRSF2*, *U2AF1* and *SF3B1* mutations was initially assessed using
184 Kaplan-Meier graphs and log-rank testing. All SF mutations presented with both inferior relapse-
185 free survival (RFS) and overall survival (OS) compared to SF wildtype patients (Figures S3.1A-
186 C and S3.2). The effect was most pronounced in *U2AF1* mutated patients with an one-year
187 survival rate of only 29.1%, followed by *SF3B1* (40.6%) and *SRSF2* mutated patients (49.2%).
188 Different point mutations inside the same SF gene did not differ significantly in their effect on
189 OS.
190 To confirm the observed prognostic impact of SF mutations we performed single Cox
191 regressions on all available clinical and genetic parameters. In agreement with the Kaplan-
192 Meier estimates, patients harboring *SRSF2*(P95H), *SRSF2*(P95L), *U2AF1*(S34F) and
193 *SF3B1*(K700E) mutations had significantly reduced RFS and OS (Figures S3.1D and S4.1). To

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

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

207 We next assessed the impact of SF mutations on mRNA expression. To this end, whole-
208 transcriptome RNA-sequencing was performed on 246 AML patients, 29 of which harbored a
209 mutation in the SF genes of interest, while 199 SF wildtype patients were used as a control
210 (Figure S2). The remaining patients either presented with a different SF mutation (n=17) or
211 exhibited more than one SF mutation (n=1) and were excluded. In addition, a subset of the Beat
212 AML cohort (n=177) with matched DNA- and RNA-sequencing data was used for validation
213 [20].

214 After low-coverage filtering we performed a differential isoform expression analysis for ~90 000
215 isoforms. Differential expression was restricted to a small fraction of all expressed isoforms
216 (<0.5%; Figure 4A and Table S6.1). Little overlap of differentially expressed (DE) isoforms was
217 found when different SF mutation groups were compared to the control, consistent with previous

218 observations [33]. However, 10 isoforms were reported as DE in both *SRSF2*(P95H) and
219 *SRSF2*(P95L) mutated samples, all with the same fold-change direction (Figure 4B). Out of
220 those, the isoforms in *GTF2I*, *H1F0*, *INHBC*, *LAMC1* and one of the isoforms of *METTL22*
221 (*ENST00000562151*) were also significant in the validation cohort for both *SRSF2*(P95H) and
222 *SRSF2*(P95L). Additionally, the isoform of *H1F0* was also reported as DE for *U2AF1*(S34F)
223 mutants in both cohorts. For *SRSF2*(P95H) mutants 107 of all DE isoforms also reached
224 significance in the validation cohort (40.1%), while for the other SF mutation subgroups
225 validation rates ranged from 15.1 to 27.3% increasing with larger mutant sample sizes. Notably,
226 mutated and wildtype samples showed large differences in the expression levels of several
227 isoforms (Figure 4C and Figure S5). The top two overexpressed isoforms in *SRSF2*(P95H) both
228 corresponded to *INTS3*, which was recently reported as dysregulated in *SRSF2*(P95H) mutants
229 co-expressing *IDH2* mutations [32]. Several DE isoforms identified in SF mutated patients
230 correspond to cancer-related genes, many of which have a known role in AML. Specifically,
231 genes with DE isoforms included, but were not limited to *BRD4* [34], *EWSR1* [35] and *YBX1*
232 [36] in *SRSF2*(P95H) mutated samples, *CUX1* [37], *DEK* [15,38] and *EZH1* [39] in
233 *U2AF1*(S34F) mutated samples, as well as *PTK2* [40] in *SF3B1*(K700E) mutated patients
234 (Tables S7.1-S7.2).
235 Hierarchical clustering using DE isoforms was performed for all samples to assess the
236 expression homogeneity of SF mutations. A tight clustering of samples harboring identical SF
237 point mutations was observed, indicating an isoform expression profile highly characteristic for
238 each individual SF mutation (Figures S6.1-S6.3). When using DE isoforms resulting from the
239 comparison of all *SRSF2* mutated samples against SF wildtype samples, the samples did not
240 cluster as well. This stands in agreement with the limited overlap of differentially expressed
241 isoforms found between the two *SRSF2* point mutations examined and suggests at least some

242 heterogeneity among them. The same also held true for *U2AF1* mutated samples, however all
243 *SF3B1* mutated samples still clustered together when compared as a single group to the
244 control.

245 **Differential splicing in SF mutants**

246 Previous studies have reported differential splicing as causal for isoform dysregulation in SF
247 mutants [41,42]. To detect aberrant splicing in our dataset, we quantified the usage of all unique
248 splice junctions by pooling information from all samples (Figure S7). After filtering out junctions
249 with low expression, 235 730 junctions were assigned to genes. Only junctions within an
250 annotated gene were considered, leading to the exclusion of 11 617 (4.9%) junctions. Junctions
251 present in genes with low-expression or genes with a single junction were excluded, leaving
252 221 249 unique junctions (19.3% novel) across 15 526 genes (Table S6.1). Applying the same
253 workflow to the Beat AML cohort yielded 194 158 junctions (8.3% novel). Notably, of the 172
254 518 junctions shared across both datasets, 10 029 (5.8%) were novel. The novel junctions
255 passing our filtering criteria were supported by a high amount of reads and samples with a
256 distribution comparable to that of annotated junctions (Figure 5A). Neither the number of novel
257 junctions nor the number of reads supporting them correlated with the presence of SF
258 mutations, suggesting that novel splicing events are not increased in SF mutants.

259 In consideration of the high proportion of novel junctions in both datasets, we employed a
260 customized pipeline that can quantify the differential splice junction usage (DSJU) of each
261 individual junction, by harnessing usage information from all junctions inside one gene. Of the
262 several hundred junctions reported as differentially used in our primary cohort ($p<0.05$, $\log_2$(fold
263 change) > 1), 20.2-45.9% constituted novel junctions (Tables S7.1-S7.3 and Tables S9.1-S9.2)
264 and were classified according to their relationship with known acceptor and donor sites as
265 described previously (Figure 5B) [15]. Unsurprisingly, validation rates increased with larger

266 mutant sample sizes, ranging from 9.3% (*SF3B1*(K700E); n = 3) to 74.0% (all *SRSF2* mutants;
267 n = 26). Furthermore, validation rates were higher for novel junctions (mean 39.3% vs. 21.5%
268 known junctions), likely due to the stricter initial filtering criteria applied. By performing nanopore
269 sequencing of one *SRSF2*(P95H) mutant and one SF wildtype sample we were able to confirm
270 the usage of several novel junctions and detect resulting novel isoforms as exemplified for
271 *IDH3G* in Figure 6A-D. A tendency towards decreased junction usage was observed for all SF
272 point mutations and was most evident in *SF3B1*(K700E) mutants (1 423/1 927; 73.9% of
273 differentially used junctions). The total number of splicing events, however, was not reduced in
274 SF mutants (mean 9,275,359 events vs. 9 192 697 in wildtype patients), suggesting that
275 decreased splicing is limited to selected junctions rather than being a global effect.
276 We systematically compared the genes with at least one DE isoform and those reported as
277 differentially spliced in all SF mutation subgroups (Tables S9.3-S9.4). For *SRSF2* mutants,
278 genes significant in both analyses included *EWSR1*, *H1F0*, *INTS3* and *YBX1*. In general, out
279 of the genes examined in both analyses only 9.8-23.3% (depending on the SF mutation) of
280 genes reported as having a DE isoform were also reported as being differentially spliced.
281 Conversely, 3.3-28.5% of differentially spliced genes were also reported as having a DE
282 isoform. These findings suggest that differential gene splicing does not always lead to altered
283 isoform expression while at the same time differential isoform expression cannot always be
284 attributed to an explicit splicing alteration. Considering the complementary nature of the
285 analyses, we performed gene ontology (GO) analysis by combining the genes with evidence of
286 differential isoform usage or differential splicing (Tables S10.1-S10.7). Interestingly, GO terms
287 enriched for both *SRSF2* mutants included “mRNA splicing, via spliceosome” ($p<0.001$ and
288 $p=0.046$, respectively) and “mRNA splice site selection” ($p=0.022$ and $p=0.019$, respectively)
289 (Figure 5C).

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

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

299 It is well established that SF mutations dysregulate splicing by changing RNA binding affinities
300 or altering 3' splice site recognition of the corresponding splicing factors. However, SRSF2 also
301 plays a splicing-independent role in transcriptional pausing by translocating the positive
302 transcription elongation factor complex (P-TEFb) from the 7SK complex to RNA polymerase II
303 [43]. A recent study reported that mutant SRSF2(P95H) enhances R-loop formation due to
304 impaired transcriptional pause release, thus providing evidence that altered splicing does not
305 account for the entirety of the *SRSF2*(P95H) mutant phenotype [33]. In another study no
306 difference was found in the total mRNA of 12 Serine/arginine rich splicing factors and 14 out of
307 16 major heterogeneous nuclear ribonucleoprotein (hnRNP) splicing factors between
308 *SRSF2*(P95H) mutant and WT CRISPR clones [44].

309 We cross-referenced our differential expression and differential splicing analysis results with a
310 list of all genes involved in splicing (GO:0000398; mRNA splicing, via spliceosome). Of the 347
311 splicing-related genes (317 of which were expressed in our dataset) 101 were dysregulated in
312 at least one SF mutant group. On average 30.5 (range 6-52) splicing-related genes were
313 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.

326 **Discussion**

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

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

338 Previous studies have demonstrated the predictive value of SF mutations in clonal
339 haematopoiesis of indeterminate potential (CHIP) [47], MDS [6,8,48–50] and AML
340 [10,18,19,51]. However, survival analyses in AML were, in their majority, hampered by small
341 sample sizes and limited availability of further risk factors. Therefore, we examined whether SF
342 mutations impact survival while accounting for recently proposed risk parameters included in
343 the ELN 2017 classification [52]. In our analysis, *SRSF2* and *SF3B1* mutations were no
344 independent prognostic markers for OS in AML. *U2AF1*(S34F) mutations displayed poor OS in
345 the AMLCG cohort, which we were unable to validate in the AMLSG cohort. The discrepancy in
346 survival of SF mutated patients between the two cohorts lied most likely in the large age
347 difference of the participants (median age difference of 8 years), which also led to a higher
348 percentage of patients receiving allogeneic transplants in the AMLSG cohort (56.5% vs. 30.6%
349 in the AMLCG cohort). In summary, SF mutations are early evolutionary events and define

350 prognosis and transformation risk in CHIP and MDS patients, yet there is no clear independent
351 prognostic value of SF mutations in AML.
352 Two large RNA-sequencing studies have been performed previously, to detect aberrantly
353 spliced genes in SF mutants, both of which focused on MDS patients [41,42]. In this study we
354 described a distinct differential isoform expression profile for each SF point mutation.
355 Furthermore, we evaluated differential splicing for the four most common SF point mutations
356 via a customized pipeline to determine differential usage of both known and novel splice
357 junctions. Our pipeline enables the differential quantification of individual splice junctions
358 without restricting the analysis to annotated alternative splicing events. We argue that the
359 strength of our analysis lies in the accurate detection of single dysregulated junctions
360 (especially in cases where splice sites are shared by multiple junctions) in an annotation-
361 independent manner achieving validation rates up to 74.0% in our largest mutant sample group
362 (*SRSF2*, n=19). Limitations of the analysis include the restriction to junctions with both splice
363 sites within the same gene (a restriction shared by most differential splicing algorithms) and
364 genes with at least two junctions. However, the reduced requirements of our analysis could
365 prove valuable in the study of differential splicing in organisms with lacking annotation.
366 All SF point mutations shared a tendency towards decreased splice junction usage, which did
367 not affect the global number of splicing events in SF mutants. Surprisingly, we observed a
368 limited overlap between genes with differentially expressed isoforms and differentially spliced
369 genes. In addition, a recent study by *Liang et al.* reported that the majority of differential binding
370 events in *SRSF2*(P95H) mutants do not translate to alternative splicing [53]. Taken together,
371 these findings indicate a “selection” or possibly a compensation of deregulatory events from
372 differential binding through differential splicing to finally differential isoform expression.
373 Furthermore, the enrichment of aberrant splicing in splicing-related genes opens the possibility

374 of a cascading effect on transcription via the differential alternative splicing of transcriptional
375 components. A congruent hypothesis was stated by *Liang et al.*, where an enrichment of
376 *SRSF2*(P95H) targets in RNA processing and splicing was shown, further supporting the notion
377 of an indirect effect of mutant SRSF2 facilitated through additional splicing components. Further
378 investigations may provide a mechanistic link between the differential splicing of selected genes
379 and the impairment of transcription and specifically transcriptional pausing observed in SF
380 mutant cells, which contributes to the MDS phenotype [33].
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 enabled
383 us to study *SRSF2*(P95H) and *SRSF2*(P95L) separately, thereby not only outlining their
384 differences but also identifying common and likely core targets of differential splicing in *SRSF2*
385 mutants. We conclude that SF mutated patients represent a distinct subgroup of AML patients
386 with poor prognosis that is not attributable solely to the presence of SF mutations. SF mutations
387 induce aberrant splicing throughout the genome including the dysregulation of several genes
388 associated with AML pathogenesis, as well as a number of genes with immediate, functional
389 implications on splicing and transcription. Further studies are required to identify which splicing
390 events are critical in leukaemogenesis and whether they are accessible to new treatments
391 options, such as splicing inhibitors [54] and immunotherapeutic approaches.

392 **Acknowledgments**

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

395

396 **Funding**

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

404

405 **Author Contributions**

406 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.-
407 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. and
408 U.M. provided bioinformatics support. J.P.-M., S.K. and H.B. managed the Genome Analyzer
409 IIx platform and the RNA-sequencing of the AMLCG samples. M.R.-T., H.J., B.K., S.S., N.K.,
410 S.K.B., K.H.M. and K.S. characterized patient samples; M.C.S., D.G., W.B., B.W., J.B. and W.H.
411 coordinated the AMLCG clinical trials. S.A.B. and T.H. wrote the manuscript. All authors
412 approved the final manuscript.

413

414 **Additional Information**

415 The authors declare no conflicts of interest.

416 **References**

417 1. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al.
418 Clinical and biological implications of driver mutations in myelodysplastic syndromes.
419 Blood. 2013;122(22):3616–27.

420 2. Metzeler KH, Herold T, Rothenberg-Thurley M, Amler S, Sauerland MC, Görlich D, et al.
421 Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia.
422 Blood [Internet]. 2016 Aug 4 [cited 2018 May 23];128(5):686–98. Available from:
423 http://www.ncbi.nlm.nih.gov/pubmed/27288520

424 3. Makishima H, Visconte V, Sakaguchi H, Jankowska AM, Kar SA, Jerez A, et al.
425 Mutations in the spliceosome machinery, a novel and ubiquitous pathway in
426 leukemogenesis. Blood [Internet]. 2012;119(14):3203–10. Available from:
427 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3321850/

428 4. Larsson CA, Cote G, Quintás-Cardama A. The changing mutational landscape of acute
429 myeloid leukemia and myelodysplastic syndrome. Mol Cancer Res [Internet]. 2013 Aug
430 [cited 2016 Oct 25];11(8):815–27. Available from:
431 http://www.ncbi.nlm.nih.gov/pubmed/23645565

432 5. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent
433 pathway mutations of splicing machinery in myelodysplasia. Nature [Internet]. 2011 Sep
434 11 [cited 2016 Oct 23];478(7367):64–9. Available from:
435 http://www.ncbi.nlm.nih.gov/pubmed/21909114

436 6. Thol F, Kade S, Schlarmann C, Löffeld P, Morgan M, Krauter J, et al. Frequency and
437 prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with
438 myelodysplastic syndromes. Blood. 2012;119(15):3578–84.

439 7. Dolatshad H, Pellagatti A, Fernandez-Mercado M, Yip BH, Malcovati L, Attwood M, et
440 al. Disruption of SF3B1 results in deregulated expression and splicing of key genes and
441 pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells.
442 Leukemia [Internet]. 2015 May [cited 2018 May 3];29(5):1092–103. Available from:
443 http://www.ncbi.nlm.nih.gov/pubmed/25428262

444 8. Wu S, Kuo Y, Hou H, Li L, Tseng M, Huang C, et al. The clinical implication of SRSF2
445 mutation in patients with myelodysplastic syndrome and its stability during disease
446 evolution. Blood [Internet]. 2014;120(15):3106–12. Available from:
447 https://www.ncbi.nlm.nih.gov/pubmed/22932795

448 9. Graubert TA, Shen D, Ding L, Okeyo-owuor T, Cara L, Shao J, et al. Recurrent
449 Mutations in the U2AF1 Splicing Factor In Myelodysplastic Syndromes. Nat Genet
450 [Internet]. 2012 Dec 11 [cited 2016 Oct 23];44(1):53–7. Available from:
451 http://www.ncbi.nlm.nih.gov/pubmed/22158538

452 10. Hou H-A, Liu C-Y, Kuo Y-Y, Chou W-C, Tsai C-H, Lin C-C, et al. Splicing factor
453 mutations predict poor prognosis in patients with de novo acute myeloid leukemia.
454 Oncotarget [Internet]. 2016 [cited 2017 Jun 19];7(8). Available from:
455 www.impactjournals.com/oncotarget

456 11. Cho Y-U, Jang S, Seo E-J, Park C-J, Chi H-S, Kim D-Y, et al. Preferential occurrence of
457 spliceosome mutations in acute myeloid leukemia with a preceding myelodysplastic
458 syndrome and/or myelodysplasia morphology. Leuk Lymphoma [Internet]. 2014 Aug 3
459 [cited 2016 Oct 25];8194(November):1–25. Available from:
460 http://www.tandfonline.com/doi/full/10.3109/10428194.2014.995648

461 12. Moon H, Cho S, Loh TJ, Jang HN, Liu Y, Choi N, et al. SRSF2 directly inhibits intron

462 splicing to suppresses cassette exon inclusion. BMB Rep [Internet]. 2017 Aug [cited
463 2018 Jan 7];50(8):423–8. Available from:
464 http://www.ncbi.nlm.nih.gov/pubmed/28712387

465 13. Kim E, Ilagan JO, Liang Y, Daubner GM, Lee SCW, Ramakrishnan A, et al. SRSF2
466 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon
467 Recognition. Cancer Cell [Internet]. 2015 May 11 [cited 2016 Oct 25];27(5):617–30.
468 Available from: http://www.ncbi.nlm.nih.gov/pubmed/25965569

469 14. Alsafadi S, Houy A, Battistella A, Popova T, Wassef M, Henry E, et al. Cancer-
470 associated SF3B1 mutations affect alternative splicing by promoting alternative
471 branchpoint usage. Nat Commun [Internet]. 2016 Feb 4 [cited 2016 Oct 19];7:10615.
472 Available from: http://www.nature.com/doifinder/10.1038/ncomms10615

473 15. Okeyo-Owuor T, White BS, Chatrikhi R, Mohan DR, Kim S, Griffith M, et al. U2AF1
474 mutations alter sequence specificity of pre-mRNA binding and splicing. Leukemia
475 [Internet]. 2015 Apr [cited 2018 Jan 28];29(4):909–17. Available from:
476 http://www.ncbi.nlm.nih.gov/pubmed/25311244

477 16. Przychodzen B, Jerez A, Guinta K, Sekeres MA, Padgett R, Maciejewski JP, et al.
478 Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. Blood
479 [Internet]. 2013 Aug 8 [cited 2018 May 23];122(6):999–1006. Available from:
480 http://www.ncbi.nlm.nih.gov/pubmed/23775717

481 17. Shirai CL, Ley JN, White BS, Kim S, Tibbitts J, Shao J, et al. Mutant U2AF1 Expression
482 Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. Cancer Cell [Internet]. 2015 May
483 11 [cited 2018 May 23];27(5):631–43. Available from:
484 http://www.ncbi.nlm.nih.gov/pubmed/25965570

485 18. Yang J, Yao D, Ma J, Yang L, Guo H, Wen X, et al. The prognostic implication of SRSF2
486 mutations in Chinese patients with acute myeloid leukemia. Tumor Biol [Internet]. 2016
487 Aug 28 [cited 2018 May 23];37(8):10107–14. Available from:
488 http://www.ncbi.nlm.nih.gov/pubmed/26820131

489 19. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al.
490 Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med
491 [Internet]. 2016 Jun 9 [cited 2019 Jan 18];374(23):2209–21. Available from:
492 http://www.ncbi.nlm.nih.gov/pubmed/27276561

493 20. Tyner JW, Tognon CE, Bottomly D, Wilmot B, Kurtz SE, Savage SL, et al. Functional
494 genomic landscape of acute myeloid leukaemia. Nature [Internet]. 2018 Oct 17 [cited
495 2019 Sep 8];562(7728):526–31. Available from: http://www.nature.com/articles/s41586-
496 018-0623-z

497 21. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast
498 universal RNA-seq aligner. Bioinformatics [Internet]. 2013 Jan 1 [cited 2017 May
499 8];29(1):15–21. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23104886

500 22. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-
501 aware quantification of transcript expression. Nat Methods [Internet]. 2017 Apr [cited
502 2018 Jan 29];14(4):417–9. Available from:
503 http://www.ncbi.nlm.nih.gov/pubmed/28263959

504 23. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential
505 expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res
506 [Internet]. 2015 Apr 20 [cited 2017 May 8];43(7):e47–e47. Available from:
507 http://www.ncbi.nlm.nih.gov/pubmed/25605792

508 24. Robinson MD, Oshlack A. A scaling normalization method for differential expression
509 analysis of RNA-seq data. Genome Biol [Internet]. 2010 [cited 2017 May 8];11(3):R25.
510 Available from: http://download.springer.com/static/pdf/905/art%253A10.1186%252Fgb-
511 2010-11-3-
512 r25.pdf?originUrl=http%3A%2F%2Fgenomebiology.biomedcentral.com%2Farticle%2F1
513 0.1186%2Fgb-2010-11-3-
514 r25&token2=exp=1494254026~acl=%2Fstatic%2Fpdf%2F905%2Fart%25253A10.1186
515 %25252Fgb

516 25. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential
517 expression analysis of digital gene expression data. Bioinformatics [Internet]. 2009 Jan
518 1 [cited 2018 May 23];26(1):139–40. Available from:
519 http://www.ncbi.nlm.nih.gov/pubmed/19910308

520 26. Liu R, Holik AZ, Su S, Jansz N, Chen K, Leong HS an, et al. Why weight? Modelling
521 sample and observational level variability improves power in RNA-seq analyses.
522 Nucleic Acids Res [Internet]. 2015 [cited 2017 Apr 10];43(15):e97. Available from:
523 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4551905/pdf/gkv412.pdf

524 27. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model
525 analysis tools for RNA-seq read counts. Genome Biol [Internet]. 2014 [cited 2017 Feb
526 12];15(2):R29. Available from:
527 http://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-2-r29

528 28. Leek JT. Svaseq: Removing batch effects and other unwanted noise from sequencing
529 data. Nucleic Acids Res [Internet]. 2014 Dec 1 [cited 2018 Jul 3];42(21):e161. Available
530 from: https://academic.oup.com/nar/article/42/21/e161/2903156

531 29. Tang AD, Soulette CM, Baren MJ van, Hart K, Hrabeta-Robinson E, Wu CJ, et al. Full-
532 length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia
533 reveals downregulation of retained introns. bioRxiv [Internet]. 2018 Sep 6 [cited 2019
534 Dec 9];410183. Available from: https://www.biorxiv.org/content/10.1101/410183v1

535 30. R Core Team. R: A Language and Environment for Statistical Computing [Internet].
536 Vienna, Austria: R Foundation for Statistical Computing; Available from: https://www.r-
537 project.org/

538 31. Herold T, Metzeler KH, Vosberg S, Hartmann L, Ollig C, Olzel FS, et al. Isolated trisomy
539 13 defines a homogeneous AML subgroup with high frequency of mutations in
540 spliceosome genes and poor prognosis. 2014;124(8):1304–11.

541 32. Yoshimi A, Lin K-T, Wiseman DH, Rahman MA, Pastore A, Wang B, et al. Coordinated
542 alterations in RNA splicing and epigenetic regulation drive leukaemogenesis. Nature
543 [Internet]. 2019 Oct 2 [cited 2019 Dec 10];574(7777):273–7. Available from:
544 http://www.nature.com/articles/s41586-019-1618-0

545 33. Chen L, Chen J-Y, Huang Y-J, Gu Y, Qiu J, Qian H, et al. The Augmented R-Loop Is a
546 Unifying Mechanism for Myelodysplastic Syndromes Induced by High-Risk Splicing
547 Factor Mutations. Mol Cell [Internet]. 2018 Feb 1 [cited 2018 Feb 1];69(3):412-425.e6.
548 Available from: http://linkinghub.elsevier.com/retrieve/pii/S1097276517310080

549 34. Roe J-S, Vakoc CR. The Essential Transcriptional Function of BRD4 in Acute Myeloid
550 Leukemia. Cold Spring Harb Symp Quant Biol [Internet]. 2016 [cited 2019 Sep
551 8];81:61–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28174254

552 35. Endo A, Tomizawa D, Aoki Y, Morio T, Mizutani S, Takagi M. EWSR1/ELF5 induces
553 acute myeloid leukemia by inhibiting p53/p21 pathway. Cancer Sci [Internet]. 2016 Dec

554 [cited 2019 Sep 8];107(12):1745–54. Available from:
555 http://www.ncbi.nlm.nih.gov/pubmed/27627705

556 36. Perner F, Jayavelu AK, Schnoeder TM, Mashamba N, Mohr J, Hartmann M, et al. The
557 Cold-Shock Protein Ybx1 Is Required for Development and Maintenance of Acute
558 Myeloid Leukemia (AML) in Vitro and In Vivo. Blood [Internet]. 2017 [cited 2019 Sep
559 8];130(Suppl 1). Available from:
560 http://www.bloodjournal.org/content/130/Suppl_1/792?sso-checked=true

561 37. McNerney ME, Brown CD, Wang X, Bartom ET, Karmakar S, Bandlamudi C, et al.
562 CUX1 is a haploinsufficient tumor suppressor gene on chromosome 7 frequently
563 inactivated in acute myeloid leukemia. Blood [Internet]. 2013 Feb 7 [cited 2019 Sep
564 8];121(6):975–83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23212519

565 38. McGarvey T, Rosonina E, McCracken S, Li Q, Arnaout R, Mientjes E, et al. The acute
566 myeloid leukemia-associated protein, DEK, forms a splicing-dependent interaction with
567 exon-product complexes. J Cell Biol [Internet]. 2000 Jul 24 [cited 2019 Sep
568 8];150(2):309–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10908574

569 39. Fujita S, Honma D, Adachi N, Araki K, Takamatsu E, Katsumoto T, et al. Dual inhibition
570 of EZH1/2 breaks the quiescence of leukemia stem cells in acute myeloid leukemia.
571 Leukemia [Internet]. 2018 Apr 27 [cited 2019 Sep 8];32(4):855–64. Available from:
572 http://www.ncbi.nlm.nih.gov/pubmed/28951561

573 40. Pallarès V, Hoyos M, Chillón MC, Barragán E, Prieto Conde MI, Llop M, et al. Focal
574 Adhesion Genes Refine the Intermediate-Risk Cytogenetic Classification of Acute
575 Myeloid Leukemia. Cancers (Basel) [Internet]. 2018 Nov 13 [cited 2019 Sep 8];10(11).
576 Available from: http://www.ncbi.nlm.nih.gov/pubmed/30428571

577 41. Shiozawa Y, Malcovati L, Gallì A, Sato-Otsubo A, Kataoka K, Sato Y, et al. Aberrant
578 splicing and defective mRNA production induced by somatic spliceosome mutations in
579 myelodysplasia. Nat Commun [Internet]. 2018 [cited 2019 Jan 21];9(1):3649. Available
580 from: http://www.ncbi.nlm.nih.gov/pubmed/30194306

581 42. Pellagatti A, Armstrong RN, Steeples V, Sharma E, Repapi E, Singh S, et al. Impact of
582 spliceosome mutations on RNA splicing in myelodysplasia: Dysregulated
583 genes/pathways and clinical associations. Blood [Internet]. 2018 Sep 20 [cited 2019
584 Feb 6];132(12):1225–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29930011

585 43. Ji X, Zhou Y, Pandit S, Huang J, Li H, Lin CY, et al. SR Proteins Collaborate with 7SK
586 and Promoter-Associated Nascent RNA to Release Paused Polymerase. Cell [Internet].
587 2013 May 9 [cited 2019 Nov 12];153(4):855–68. Available from:
588 http://www.ncbi.nlm.nih.gov/pubmed/23663783

589 44. Zhang J, Lieu YK, Ali AM, Penson A, Reggio KS, Rabadan R, et al. Disease-associated
590 mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. Proc Natl
591 Acad Sci [Internet]. 2015 Aug 25 [cited 2017 May 31];112(34):E4726–34. Available
592 from: http://www.ncbi.nlm.nih.gov/pubmed/26261309

593 45. Skrdlant L, Stark JM, Lin R-J. Myelodysplasia-associated mutations in serine/arginine-
594 rich splicing factor SRSF2 lead to alternative splicing of CDC25C. BMC Mol Biol
595 [Internet]. 2016 [cited 2019 Nov 12];17(1):18. Available from:
596 http://www.ncbi.nlm.nih.gov/pubmed/27552991

597 46. Lemieux B, Blanchette M, Monette A, Mouland AJ, Wellinger RJ, Chabot B. A Function
598 for the hnRNP A1/A2 Proteins in Transcription Elongation. Caputi M, editor. PLoS One
599 [Internet]. 2015 May 26 [cited 2019 Nov 12];10(5):e0126654. Available from:

600 https://dx.plos.org/10.1371/journal.pone.0126654

601 47. Abelson S, Collord G, Ng SWK, Weissbrod O, Mendelson Cohen N, Niemeyer E, et al.
602 Prediction of acute myeloid leukaemia risk in healthy individuals. Nature [Internet]. 2018
603 Jul 9 [cited 2019 Feb 14];559(7714):400–4. Available from:
604 http://www.ncbi.nlm.nih.gov/pubmed/29988082

605 48. Malcovati L, Papaemmanuil E, Bowen DT, Boultwood J, Della Porta MG, Pascutto C, et
606 al. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and
607 myelodysplastic/myeloproliferative neoplasms. Blood [Internet]. 2011 Dec 8 [cited 2019
608 Feb 24];118(24):6239–46. Available from:
609 http://www.ncbi.nlm.nih.gov/pubmed/21998214

610 49. Papaemmanuil E, Cazzola M, Boultwood J, Malcovati L, Vyas P, Bowen D, et al.
611 Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N Engl J Med
612 [Internet]. 2011 Oct 13 [cited 2019 Feb 24];365(15):1384–95. Available from:
613 http://www.ncbi.nlm.nih.gov/pubmed/21995386

614 50. Wu L, Song L, Xu L, Chang C, Xu F, Wu D, et al. Genetic landscape of recurrent
615 ASXL1, U2AF1, SF3B1, SRSF2, and EZH2 mutations in 304 Chinese patients with
616 myelodysplastic syndromes. Tumor Biol [Internet]. 2016 Apr 28 [cited 2019 Feb
617 24];37(4):4633–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26508027

618 51. Zhang S-J, Rampal R, Manshouri T, Patel J, Mensah N, Kayserian A, et al. Genetic
619 analysis of patients with leukemic transformation of myeloproliferative neoplasms
620 shows recurrent SRSF2 mutations that are associated with adverse outcome. Blood
621 [Internet]. 2012 May 10 [cited 2019 Jan 31];119(19):4480–5. Available from:
622 http://www.ncbi.nlm.nih.gov/pubmed/22431577

623 52. Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T, et al.
624 Diagnosis and management of AML in adults: 2017 ELN recommendations from an
625 international expert panel [Internet]. Vol. 129, Blood. 2017 [cited 2019 Feb 14]. p. 424–
626 47. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27895058

627 53. Liang Y, Tebaldi T, Rejeski K, Joshi P, Stefani G, Taylor A, et al. SRSF2 mutations drive
628 oncogenesis by activating a global program of aberrant alternative splicing in
629 hematopoietic cells. Leukemia. 2018;

630 54. Lee SCW, Abdel-Wahab O. Therapeutic targeting of splicing in cancer [Internet]. Vol.
631 22, Nature Medicine. NIH Public Access; 2016 [cited 2019 Jan 31]. p. 976–86. Available
632 from: http://www.ncbi.nlm.nih.gov/pubmed/27603132

633

634

635 **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**

636

637 WBC: white blood cells, LDH: lactate dehydrogenase, BM: bone marrow, ECOG: Eastern Co-

638 operative Oncology Group performance score

639

640

641

642

643 **Figure Legends**

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

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

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

661 **Figure 4: Differential isoform expression analysis in the AMLCG cohort.** (A) Number of
662 differentially expressed isoforms for each SF point mutation. (B) Isoforms reported as
663 differentially expressed in the same direction among patients with different SF mutations vs. SF
664 wildtype patients. HGNC symbols of the genes in which the common isoforms are located are
665 shown. The corresponding Ensembl isoform identifiers are: 1. ENST00000340857, 2.
666 ENST00000481621, 3. ENST00000309668, 4. ENST00000258341, 5. ENST00000426532, 6.

667 ENST00000562151, 7. ENST00000564133, 8. ENST00000368817, 9. ENST00000566524,
668 10. ENST00000530211. (C) Volcano plots showing the magnitude of differential isoform
669 expression for *SRSF2*(P95H) and *SRSF2*(P95L). The x-axis corresponds to the $\log_2$(fold
670 change) of each isoform between mutated and wildtype samples, while the y-axis corresponds
671 to –log10(FDR), where FDR represents the adjusted p-value for each isoform (False Discovery
672 Rate).

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

683 **Figure 6: Splicing dysregulation in *SRSF2* mutants.** (A) DSJU of all splice junctions inside
684 *IDH3G* (ENSG00000067829) are shown for *SRSF2*(P95H) mutants compared to SF wildtype
685 patients in the AMLCG and Beat AML cohort. To determine significance the log fold-change
686 (logFC) of each splice junction (SJ) is compared to the logFC of all other junctions inside the
687 same gene. The x-axis denotes individual splice junctions defined by their chromosomal
688 coordinates (note the high number of splice sites shared by multiple splice junctions). (B)-(D)
689 Nanopore sequencing results of one *SRSF2*(P95H) mutated sample and one SF wildtype
690 sample of the AMLCG cohort. The yellow boxes highlight examples of exon skipping (same

691 exons highlighted in (C) and (D). (B) FLAIR distribution of transcripts. Only one known isoform
692 is expressed in the samples. Additionally, two novel isoforms were detected which are virtually
693 mutually exclusive in the *SRSF2*(P95H) mutant and the SF wildtype sample. (C) Exon
694 composition of known and novel isoforms detected (D). Sashimi plots showing the exon
695 sequence coverage as well as the splice junction usage of the *SRSF2*(P95H) mutated and SF
696 wildtype samples. The black arrow indicates the novel splice junction that is differentially used
697 in *SRSF2*(P95H) mutated samples compared to SF wildtype samples in both RNA-Seq
698 datasets shown in (A). (E) STRING plot depicting an interaction between SRSF2 and several
699 other splicing-related proteins including *TRA2A*. (F) Differential splice junction usage of all
700 splice junctions inside *TRA2A* (ENSG00000164548) are shown for *SRSF2*(P95H) and
701 *SRSF2*(P95L) mutants compared to SF wildtype patients. Annotation same as (A).

702

Figure 1

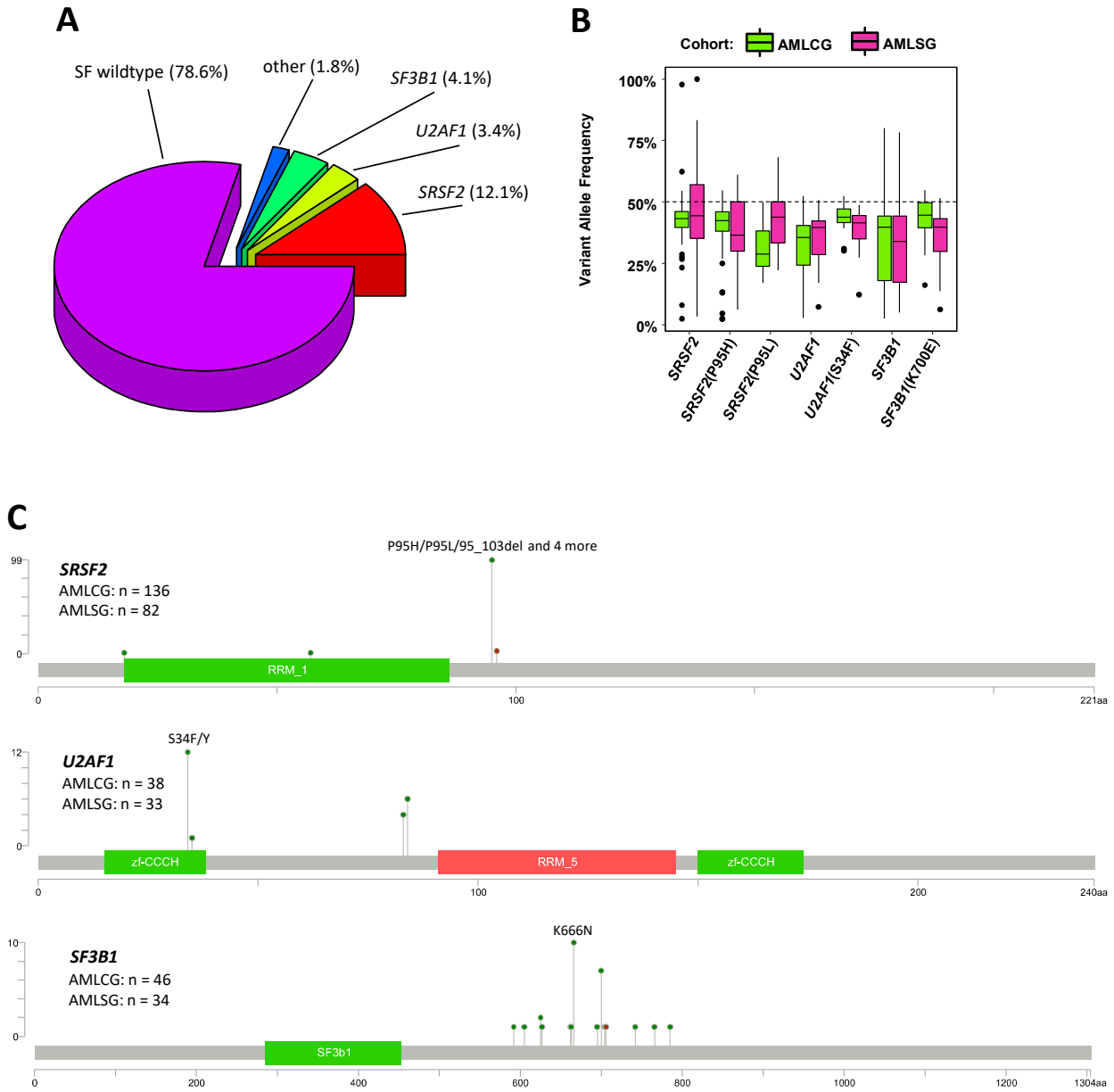

A
SF wildtype (78.6%)
other (1.8%)
SF3B1 (4.1%)
U2AF1 (3.4%)
SRSF2 (12.1%)
B
Cohort: AMLCG AMLSG
Variant Allele Frequency
100%
75%
50%
25%
0%
SRSF2
SRSF2(P95H)
SRSF2(P95L)
U2AF1
U2AF1(S34F)
SF3B1
SF3B1(K700E)
C
SRSF2
AMLCG: n = 136
AMLSG: n = 82
P95H/P95L/95_103del and 4 more
RRM_1
221aa
U2AF1
AMLCG: n = 38
AMLSG: n = 33
S34F/Y
zf-CCCH
RRM_5
zf-CCCH
240aa
SF3B1
AMLCG: n = 46
AMLSG: n = 34
K666N
SF3b1
1304aa

Figure 2

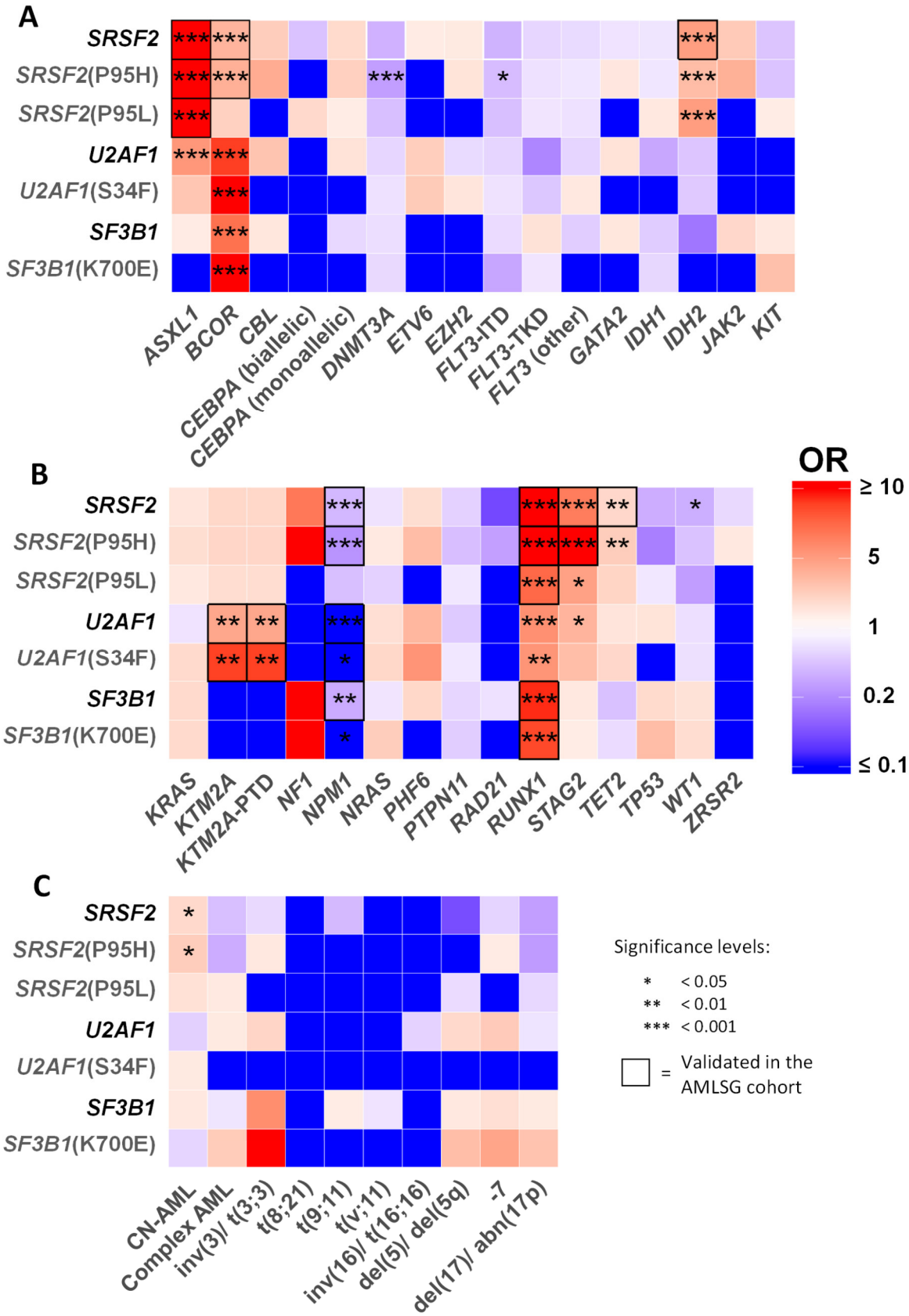

A
SRSF2
SRSF2(P95H)
SRSF2(P95L)
U2AF1
U2AF1(S34F)
SF3B1
SF3B1(K700E)
ASXL1
BCOR
CBL
CEBPA (biallelic)
CEBPA (monoallelic)
DNMT3A
ETV6
EZH2
FLT3-ITD
FLT3-TKD
FLT3 (other)
GATA2
IDH1
IDH2
JAK2
KIT
B
KRAS
KTM2A
KTM2A-PTD
NF1
NPM1
NRAS
PHF6
PTPN11
RAD21
RUNX1
STAG2
TET2
TP53
WT1
ZRSR2
OR
≥ 10
5
1
0.2
≤ 0.1
C
CN-AML
Complex AML
inv(3)/ t(3;3)
t(8;21)
t(9;11)
t(v;11)
inv(16)/ t(16;16)
del(5)/ del(5q)
-7
del(17)/ abn(17p)
Significance levels:
* < 0.05
** < 0.01
*** < 0.001
= Validated in the AMLSG cohort

Figure 3

Overall Survival

Variables	AMLCG cohort: Hazard Ratio (CI 95%)	AMLCG cohort: P-Value	AMLSG cohort: Hazard Ratio (CI 95%)	AMLSG cohort: P-Value
AMLHD98B vs AMLHD98A trial	-	-	0.80 (0.56 - 1.14)	0.212
AMLSG0704 vs AMLHD98A trial	-	-	0.40 (0.31 - 0.52)	<0.001*
AMLCG08 vs AMLCG99 trial	0.69 (0.54 - 0.87)	0.002*	-	-
sAML vs pAML	1.25 (0.96 - 1.63)	0.097	1.81 (1.16 - 2.83)	0.009*
tAML vs pAML	1.56 (1.08 - 2.25)	0.018*	1.66 (1.04 - 2.65)	0.034*
Female sex	0.97 (0.80 - 1.17)	0.738	1.01 (0.80 - 1.27)	0.959
Age (years), per 10 years	1.43 (1.32 - 1.55)	<0.001*	1.68 (1.46 - 1.95)	<0.001*
WBC (10^9/L), per 50 units	1.16 (1.11 - 1.22)	<0.001*	1.35 (1.21 - 1.51)	<0.001*
LDH (U/L), per 500 units	1.07 (1.04 - 1.10)	<0.001*	1.13 (1.03 - 1.25)	0.011*
ASXL1	0.95 (0.71 - 1.27)	0.723	1.15 (0.70 - 1.89)	0.577
BCOR	1.01 (0.73 - 1.39)	0.953	0.86 (0.44 - 1.66)	0.647
CEBPA (biallelic)	0.44 (0.22 - 0.88)	0.019*	0.45 (0.21 - 0.98)	0.045*
JAK2	1.91 (1.00 - 3.68)	0.052	0.69 (0.29 - 1.68)	0.418
KIT	0.67 (0.39 - 1.17)	0.159	0.86 (0.47 - 1.57)	0.624
KRAS	1.61 (1.16 - 2.24)	0.004*	1.09 (0.62 - 1.91)	0.755
NPM1	0.66 (0.51 - 0.85)	0.001*	0.76 (0.56 - 1.05)	0.096
NRAS	0.89 (0.71 - 1.13)	0.334	0.88 (0.64 - 1.20)	0.413
PTPN11	0.89 (0.66 - 1.20)	0.445	0.90 (0.59 - 1.36)	0.608
RUNX1	1.40 (1.07 - 1.84)	0.015*	1.11 (0.77 - 1.59)	0.582
SF3B1 (K700E)	0.98 (0.55 - 1.76)	0.951	0.64 (0.17 - 2.38)	0.506
SF3B1 (other)	1.26 (0.76 - 2.11)	0.370	1.06 (0.52 - 2.17)	0.870
SRSF2 (P95H)	0.91 (0.63 - 1.33)	0.639	1.36 (0.77 - 2.41)	0.294
SRSF2 (P95L)	1.33 (0.83 - 2.13)	0.232	1.26 (0.54 - 2.89)	0.593
SRSF2 (other)	1.02 (0.64 - 1.62)	0.945	0.89 (0.52 - 1.53)	0.677
STAG2	1.08 (0.79 - 1.47)	0.636	1.35 (0.87 - 2.08)	0.176
TET2	0.87 (0.68 - 1.10)	0.230	0.99 (0.70 - 1.39)	0.957
TP53	1.25 (0.81 - 1.95)	0.315	1.95 (1.12 - 3.39)	0.018*
U2AF1 (S34F)	1.90 (1.05 - 3.45)	0.034*	1.39 (0.63 - 3.10)	0.416
U2AF1 (other)	1.11 (0.61 - 2.05)	0.726	0.74 (0.26 - 2.13)	0.577
ZRSR2	1.38 (0.65 - 2.91)	0.404	2.35 (0.92 - 6.03)	0.075
CN-AML	0.84 (0.66 - 1.08)	0.183	0.91 (0.69 - 1.20)	0.506
Complex AML	1.73 (1.02 - 2.93)	0.043*	1.63 (1.00 - 2.68)	0.051
inv(3)/ t(3;3)	2.35 (1.11 - 4.99)	0.026*	2.07 (0.66 - 6.53)	0.214
inv(16)/ t(16;16)	0.32 (0.17 - 0.62)	0.001*	0.31 (0.14 - 0.68)	0.004*
del(5)/ del(5q)	1.14 (0.72 - 1.80)	0.567	1.26 (0.70 - 2.29)	0.445
-7	1.11 (0.73 - 1.68)	0.620	1.77 (1.01 - 3.10)	0.048*
del(17)/ abn(17p)	1.00 (0.60 - 1.64)	0.986	0.79 (0.44 - 1.44)	0.448

Figure 4

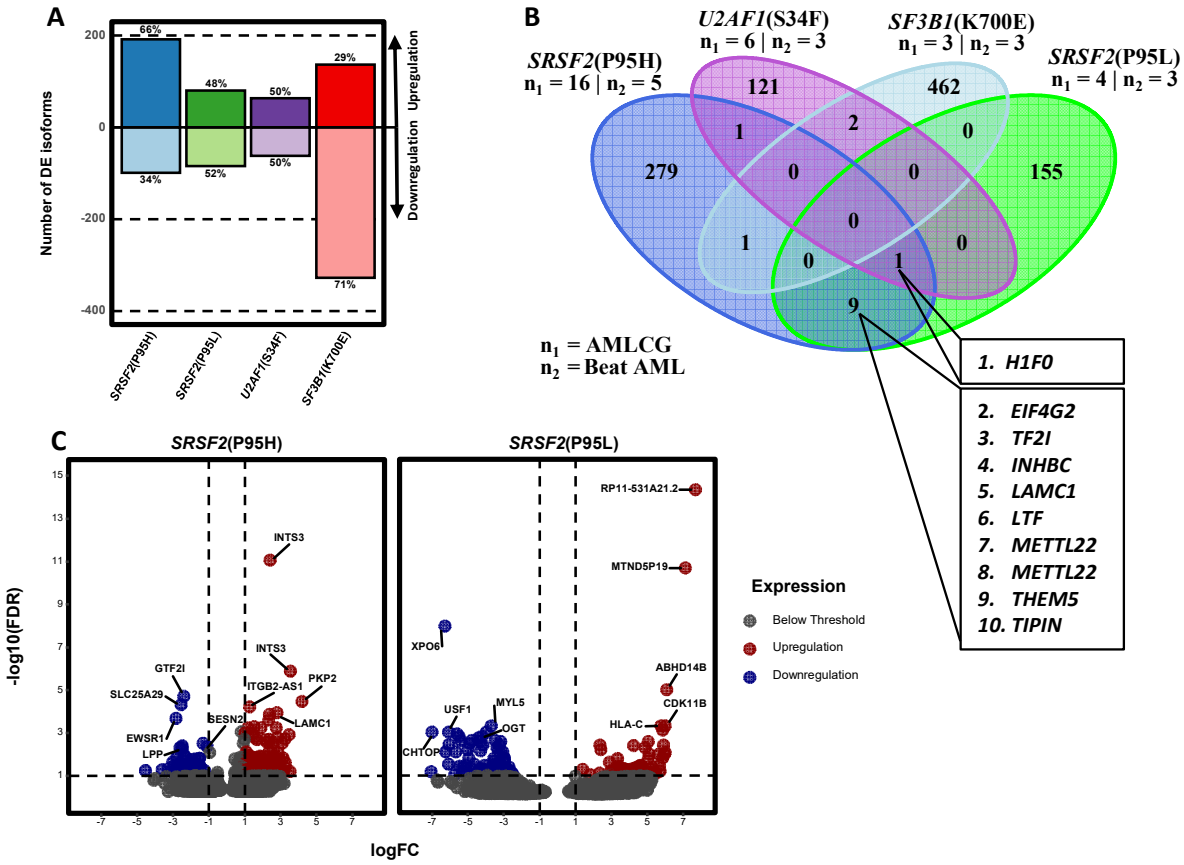
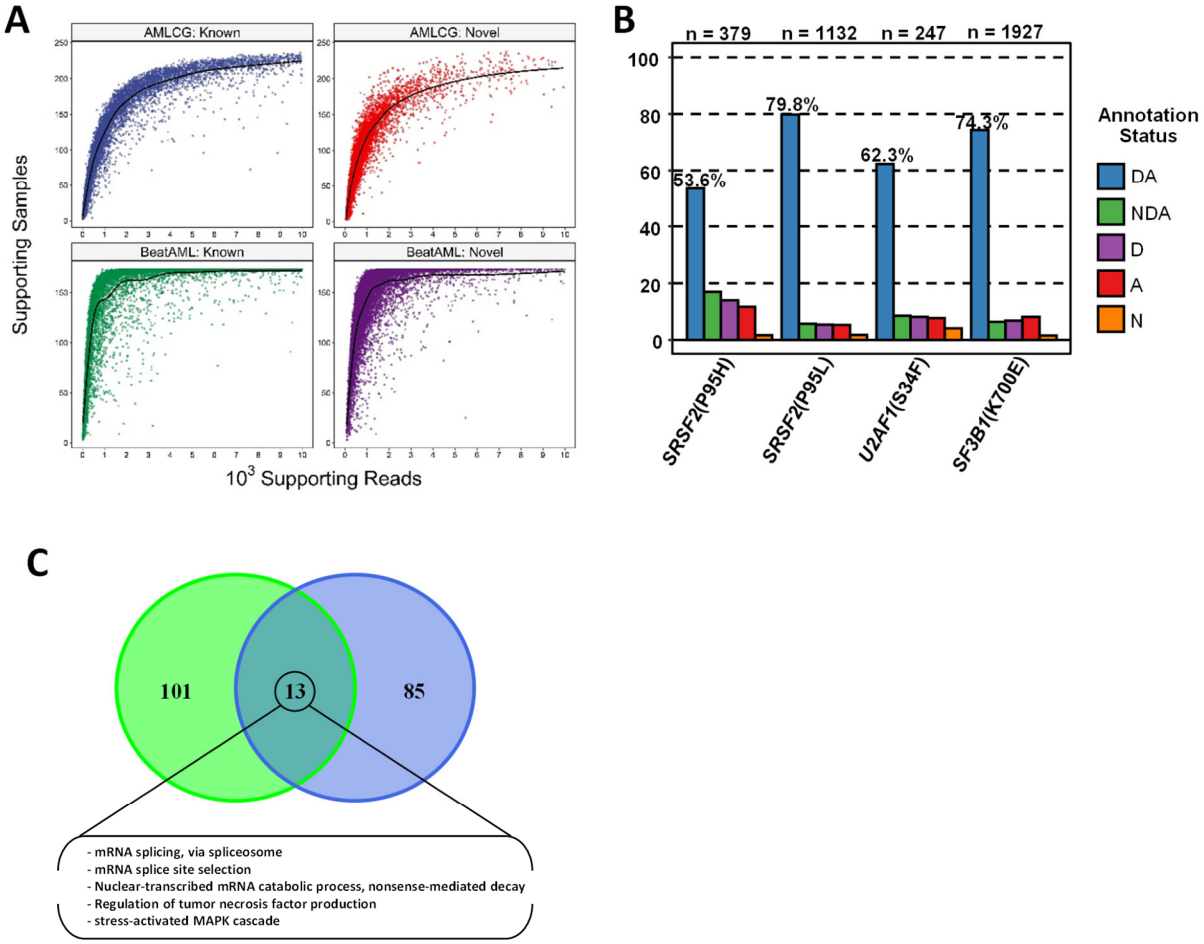

A
Number of DE isoforms
200
0
-200
-400
66%
34%
48%
52%
50%
50%
29%
71%
Upregulation
Downregulation
SRSF2(P95H)
SRSF2(P95L)
U2AF1(S34F)
SF3B1(K700E)
B
SRSF2(P95H)
n1 = 16 | n2 = 5
U2AF1(S34F)
n1 = 6 | n2 = 3
SF3B1(K700E)
n1 = 3 | n2 = 3
SRSF2(P95L)
n1 = 4 | n2 = 3
279
121
462
155
1
2
0
0
0
0
1
0
0
1
9
n1 = AMLCG
n2 = Beat AML
1. H1F0
2. EIF4G2
3. TF2I
4. INHBC
5. LAMC1
6. LTF
7. METTL22
8. METTL22
9. THEM5
10. TIPIN
C
SRSF2(P95H)
SRSF2(P95L)
-log10(FDR)
logFC
INTS3
INTS3
GTF2I
SLC25A29
ITGB2-AS1
PKP2
LAMC1
BESN2
EWSR1
LPP
RP11-531A21.2
MTND5P19
XPO6
ABHD14B
CDK11B
USF1
MYL5
OGT
HLA-C
CHTOP
Expression
Below Threshold
Upregulation
Downregulation

Figure 5


A
AMLCG: Known
AMLCG: Novel
BeatAML: Known
BeatAML: Novel
Supporting Samples
10^3 Supporting Reads
B
n = 379
n = 1132
n = 247
n = 1927
53.6%
79.8%
62.3%
74.3%
SRSF2(P95H)
SRSF2(P95L)
U2AF1(S34F)
SF3B1(K700E)
Annotation Status
DA
NDA
D
A
N
C
101
13
85
- mRNA splicing, via spliceosome
- mRNA splice site selection
- Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay
- Regulation of tumor necrosis factor production
- stress-activated MAPK cascade

928
929
930

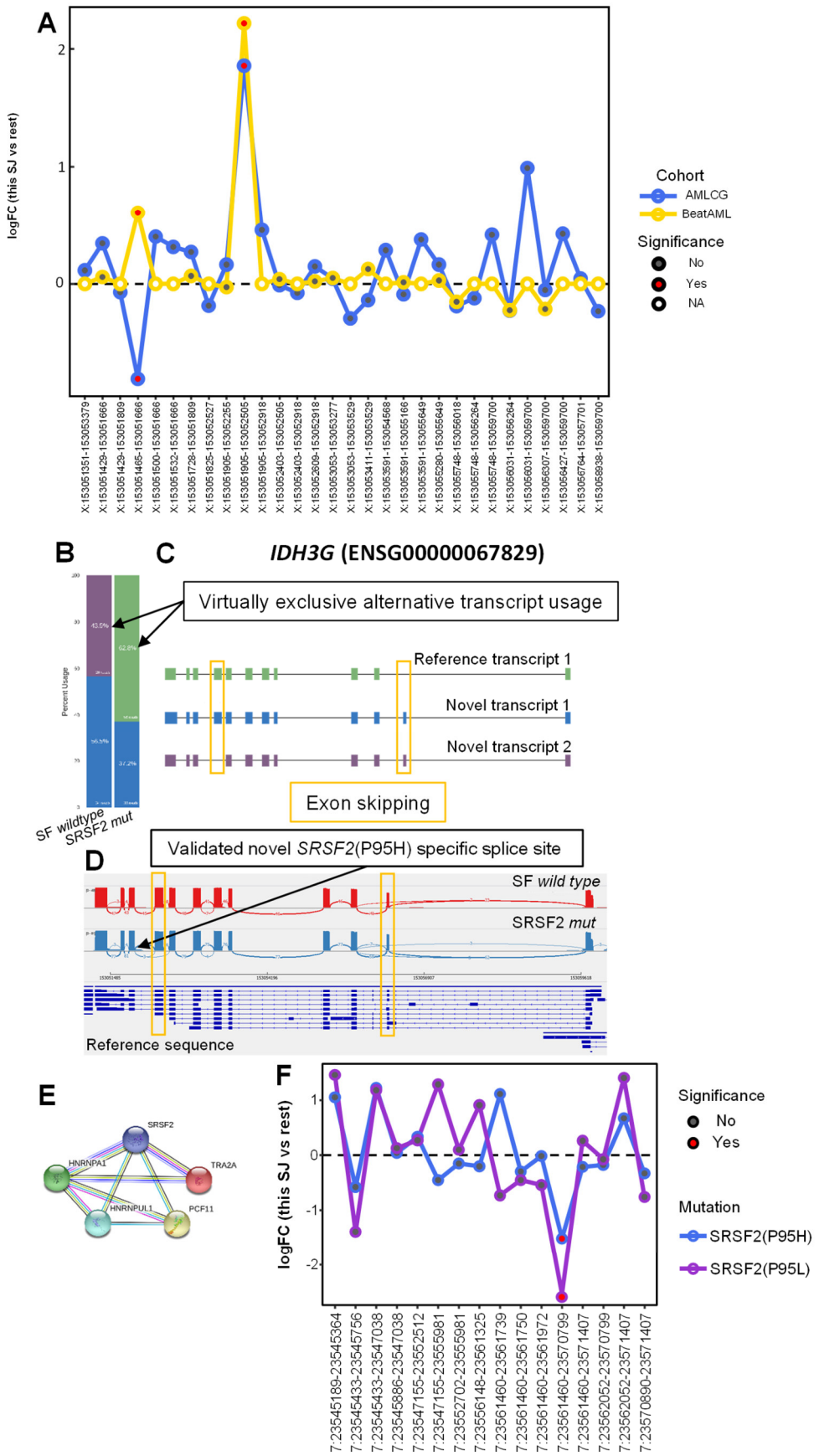



Figure 6