Splicing factor gene mutations in hematologic malignancies

AUTHORS:
Borja Saez¹,², Matthew J. Walter³*, Timothy A. Graubert¹*

AFFILIATIONS:
¹ Massachusetts General Hospital Cancer Center, Boston, MA
² Center for Applied Medical Research, University of Navarra, Spain
³ Washington University School of Medicine, Division of Oncology, St. Louis, MO

RUNNING TITLE: Spliceosome mutations in hematologic malignancies

CORRESPONDING AUTHORS:
*Timothy A. Graubert
Massachusetts General Hospital Cancer Center
149 13th Street
Charlestown, MA, 02129
tgraubert@partners.org
Phone: 617-643-0670
Fax: 617-724-3166

*Matthew J. Walter, MD
Washington University School of Medicine
660 S. Euclid Avenue
Campus Box 8007
St. Louis, MO, 63110
mjwalter@wustl.edu
Phone: 314-362-9409
Fax: 314-362-9333

WORD COUNTS: Abstract: 154; Main Text: 4410
ABSTRACT

Alternative splicing generates a diversity of mRNA transcripts from a single mRNA precursor and contributes to the complexity of our proteome. Splicing is perturbed by a variety of mechanisms in cancer. Recurrent mutations in splicing factors have emerged as a hallmark of several hematologic malignancies. Splicing factor mutations tend to occur in the founding clone of myeloid cancers and these mutations have recently been identified in blood cells from normal healthy elderly individuals with clonal hematopoiesis who are at increased risk of subsequently developing a hematopoietic malignancy, suggesting these mutations contribute to disease initiation. Splicing factor mutations change the pattern of splicing in primary patient and mouse hematopoietic cells and alter hematopoietic differentiation and maturation in animal models. Recent developments in this field are reviewed here, with an emphasis on the clinical consequences of splicing factor mutations, mechanistic insights from animal models, and implications for development of novel therapies targeting the pre-mRNA splicing pathway.
INTRODUCTION

The control of mRNA processing is a crucial component of gene regulation. The precursor messenger RNA (pre-mRNA) sequence contains exons that flank introns, which are removed during splicing.\(^1\) Human genes are remarkably complex, containing an average of 8 exons, with introns comprising up to 90% of the pre-mRNA sequence.\(^2,3\) Up to 94% of human genes generate multiple mRNA isoforms.\(^4-6\) The spliceosome is a multiprotein complex consisting of 5 small nuclear ribonucleoproteins (snRNPs) and more than 150 proteins that recognizes the elements near intron-exon boundaries and the branch site that catalyzes the excision of intronic regions. Additional trans-acting factors including SR proteins bind the cis-regulatory sequences (exonic and intronic silencers and enhancers) to promote or prevent spliceosome assembly.\(^7-9\) The complexity of regulatory elements that control proper splicing makes this process not only critical to maintain tissue homeostasis but also highly susceptible to hereditary and somatic mutations that are involved in disease. This review will discuss the implications of aberrant splicing on human disease, with a focus on the impact of somatic mutations in trans-acting splicing factors in hematologic malignancies.

SPlicing and Disease

A comprehensive analysis of TCGA RNA-seq data from 16 tumor types and matched normal tissue revealed evidence of global alternative splicing involving a variety of splicing junctions, including cassette exons, retrained introns, and competing 5' and 3' splice sites.\(^10\) In particular, AML cells showed the largest number of alternative spliced events among tumor types.\(^10\) Recent evidence from analysis of paired DNA and RNA sequencing data from TCGA has shown that somatic mutations affecting splicing regulatory sequences (splice donor/acceptor sites, last-base exon sites, and exonic splicing enhancer and silencer sites) can alter pre-mRNA splicing in tumor cells (Figure 1A-B).\(^11-13\)

Somatic mutations in the non-coding intronic 5' splice donor (e.g., GT dinucleotide) and 3' splice acceptor (e.g., AG dinucleotide) sites account for \(~1.6\)% of mutations in 12 TCGA tumor types, including AML.\(^12\) These mutations predominantly result in alterations of cassette exon splicing (i.e., exon skipping) and intron retention
(Figure 1C). Jung et al. analyzed paired DNA and RNA-seq results from 6 TCGA tumor types and identified coding somatic single nucleotide variants (SNVs) that affected RNA splicing. Seventy percent of these SNVs occurred on the last base of exons, termed last-based exonic mutations (LBEM). Collectively, 38% of LBEM were associated with abnormal splicing. In contrast, only 4% of the penultimate position relative to the last base of exons, and only 3% of the first base of exons, caused abnormal splicing when mutated. Many of the LBEM were synonymous and often involved tumor suppressor genes, suggesting that they may contribute to tumor pathogenesis (Figure 1C). Supek et al. evaluated 11 TCGA tumor types using paired DNA and RNA-seq data and identified that ~24% of all mutations were synonymous. Synonymous mutations were enriched within 30 base pairs of exon-intron junctions and often resulted in the gain of an exonic splice enhancer (ESE) or loss of an exonic splice silencer (ESS), both resulting in exon inclusion events. Somatic mutations affecting ESEs and ESSs were enriched in oncogenes, suggesting that they too may be important in disease pathogenesis (Figure 1C). In contrast, tumor suppressors such as p53 are also enriched in synonymous mutations. Some of these mutations affect sequences adjacent to splice sites, which can lead to tumor suppressor inactivation. Collectively, somatic mutations in cis-acting regulatory sequences are likely to play a role in tumor pathogenesis by altering specific genes that have functional consequences.

RECURRENT SPlicing MUTATIONS IN HEMATOLOGICAL MALIGNANCIES

Mutation frequencies and clinical associations.
In 2011, next-generation sequencing studies first revealed recurrent, somatic mutations in core components of the spliceosome. The initial studies focused on MDS and CLL, where these mutations are most prevalent. Subsequent studies have clarified the frequency of these mutations across hematologic malignancies and solid tumors, and begun to define their associations with disease subtypes and natural history. This information has been reviewed in detail elsewhere. The major clinical associations relevant to hematologic malignancies are summarized and updated here (Table 1). Mutations in more than 30 factors with established or putative roles in splicing have been reported in hematologic malignancies. The focus here will be on factors with the highest mutation frequency, since much larger datasets will be required to derive consistent and robust clinical associations for rare mutations.
**SF3B1** is mutated in ~5-15% of CLL cases at initial diagnosis, and is associated with shorter time to initial treatment and reduced overall survival.\textsuperscript{16,17} The frequency of mutations is higher in fludarabine-refractory cases.\textsuperscript{23} In the UK LRF CLL4 trial, **SF3B1** mutations were associated with inferior overall survival in both treatment arms, but were not associated with progression-free survival in multivariate analysis.\textsuperscript{24} In myeloid malignancies, the frequency of **SF3B1** mutations is highest in MDS, but recurrent mutations are also present in AML, MPNs, and MDS/MPN overlap syndromes.\textsuperscript{25-27} Within MDS, the mutations are strongly enriched in cases with ring sideroblasts, and are present at lower frequencies in RA, RCMD, and RAEB.\textsuperscript{25,26} The prognostic significance of **SF3B1** mutations in MDS has been controversial, but several studies have shown that they are associated with longer overall survival in multivariate analysis controlling for WHO category.\textsuperscript{26,28,29} The similarity in disease phenotype and clinical behavior for **SF3B1**-mutated cases with ≥15% or <15% ring sideroblasts provided the justification for lowering the threshold in the revised WHO classification to 5% ring sideroblasts for the diagnosis of MDS-RS if an **SF3B1** mutation is present.\textsuperscript{30}

**SRSF2** is mutated in ~5-15% of MDS cases overall, but is highly enriched in CMML.\textsuperscript{15,31,32} CMML-1 and CMML-2 phenotypes are equally represented. **SRSF2** mutations are associated with older age at CMML diagnosis, but have no significant consistent impact on survival; whereas, OS is reduced in lower risk MDS cases with **SRSF2** mutations.\textsuperscript{31,33} In primary myelofibrosis, **SRSF2** is mutated in ~17% of cases and is associated with worse OS, independent of other prognostic features.\textsuperscript{34} Although **SRSF2** mutations are rare in de novo AML, they are more frequent in secondary AML derived from MDS or MPNs.\textsuperscript{35}

**U2AF1** is mutated in 5-10% of MDS cases, across all WHO and IPSS subtypes.\textsuperscript{15} **U2AF1** mutations are generally associated with adverse prognosis in MDS, including reduced overall survival and increased risk of progression to secondary AML reported in some studies.\textsuperscript{14,36}

Other splicing factors are mutated at lower frequencies in hematologic malignancies, so there is less certainty about their associations with disease phenotype and clinical behavior. **ZRSR2** is mutated in ~5% of MDS
patients.\textsuperscript{15,37} These mutations are associated with a higher risk of AML transformation and appear to be an independent unfavorable prognostic factor for overall survival when co-occurring with \textit{TET2} mutations.\textsuperscript{38} Recurrent mutations in \textit{SF3A1}, \textit{U2AF2}, \textit{SF1}, \textit{PRPF40B} have also been reported in hematologic malignancies, but at insufficient frequency to have confidence in their clinical associations.\textsuperscript{15}

Collectively, the frequency of splicing factor mutations is highest in MDS (>50\% of cases), followed by MDS/MPN and CMML (~50-75\%), primary myelofibrosis (~50\%), secondary AML derived from MDS or MPN (~10-55\%), MPN (~10-30\%), and de novo AML (~10\%).\textsuperscript{15,39-43} The relative enrichment of splicing factor mutations in antecedent hematologic diseases with a propensity to evolve to AML (i.e., MDS and MPN) and their paucity in \textit{de novo} AML provided an opportunity to develop a genetic classifier than can objectively distinguish secondary AML arising from an known antecedent MDS from \textit{de novo} AML, agnostic of clinical history.\textsuperscript{43} In this study, 8 genes (4 of which are splicing factors) could differentiate secondary AML from \textit{de novo} AML with >95\% specificity.\textsuperscript{43} The authors demonstrate that patients classified as \textit{de novo} AML that harbor secondary-type mutations share many clinical and pathological features with secondary AML, suggesting that they may have had an unrecognized period of antecedent MDS before AML diagnosis. Furthermore, myeloid malignancies with 5-20\% blasts and >20-30\% blasts (defined as MDS-EB and AML, respectively, in the WHO classification) have a similar frequency of splicing factor mutations and can be further stratified using gene expression and DNA methylation profiling.\textsuperscript{44} These studies provide additional evidence that splicing factor mutations define myeloid malignancies with a consistent phenotype and clinical behavior that may not be apparent using conventional clinico-pathologic information.

**Characteristics of splicing factor mutations.** Several principles emerged from the initial descriptions of splicing factor mutations and now have considerable support in the literature. First, the mutations tend to be acquired early in the disease course in MDS (i.e., in the “founding” clone), and less often as subclonal mutations acquired at progression or relapse.\textsuperscript{45,46} In fact, splicing factor mutations have been found in the peripheral blood of aging individuals with clonally-skewed hematopoiesis who do not have MDS or AML, implying that these mutations are sufficient to confer a selective growth advantage.\textsuperscript{47} Conversely, splicing
factor mutations in CLL (principally, in \textit{SF3B1}) are more often subclonal.\textsuperscript{48} Second, the mutations are mutually exclusive. That is, more than one splicing factor mutation is rarely seen in a single tumor (and rare exceptions may represent independent subclones harboring a single splicing factor mutation).\textsuperscript{15,45,49} Third, mutations in the three most frequently mutated genes (\textit{SF3B1, SRSF2, U2AF1}) are heterozygous and are enriched at ‘hotspot’ codons.\textsuperscript{14,15,25} Fourth, the spectrum of mutant alleles for these genes is consistent across different hematologic malignancy histologies, although there are notable exceptions in solid tumors (for example, distinct \textit{SF3B1} alleles in uveal melanoma).\textsuperscript{50,51} Finally, mutations in \textit{SF3B1} and \textit{U2AF1} affect 3’ splice site usage and \textit{SRSF2} mutations affect exon recognition. Why the 5’ complex is largely spared, although it is also comprised of a large number of ubiquitously expressed proteins, is not understood. The main lessons will be reviewed and updated here, as this information has also been the subject of other recent reviews.\textsuperscript{18,19,21,22}

\textit{SF3B1} mutations in all hematologic malignancies are clustered in exons 14-16, which encode the 4\textsuperscript{th}-6\textsuperscript{th} HEAT domains, with a hotspot (~50\% of mutations) at codon 700 (K700E).\textsuperscript{25} In CLL, \textit{SF3B1}-mutated cases are less likely to co-occur with trisomy 12 and are largely mutually exclusive with mutations in \textit{NOTCH1}.\textsuperscript{23,24} In MDS, \textit{SF3B1} mutations frequently co-occur with mutations in factors regulating DNA methylation (e.g., \textit{DNMT3A, TET2}).\textsuperscript{29}

\textit{SRSF2} mutations invariably affect codon 95, most often as missense mutations (P95H>P95L>P95R>P95A/T) and less commonly as insertion/deletions at this location.\textsuperscript{15,31} In MDS, \textit{SRSF2} mutations co-occur with mutant \textit{IDH1/2, TET2, RUNX1, ASXL1} and \textit{STAG2}.\textsuperscript{33,37,38,45} In CMML, \textit{SRSF2} mutations are largely mutually exclusive with \textit{EZH2} mutations, but frequently co-occur with mutations in \textit{TET2}.\textsuperscript{31} In myelofibrosis, \textit{SRSF2} mutations frequently co-occur with mutant \textit{IDH1/2}.\textsuperscript{34}

\textit{U2AF1} mutations are almost exclusively missense substitutions at two codons, codon 34 (S34F>S34Y) and 157 (Q157P/R/G), and rare biallelic mutations in \textit{cis}.\textsuperscript{14,15} \textit{U2AF1}-mutated cases are enriched for co-occurring \textit{ASXL1} mutations and del(20q).\textsuperscript{38,45,49}
Less frequently mutated splicing factors have distinct mutation spectra. \textbf{ZRSR2} mutations are distributed across the coding region and include missense, nonsense, frameshift, and splice site mutations, consistent with a putative loss-of-function role for this X-linked factor.\textsuperscript{15}

\textbf{Functional consequences of splicing factor mutations.} To investigate the functional consequences of splicing factor mutations, a number of studies have been performed using isogenic human cell lines and genetically engineered mouse models (summarized in Table 2).

Despite the predicted oncogenic role of U2AF1 mutations, Yoshida and colleagues first demonstrated that they have growth suppressive properties.\textsuperscript{15} Expression of the mutant \textit{U2AF1} (S34F) allele in HeLa cells led to growth delay and increased apoptosis. Furthermore, expression of \textit{U2AF1} mutations in highly purified hematopoietic stem and progenitor cells (HSPCs) reduced repopulation capacity in transplantation assays compared to WT-U2AF1 transduced cells.\textsuperscript{15} In a first generation mouse model, a single copy of a human \textit{U2AF1} cDNA (wild-type or S34F) was inserted in the collagen 1a1 locus downstream of a tetracycline-inducible promoter. This system allows for the temporal expression of the S34F mutation while maintaining its ratio similar to the endogenous U2af1 protein levels. Expression of the S34F mutation in the hematopoietic compartment led to a persistent reduction of the total white blood cell count (WBC) without changes in the red blood cell (RBC) parameters and platelet counts in the peripheral blood.\textsuperscript{52} Concomitantly, mutant mice showed a reduction in the numbers of monocytes and B-lymphocytes and an accumulation of myeloid progenitors in the bone marrow. Nevertheless, this accumulation of HSPCs did not correlate with a functional advantage since competitive transplantation assays demonstrated a repopulation disadvantage of mutant cells in primary and secondary recipients. Furthermore, expression of the S34F mutation for up to a year was not sufficient to induce transformation to AML.\textsuperscript{52}

A floxed SRSF2-P95H allele inserted in the endogenous Srsf2 locus in mice provided a conditional model to study the functional impact of this mutation on hematopoiesis.\textsuperscript{53} SRSF2-P95H mice developed a myelodysplastic phenotype characterized by leukopenia, macrocytic anemia, myeloid and erythroid dysplasia.
as well as accumulation of stem and progenitor cells in the bone marrow. Similar to the mutant U2AF1 mouse, SRSF2-P95H HSCs show limited repopulation ability in transplantation assays compared to their wild-type counterpart. A conditional murine knock-in of the SF3B1-K700E mutation showed similar features to those observed in humans. Heterozygous expression of the mutation led to macrocytic anemia associated with elevated levels of erythropoietin, the block in erythroid differentiation and no differences in the relative abundance of multipotent progenitor cells (while HSC quantification differs between both studied, most likely due to the use of alternative immunophenotypic definition for HSCs). However, these animals did not exhibit ring sideroblasts, the defining characteristic of SF3B1 mutant cells in human MDS. In competitive transplantation settings mutant cells showed an engraftment disadvantage compared to wild-type stem cells. Limiting SF3B1-K700E expression to the B-cell compartment led to global splicing changes, defects in B lymphopoiesis and cell senescence but did not result in transformation to overt CLL (at least with the reported follow-up times).

**Molecular consequences of splicing factor mutations.** Most splicing factors mutations found in hematological malignancies affect proteins implicated in the early exon recognition steps of pre-mRNA processing. These mutations universally lead to splicing defects affecting a specific subset of genes rather than globally affecting multi-exonic genes.

**SF3B1** mutations cause alternative 3’ splice site (SS) usage due to the increased recognition of a cryptic SS between the branch point and the canonical 3’ SS *(Figure 2A).* How this 3’ alternative SS is selected is being actively studied. Recent findings have suggested that the cryptic 3’ SS is associated with a shorter pyrimidine tract and enrichment in adenines upstream of the canonical 3’ SS leading to the use of different branch sites. In addition, mutant SF3B1 may potentially induce a conformational change that would facilitate the use of the cryptic 3’ SS by modifying the ability to select splice sites immediately downstream of the branch site, normally protected by steric hindrance upon binding of the U2 snRNP to the branch site. Analysis of a crystal
structure for the SF3b core complex suggests that the hotspot mutations may induce conformational changes that alter interactions with pre-mRNA and other splicing factors.\textsuperscript{62}

\textbf{U2AF1} gene mutations also alter 3' splice site recognition of a subset of AG-dependent 3' splice sites leading predominantly to exon skipping and alternative 3' acceptor site use (\textbf{Figure 2B}).\textsuperscript{57,58,63,64} These alterations are allele-specific, since S34 and Q157 mutations show alternative sequence specificity at position -3 and +1 in the 3' splice site, respectively.\textsuperscript{58} U2AF1 recognizes a pyrimidine-AG-purine motif. As a consequence of the S34F mutation, U2AF1 losses affinity for the motifs when a T (U) is present in position -3 leading to an increase in exon skipping and decreased recognition of the canonical junction resulting in the use of an alternative splice site.\textsuperscript{57} Conversely, mutations at the Q157 residue promote the recognition of 3' SS when a G is present in position +1 and repress those bearing an A at this position.\textsuperscript{58} Changes in RNA binding affinity for mutant U2AF1 can, in general, account for observed changes in splicing preference.\textsuperscript{65} Furthermore, these sequence specific effects are distinct from those observed in U2AF1 deficient cells, reinforcing the concept that they do not behave as loss-of-function but rather have neomorphic activity.\textsuperscript{58} In addition, one recent study demonstrated that ectopic expression of S34F U2AF1 in Ba/F3 cells resulted in increased use of distal mRNA cleavage and polyadenylation sites.\textsuperscript{66} One of the affected transcripts was \textit{Atg7}, which led to reduced protein levels and a functional autophagy defect.

Deep sequencing analysis of U2AF1 mutant patient samples, transgenic mice and isogenic cell lines have demonstrated the presence of global changes that affect 5-10% of the transcriptome.\textsuperscript{52,57} Some of these recurrent alternatively spliced genes are implicated in essential cellular processes such as RNA processing and splicing, protein translation, DNA methylation, DNA damage response and apoptosis, and can also occur in genes that are recurrently mutated in MDS.\textsuperscript{52,57,58,63,64,67} However, the functional contribution to the disease phenotype of alternatively spliced genes remains an active area of research.

\textbf{SRSF2} binds to ESEs to promote exon recognition in a sequence specific manner. SRSF2 binds a consensus ESE motif (SSNG, where S represents a C or G) through its RNA recognition motif (RRM) domain.\textsuperscript{58-70} The
P95 mutation affects SRSF2 normal sequence-specific RNA binding activity. Mutant SRSF2 recognizes with higher affinity the CCNG motif versus GGNG, whereas the wild-type protein recognizes both with similar affinity (Figure 2C). As a result, mutant SRSF2 promotes or represses the inclusion of exons containing C or G rich motifs causing splicing changes in hundreds of genes. Some of these alternative splicing events affect genes commonly mutated in MDS like EZH2 and BCOR, suggesting that SRSF2 may contribute to MDS pathogenesis by, at least in part, controlling the expression of these genes. Similar to U2AF1 mutations, P95 mutant SRSF2 induces distinct splicing changes compared to those induced by the depletion of SRSF2, again suggesting that SRSF2 mutations are not loss-of-function but rather have neomorphic activity.

Less is known about the molecular consequences of ZRSR2 mutations. Due to its critical role in the minor spliceosome, ZRSR2 mutations are characterized by the accumulation of alternatively spliced U12-type introns (Figure 2D). These defects are mainly retention of U12-type introns. Furthermore, splicing defects due to ZRS2 mutations are enriched for pathways including MAPK, ErbB signaling, and focal adhesion kinases among others.

While the splicing junctions that are affected by these mutations are typically different, it is not known whether a shared set of downstream genes are affected. Although the discovery of common target genes has been limited by differences in the cell types that were sequenced and analytic approaches used in current studies, this remains an active area of research.

**Therapeutic Opportunities**

The strong selective pressure favoring acquisition of mutations in splicing factors suggests not only that these alterations are important for the biology of hematologic malignancies, but also that they may create vulnerabilities that could be exploited therapeutically. Preclinical work in this area has focused on three potential entry points (Figure 3): 1) perturbation of splicing, 2) modifying signaling pathways upstream of splicing, and 3) correcting downstream aberrant splicing. Recent developments and initial steps towards clinical translation will be reviewed here. Readers are referred to other reviews and primary sources for
additional details on the medicinal chemistry, pharmacology, and biochemical properties of these compounds.74-76

**Splicing modulators.** A variety of natural plant and bacterial products have been isolated that have anti-tumor and splicing modulator properties (Figure 3A). Several of these molecules have been modified to improve potency and drug-like properties, either by generating derivatives of the parent compounds or by *de novo* synthesis of structural analogs. FR901464 (FR) was isolated from *Pseudomonas* sp. Its methyl ketal derivative, Spliceostatin A (SSA), physically interacts with components of the SF3b subcomplex and induces accumulation of unspliced products.77 Further study indicated that SSA alters splicing fidelity, rather than acting as a general splicing inhibitor, resulting in accumulation of alternatively spliced products containing premature termination codons, leading to their depletion by nonsense-mediated decay.78 In CLL cells, SSA potently depletes MCL-1 via alternative splicing, leading to induction of apoptosis at low nanomolar concentrations *in vitro*, with relative sparing of normal T and B cells.79

Pladienolide B (PB), FD-895, and herboxidiene are natural products of *Streptomyces* sp. that share a common pharmacophore with FR. The elaboration of structurally-related splicing modulators by genetically diverse prokaryote species is a fascinating and unexplained example of convergent evolution. An elegant genetic study demonstrated unequivocally that SF3B1 is a direct target of PB.80 E7107 is a urethane derivative of PB that interferes with spliceosome assembly.81 In preclinical studies, E7107 reduced the leukemic burden in the *MLL/AF9* retroviral transduction/transplantation model on a *Srsf2*P95H/+ background and impaired the growth of primary human AML samples with splicing factor mutations in a xenograft model.82 Similarly, in a competitive repopulation model in mice, *Sf3b1*K700E/+ cells were selectively depleted *in vivo* by E7107.54

Parallel phase I studies of E7107 with slightly different administration schedules were initiated in Europe and the US for patients with advanced non-hematologic malignancies.83,84 Although the adverse effects were generally mild to moderate and reversible in 66 enrolled patients, both studies were terminated when 2 patients on the US study (NCT00499499) experienced impaired visual acuity.84 One patient on the European study...
(NCT00459823) developed visual disturbance at the time of treatment discontinuation, felt to be consistent with optic neuritis that partially improved with corticosteroid treatment. A phase I study of H3B-8800, an orally-bioavailable splicing modulator, recently opened to accrual for patients with myeloid malignancies (NCT02841540). Whether the ocular toxicity in the E7107 studies is unique and attributable to this compound, the pharmacokinetics at these schedules, a class effect of PB derivatives, sensitivity of optic neuronal cells to splicing modulators in general, or other factors is unknown, but certainly merits close attention during clinical investigation with this class of compounds.

The sudemycins are synthetic analogs of PB and FR that share a common pharmacophore, but have fewer stereocenters and have improved solubility, chemical stability, and plasma half-life. Primary human CLL cells undergo apoptosis in a dose-dependent fashion after exposure to Sudemycin D1 (SD1) in vitro. $SF3B1$ mutant samples were more sensitive in vitro than CLL cells lacking splicing factor mutations. SD1 and the BTK inhibitor, Ibrutinib, act synergistically to induce apoptosis in CLL cells, possibly due to inactivation of the endogenous BTK inhibitor, IBTK, by SD1 via alternative splicing. Primary human MDS/AML cells with mutant $U2AF1$ had impaired proliferation after treatment with SD1 in vitro compared to cells lacking splicing factor mutations and in vivo treatment of mice with SD6 reverted aberrant hematopoietic progenitor expansion in an inducible mouse model of mutant U2AF1.

FD-895 caused induction of apoptosis in CLL cells in vitro (independent of SF3B1 genotype) at nanomolar concentrations, similar to PB. A semisynthetic analog (17S-FD-895) with improved pharmacokinetic properties administered to immunodeficient mice engrafted with primary human AML samples reduced their repopulating activity in secondary transplants and reverted the splicing signature in LSCs from “secondary AML-associated” to “healthy aging”.

The genetic data from patients provided a priori support for the hypothesis that splicing factor mutant hematologic malignancies might be uniquely sensitive to splicing modulators. As discussed above, the wild-type allele is retained and expressed in cells with mutations in the 3 most frequently mutated splicing factors,
suggesting that some level of normal splicing is required for viability. Furthermore, the mutually exclusive nature of these mutations suggests that a cell with a splicing factor mutation cannot tolerate further perturbation of splicing induced by a second mutation in this pathway. Further support comes from the Srsf2<sup>P95H</sup>/+ mouse model, where deletion of the wild-type allele caused rapid loss of hemizygous mutant cells.<sup>82</sup> These preclinical studies provide support for the initial hypothesis and lead to the following predictions that will be tested in planned and ongoing clinical trials: 1) patients with splicing factor mutant malignancies may derive the largest clinical benefit from splicing modulators, and 2) a general splicing modulator may have activity across a broad range of splicing factor mutant genotypes. In fact, sensitivity to splicing modulators may extend to malignancies that lack splicing factor mutations. Myc-driven breast cancers show an exquisite sensitivity to pharmacologic splicing modulation that has been attributed to the increased burden on the spliceosome to process Myc-induced pre-mRNA production.<sup>93</sup> In addition, there is emerging evidence that some B-cell lymphomas (mantle cell and Myc-driven lymphomas) are sensitive to pharmacologic modulation of splicing.<sup>94,95</sup> Future studies will be needed to explore the potential use of splicing modulator therapy in these diseases.

**Targeting signaling inputs to splicing.** A family of SR factors influence splice site selection by the spliceosome (Figure 3B). The intracellular trafficking and protein-protein interactions of SR factors are regulated by the phosphorylation status of their arginine-serine (RS) domains. Hence, the upstream serine kinases (e.g., SRPK1, SRPK2, CLK1) are potential targets for small molecules that could modulate splicing. A number of compounds with these properties have been identified, including chlorhexidine, TG003, diospyrin, and indolocarbazole.<sup>96-99</sup>

**Reverting aberrant splicing.** Considerable progress has been made towards pharmaceutical approaches for diseases where a single aberrant splice isoform is pathogenic. These include antisense oligonucleotides that target splice sites or regulatory sequences and trans-splicing molecules that mediate exchange of a target sequence of an endogenous sequence containing a deleterious variant (Figure 3C) (reviewed in: Daguenet, et al<sup>100</sup>). These approaches may have value for splicing factor mutant hematologic malignancies where one or a
few aberrant splice isoforms prove to be pathogenic. Correction of a larger number of alternative splicing events will be challenging, and may require the development of novel strategies and technologies (Figure 3D).

The tight coordination of splicing with other cellular processes (e.g., transcription initiation and elongation, histone modification and DNA methylation, mRNA export and translation, DNA damage response)\textsuperscript{101-109} suggests that splicing factor mutations may create vulnerabilities in other pathways. Unbiased small molecule and genetic screens may uncover additional potentially druggable targets, some of which could be candidates for drug development.

**CONCLUSION**

Rapid progress has been made to understand spliceosome gene mutations since their first description and their clinical associations in 2011. This was followed by studies focusing on the impact of mutations on splicing and then the development of genetically-engineered mouse models to characterize their contribution to cellular and molecular phenotypes. Preclinical work on splicing modulators in cellular and mouse models, leading to a first in human study of an oral agent for patients with myeloid malignancies in 2016, just 5 years after discovery of the mutations, is evidence that the field is moving rapidly. This accelerated pace builds on preceding decades of work in the basic biology of splicing and the isolation and modification of naturally occurring splice modulator products. Finally, a highly collaborative research environment has been instrumental in moving this field forward with the goal of translating this research for the benefit of patients. Current and future research directions should clarify the mechanistic consequences of splicing factor mutations, define the role of cooperating mutations in mediating transformation to overt malignancy, and uncover vulnerabilities that can be exploited for novel therapeutic strategies.

**Acknowledgment**

Support was provided by a NIH/NCI SPORE in Leukemia (P50CA171963 to TAG and MJW), the Edward P. Evans Foundation (MJW, TAG), the Lottie Caroline Hardy Trust (TAG, MJW), and a Leukemia and Lymphoma Society Scholar Award (MJW).
Author Contributions

The manuscript was written and edited by: BS, MJW, and TAG.

All co-authors reviewed and approved the submission.

Conflicts of Interest

The authors have no conflicts of interest.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Disease</th>
<th>Frequency (%)</th>
<th>Co-occurring Mutations</th>
<th>Clinical Association(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SF3B1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLL</td>
<td></td>
<td>5-17</td>
<td>del(11q)</td>
<td>Reduced PFS (clonal only) &lt;sup&gt;110&lt;/sup&gt;, Reduced OS &lt;sup&gt;24&lt;/sup&gt;, Reduced PFS/OS &lt;sup&gt;17,23&lt;/sup&gt;</td>
</tr>
<tr>
<td>MDS</td>
<td>del(11q)</td>
<td>7-28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RARS/RCMD-RS</td>
<td></td>
<td>54-83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RARS-T</td>
<td></td>
<td>81-88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RA</td>
<td></td>
<td>5-12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAEB</td>
<td></td>
<td>5-11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCMD</td>
<td></td>
<td>4-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>secondary AML</td>
<td></td>
<td>5-11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>de novo AML</td>
<td></td>
<td>3-5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tAML</td>
<td></td>
<td>35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMML</td>
<td></td>
<td>5-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDS/MPN</td>
<td></td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMF</td>
<td></td>
<td>4-8</td>
<td></td>
<td>No impact on OS &lt;sup&gt;27&lt;/sup&gt;</td>
</tr>
<tr>
<td>ET</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>secondary AML</td>
<td></td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SRSF2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDS</td>
<td>RUNX1, IDH2, IDH1, TET2, STAG2</td>
<td>1-12</td>
<td></td>
<td>Reduced PFS &lt;sup&gt;112&lt;/sup&gt;, Reduced LFS &lt;sup&gt;45&lt;/sup&gt;, Reduced OS (lower risk IPSS) &lt;sup&gt;33&lt;/sup&gt;, Reduced OS (multivariate) &lt;sup&gt;13&lt;/sup&gt;, Reduced LFS/OS &lt;sup&gt;37&lt;/sup&gt;</td>
</tr>
<tr>
<td>secondary AML</td>
<td></td>
<td>5-20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMML</td>
<td>TET2</td>
<td>30-47</td>
<td></td>
<td>No impact on OS &lt;sup&gt;31,34&lt;/sup&gt;, Reduced LFS/OS &lt;sup&gt;39&lt;/sup&gt;</td>
</tr>
<tr>
<td>MPN</td>
<td>IDH1/2</td>
<td>19-34</td>
<td></td>
<td>Reduced LFS/OS &lt;sup&gt;34&lt;/sup&gt;</td>
</tr>
<tr>
<td>secondary AML</td>
<td></td>
<td>3-17</td>
<td></td>
<td>Reduced OS &lt;sup&gt;35&lt;/sup&gt;</td>
</tr>
<tr>
<td>de novo AML</td>
<td></td>
<td>5-6</td>
<td></td>
<td>Reduced OS (univariate) &lt;sup&gt;31&lt;/sup&gt;</td>
</tr>
<tr>
<td>tAML</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>U2AF1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDS</td>
<td>del(20q), ASXL1</td>
<td>5-11</td>
<td></td>
<td>Reduced PFS &lt;sup&gt;112&lt;/sup&gt;, Reduced LFS &lt;sup&gt;45&lt;/sup&gt;, Reduced OS (univariate) &lt;sup&gt;113&lt;/sup&gt;, Inferior OS, Inferior LFS (lower risk) &lt;sup&gt;36&lt;/sup&gt;</td>
</tr>
<tr>
<td>secondary AML</td>
<td></td>
<td>9-16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDS/MPN</td>
<td>DNMT3A</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMML</td>
<td>5-9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPN</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMF</td>
<td>del(20q), ASXL1</td>
<td>3-16</td>
<td></td>
<td>Reduced OS (univariate) &lt;sup&gt;32&lt;/sup&gt;</td>
</tr>
<tr>
<td>secondary AML</td>
<td></td>
<td>5-7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>de novo AML</td>
<td></td>
<td>2-7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tAML</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ZRSR2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disorder</td>
<td>TET2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDS</td>
<td>1-11</td>
<td>13-37, 45, 49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>secondary AML</td>
<td>2-8</td>
<td>30, 43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMF</td>
<td>6-11</td>
<td>30, 42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>secondary AML</td>
<td>1.8</td>
<td>35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>de novo AML</td>
<td>5.6</td>
<td>35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tAML</td>
<td>1.13</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PFS: progression-free survival; OS, overall survival; LFS: leukemia-free survival; CLL: chronic lymphocytic leukemia; MDS: myelodysplastic syndromes; RARS: refractory anemia with ringed sideroblasts; RCMD-RS: refractory cytopenia with multilineage dysplasia and ringed sideroblasts; RARS-T: refractory anemia with ringed sideroblasts and thrombocytosis; RA: refractory anemia; RAEB: refractory anemia with excess blasts; AML: acute myeloid leukemia; tAML: therapy-related AML; CMML: chronic myelomonocytic leukemia; MPN: myeloproliferative neoplasm; PMF: primary myelofibrosis; ET: essential thrombocytosis.
Table 2. Summary of hematopoietic changes observed in mouse models of mutant splicing factors.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Mutation</th>
<th>Mouse model</th>
<th>WBC</th>
<th>RBC</th>
<th>PLT</th>
<th>HSPCs</th>
<th>Competitive Repopulation</th>
<th>MDS/AML</th>
<th>Pre-mRNA splicing</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sf3b1</td>
<td>Knockin (activated by cre)</td>
<td>K700E</td>
<td>=</td>
<td>↓</td>
<td>Hg</td>
<td>=</td>
<td>↑&lt;sup&gt;54&lt;/sup&gt;, =&lt;sup&gt;55&lt;/sup&gt; (LT-HSCs)</td>
<td>↓</td>
<td>MDS-like (Erythroid dysplasia)</td>
<td>Alternative 3' SS</td>
</tr>
<tr>
<td>Srsf2</td>
<td>Knockin (activated by cre)</td>
<td>P95H</td>
<td>=</td>
<td>↓</td>
<td>Hg</td>
<td>=</td>
<td>↑ (LSK)</td>
<td>↓</td>
<td>MDS-like (Multilineage dysplasia)</td>
<td>Altered affinity for ESE motif SSNG</td>
</tr>
<tr>
<td>U2AF1</td>
<td>Transgenic (DOX inducible)</td>
<td>S34F</td>
<td>=</td>
<td>=</td>
<td>=</td>
<td>↑ (LSK)</td>
<td>↓</td>
<td>No</td>
<td>Altered 3' consensus sequence at -3bp</td>
<td>Lunn-Shirai et al., 2015.52</td>
</tr>
</tbody>
</table>

Hg: Hemoglobin; MCV: mean corpuscular; WBC: white blood cells; RBC: red blood cell; PLT: platelets; HSPC: hematopoietic stem and progenitor cell; LKS: Lin<low>, Kit<+>, Sca1<+>; LT-HSC: Lin<low>, Kit<+>, Sca1<+>, CD48<+>, CD150<+> or Lin<low>, Kit<+>, Sca1<+>, CD34<+>, CD135<+>; SS: splicing site; DOX: doxycycline; cre: cre recombinase.
FIGURE LEGENDS

Figure 1. Consequences of somatic mutations affecting cis-acting pre-mRNA sequences. (A) A representative 3-exon pre-mRNA model is depicted with annotation of the location of common cis-acting sequences affected by somatic mutations. (B) Common splicing outcomes that are caused by somatic mutations in cis-acting sequences or splicing factors include exon inclusion, exon skipping and intron retention. Other splicing outcomes include alternative 5’ or 3’ splice sites (mutually exclusive exons, coordinate cassette exons, and alternative first and last exons are not shown). The yellow color in exons 1 and 2 represents exonic sequence that is excluded when splicing into an alternative cryptic splice site occurs, as indicated by the dashed lines at the bottom of each cartoon. (C) The predominant type of pre-mRNA splicing alteration induced by mutations in various cis-acting sequences. Last-base exonic mutations (LBEM), exonic-splice enhancer (ESE), exonic-splice silencer (ESS).

Figure 2. Splicing factor (trans-acting) mutations and their impact on splicing in hematological malignancies. (A) Sf3b1 mutations lead to alternative 3’ splicing site usage due to the increased recognition of cryptic splice sites between the branch point and the canonical 3’ splice site. (B) U2AF1 mutations affect 3’ splice site recognition leading to an increase in exon skipping due to use of an alternative splice sites. S34F-mutant-U2AF1 losses affinity for the 3’ SS motif when a T(U) is present in position -3. Conversely, mutations at the Q157 residue promote the recognition of 3’ SS when a G is present in position +1 and repress those bearing an A. (C) Srsf2 mutations affect its normal RNA binding activity to the consensus ESE motif (SSNG). Mutant SRSF2 recognizes with higher affinity the CCNG motif versus the GGNG, promoting or repressing the inclusion of exons containing C or G rich motifs. (D) Zrsr2 loss-of-function mutations specifically affect splicing of U12-type introns leading to intron retention.

SS: splicing site; BP: branch point; Py-tract: polypyrimidine-tract

Figure 3. Potential therapeutic strategies targeting splicing factor mutations and splicing changes. A schematic of the major spliceosome components illustrates potential strategies to modulate splicing. (A) Small molecules (e.g., PB and FR derivatives and analogs) alter the mRNA binding affinity of the U2 snRNP, leading
to the accumulation of unspliced products. (B) Inhibition of SR protein phosphorylation alters intracellular trafficking and binding of SR proteins to exonic splice enhancers, leading to altered utilization of cassette exons. (C) Oligonucleotides and transsplicing modulators can change splicing preference at specific targets. (D) Depletion or replacement of spliced products. BPS, branch point sequence; PPT, polypyrimidine tract; CLS, CDC-like kinase; SRPK, SR protein kinase.
REFERENCES


A

Exon inclusion
Exon skipping
Intron retention

B

Splicing outcomes

Exon inclusion
Exon skipping
Intron retention
Alternative 5’ splice site
Alternative 3’ splice site

C

<table>
<thead>
<tr>
<th></th>
<th>Splice site mutations</th>
<th>LBEM mutations</th>
<th>Gain ESE or lose EES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon inclusion</td>
<td></td>
<td></td>
<td>✓ Enriched in oncogenes</td>
</tr>
<tr>
<td>Exon skipping</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron retention</td>
<td>✓</td>
<td>✓</td>
<td>✓ Enriched in tumor suppressor genes</td>
</tr>
</tbody>
</table>

Figure 1
Figure 2

A

U2-snRNP

ESE
GU
A
AG
U2AF2
AG
ESE
GU

K700E

TTAG

Alternative 3' SS usage

B

U2-snRNP

ESE
GU
A
AG
U2AF2
AG
ESE
GU

S34(F/Y)

TTAG

Exon Skipping

Q157(P/R)

TTAG

Exon Skipping

C

U2-snRNP

ESE
GU
A
AG
U2AF2
AG
ESE
GU

P95(H)

TTAG

Exon Inclusion

D

U2-snRNP

ESE
GU
A
AG
U2AF2
AG
ESE
GU

WT

TTAG

Mut

U12-type intron

Intron Retention
Figure 3

A. Modulation of the U2 snRNP

B. Inhibition of SR protein phosphorylation

C. Modification of splicing preferences

D. Targeting of aberrantly spliced products

U1, U2, U2AF1, U2AF2, SR, CLKs, SRPKs, ESE, GU, AG, BPS, PPT, 5' SS, 3' SS
Splicing factor gene mutations in hematologic malignancies

Borja Saez, Matthew J. Walter and Timothy A. Graubert