Altered RNA Processing in Cancer Pathogenesis and Therapy

Esther A. Obeng¹, Connor Stewart², and Omar Abdel-Wahab²

ABSTRACT

Major advances in our understanding of cancer pathogenesis and therapy have come from efforts to catalog genomic alterations in cancer. A growing number of large-scale genomic studies have uncovered mutations that drive cancer by perturbing cotranscriptional and post-transcriptional regulation of gene expression. These include alterations that affect each phase of RNA processing, including splicing, transport, editing, and decay of messenger RNA. The discovery of these events illuminates a number of novel therapeutic vulnerabilities generated by aberrant RNA processing in cancer, several of which have progressed to clinical development.

Significance: There is increased recognition that genetic alterations affecting RNA splicing and polyadenylation are common in cancer and may generate novel therapeutic opportunities. Such mutations may occur within an individual gene or in RNA processing factors themselves, thereby influencing splicing of many downstream target genes. This review discusses the biological impact of these mutations on tumorigenesis and the therapeutic approaches targeting cells bearing these mutations.

INTRODUCTION

The advent of high-throughput transcriptome sequencing (RNA-seq) has provided a wealth of information on RNA splicing on a genome-wide scale. It is now understood from RNA-seq that >95% of human genes are subject to alternative splicing (AS), the enzymatic process by which a single gene has the potential to produce multiple, potentially functionally distinct pre-mRNA and protein isoforms. Splicing is considered to be a major mediator of proteome diversity through its ability to generate multiple transcripts with differing amino acid sequences from a single gene. Moreover, due to the link between splicing and nonsense-mediated mRNA decay (NMD), splicing also provides a means to regulate gene expression.

Systematic application of RNA-seq to an ever-expanding number of human tumors and matched normal tissues has now identified numerous means by which RNA processing is dysregulated in cancer. In this review, we focus on how RNA splicing and polyadenylation are altered in cancer and functionally drive cancer initiation and maintenance. This is notably only a subset of the means by which RNA processing goes awry in cancer. Cancer-associated changes in RNA editing, RNA modifications, and expression of noncoding RNA species including micro-RNAs and long noncoding RNAs are covered in recent excellent publications (1–4). Finally, motivated by recent data illustrating individual splicing alterations as well as mutations in RNA-splicing factors as therapeutic vulnerabilities in cancer, we discuss ongoing and future efforts to target RNA splicing for cancer therapy.

BASIC MECHANISMS OF RNA SPLICING CATALYSIS

RNA splicing is a nuclear enzymatic process accomplished by a macromolecular machine composed of a large constellation of RNA binding proteins (RBP) and additional splicing proteins combined with five small nuclear RNAs (snRNA) in complexes known as small nuclear ribonucleoproteins (snRNP; reviewed recently; refs. 5–7). snRNAs base pair with sequences within pre-mRNA that are critical in delineating exons from introns. These include the dinucleotides at the first two and last two positions of an intron [known as the 5′ and 3′ splice sites (ss), respectively] and a poorly conserved sequence within the intron known as the branchpoint (Fig. 1A). The branchpoint is typically close to the 3′ ss and is usually an adenosine nucleotide. A stretch of pyrimidine nucleotides (the polypyrimidine tract) is often adjacent to the branchpoint and promotes the spliceosome’s recognition of the branchpoint. Deletion or mutation of these intronic
A

![Diagram of splicing events](image)

1. **Normal splicing event**
2. **Exon skipping**
3. **Intron retention**
4. **Splice-site creating mutation**

B

![Diagram of splicing mechanisms](image)

1. **Exonic splicing enhancer**
2. **Exonic splicing suppressor**
3. **Intronic splicing enhancer**
4. **Intronic splicing suppressor**

C

1. **Cassette exon**
2. **Alternative 5’ splice site**
3. **Alternative 3’ splice site**
4. **Intron retention**
5. **Mutually exclusive exon**

D

1. **RefSeq**
2. **GenCode**
3. **Ensembl**

- **Cassette exon**
- **Alternative 5’ or 3’ ss**
- **Retained intron**
- **Other**
sequences typically greatly abrogates splicing usage at these sites. In addition, sequences throughout the exon and intron, referred to as splicing enhancers and silencers, are bound by RBPs that recruit or repel spliceosome assembly to promote or inhibit splicing, respectively.

The enzymatic process of splicing consists of two sequential transesterification reactions (Fig. 1B). In the first step, the 2′-hydroxyl group of the branchpoint performs a nucleophilic attack on the phosphorous atom of the 5′ ss to generate a linear left exon and a right intron–exon branched sequence known as a lariat. In the second step, the 2′-hydroxyl group of the linear left exon attacks the phosphorous atom of the 3′ ss, thereby concatenating two exons. An excised intron is simultaneously released within the lariat and later degraded by RNA debranching enzymes.

Over the last four years there have been major advances in the mechanistic understanding of spliceosome function, with the publication of at least 18 independent structures of the spliceosome at a resolution of 3.3 to 9.9 Å (reviewed recently; refs. 8–10). Although these initial efforts were mostly carried out using cryoelectron microscopy (cryo-EM) of yeast spliceosomes, initial structures of the human spliceosome followed in 2017. It is almost certain that high-resolution structures of the human spliceosome in each step of the splicing reaction will be elucidated in the very near future.

**ALTERATIONS IN RNA SPlicing IN CANCER**

**Splicing Changes in Cancer Relative to Normal Tissues**

Prior to the advent of RNA-seq, evaluation of expressed sequence tag libraries from human cancer cells and other tissue types revealed that cancer cells have an elevated rate of stop codons relative to noncancer cells (11). These data suggested increased missplicing in cancer transcriptions relative to normal tissues; the basis for which was not entirely clear. Over the last 10 years, however, a wealth of DNA- and RNA-sequencing data from tumors and paired normal tissues as part of The Cancer Genome Atlas (TCGA) effort, as well as distinct normal human tissues from the Genotype-Tissue Expression Program (GTEx), have enabled systematic interrogation of splicing across cancers and evaluation for a potential genetic basis for splicing alterations in cancer. The largest of such studies, recently published by Kahles and colleagues, reanalyzed whole-exome sequencing (WES) and RNA-seq data across 32 TCGA cancer types from 8,705 patients (12). This effort showed that tumor samples, on average, harbor ~20% more AS events than normal tissues. As such, many splicing changes identified in cancer cells can be linked to single-nucleotide variants (SNV) within that gene that disrupt mRNA sequences required for normal splicing (Fig. 1C). Heterozygous cis-acting splicing alterations act in an allele-specific manner such that the mutant allele shows abnormal splicing, whereas the wild-type allele supports normal splicing. Interestingly, prior work from Supek and colleagues and Jung and colleagues found that SNVs resulting in frameshifts in mRNA sequence due to intron retention (IR) occur most commonly in tumor suppressors while sparing oncogenes (13, 14). In contrast, in-frame exonic SNVs disrupting exon usage are most commonly enriched in oncogenes.

Although most studies evaluating the effects of SNVs on splicing have focused on mutations abolishing splice sites, recent reanalysis by Jayasinghe and colleagues of 8,000 tumor samples from 33 TCGA cancer types identified many examples of mutations creating novel splice sites (Fig. 1C; ref. 15). These splice site-creating mutations affect many known cancer genes. Intriguingly, tumors bearing these mutations also appeared to be associated with increased expression of T cell–associated genes and mRNAs encoding the immune-checkpoint blockade molecules PD-1 and PD-L1. These data suggest the potential for splicing to generate novel, immunogenic peptides (a topic discussed later in this review).

It is important to note that most studies evaluating the impact of cis-acting mutations on splicing have utilized WES data and focused on mutations that occur at 5′ or 3′ exon–intron boundaries. However, in vitro screening of sequence variants across every nucleotide within model exons has shown that >50% of nucleotide substitutions in exons can induce splicing changes with similar effects expected from coding and synonymous (or “silent”) mutations (16, 17). Consistent with this, analysis of >3,000 cancer exomes and >300 cancer genomes by Supek and colleagues identified recurrent selection for synonymous mutations in cancer, the majority of which affected RNA splicing and occurred in oncogenes (14). Notably, many regulatory elements required for ss recognition occur deep within introns (at sites including the branchpoint, polypyrimidine tract, and intronic splicing enhancers and silencers), and longer introns are more prone to splicing errors (18). Thus, increased use of whole-genome sequencing paired with RNA-seq of the same tumor samples will be important in improving our knowledge of how mutations in these noncoding regions affect expression of protein-coding isoforms through splicing.
There have been a variety of efforts to quantify the distribution of splicing changes in cancer cells (Fig. 1D). At least one report found no biases in recognition of cassette exons or 5′ or 3′ ss globally in cancer, but almost all cancer types exhibit widespread increased IR (19). Although these findings are intriguing, they have not been uniformly replicated using alternative methods of splicing analysis (20). It is important to note that the detection and quantification of splicing changes from RNA-seq data remains a formidable challenge for several reasons. First, although numerous informatics tools have been described to identify and quantify splicing (reviewed recently; refs. 20–22), no single bioinformatics method is uniformly used, and these tools vary widely in their requirement for prior transcriptome annotation. Clearly, reliance on preannotated libraries of known transcripts limits detection of novel, unannotated splicing events. In addition, existing analysis tools for detecting splicing changes vary in their false-positive rate for detection of certain splicing events and ability to detect complicated splicing events (consisting of multiple simultaneous events from those shown in Fig. 1C; refs. 20, 22). For example, there are many examples where use of alternative exons residing within an intron can be mistaken for IR (20). Discrepancies in the description of global RNA-splicing changes in cancer highlight the challenges of inferring splicing changes from short-read RNA-seq data. Future efforts applying long-read RNA-seq methodologies to capture data from entire mRNA transcript isoforms and methods of RNA-seq analysis which do not rely on preannotation of splicing isoforms will be crucial in further illuminating our understanding of altered splicing in cancer.

Mutations in RNA Splicing Factors in Cancer

Although the above data cataloging how mutations within genes affect their own splicing in cis elucidated important roles for splicing in cancer, the discovery of recurrent mutations in components of the RNA-splicing machinery in 2011 further highlighted the importance of aberrant splicing in cancer (23). Currently, hotspot, heterozygous mutations in the RNA-splicing factors SF3B1, SRSF2, and U2AF1 are known to be enriched throughout myeloid leukemias, chronic lymphocytic leukemia (CLL), uveal melanoma (UVM), and mucosal melanoma (Fig. 2A and B; refs. 20–30). A recent study surveyed mutations in 404 genes encoding splicing factors from 33 cancer types in the TCGA from investigators at H3 Biomedicine and the TCGA identified putative driver mutations in 119 splicing factor genes (31). These include a number of additional splicing factors affected by recurrent hotspot mutations (PHFSA, PCBP1, and HNRNPC) as well as a larger number of splicing factors affected by loss-of-function mutations (RBMM, FUPB1). Although the overall mutational frequency in RNA-splicing factors is low across unselected solid cancer types, these mutations recur at rates higher than expected by chance in the melanoma subtypes noted above in addition to lung adenocarcinoma (LUAD) and bladder cancer.

The most commonly mutated splicing factor gene across all cancers is SF3B1. The initial discovery of mutations in SF3B1 was surprising and unexpected for a variety of reasons. First, this was the first example of mutations in a splicing factor positively selected for in clonal disorders. Second, mutations in SF3B1 have an exquisite and intriguing enrichment in otherwise unrelated cancer types, including myelodysplastic syndromes (MDS), CLL, and UVM. Finally, within each of these disorders, SF3B1 had immediate relevance for disease diagnosis and/or prognostication. For example, SF3B1 mutations in MDS are present in >90% of patients with a subtype of MDS known as refractory anemia with ring sideroblasts (RARS) characterized by anemia, the finding of iron-laden mitochondria surrounding the nuclei of erythroid precursors (“ring sideroblasts”) in the bone marrow, and an overall favorable prognosis (23, 24). This form of MDS has been morphologically recognized for decades but a genetic basis was unknown until the discovery of SF3B1 mutations. In fact, SF3B1 mutations are now part of the diagnostic criteria for RARS, as SF3B1 mutations have >97% positive predictive value for patients suspected to have this form of MDS (32).

Within MDS and other myeloid malignancies, SF3B1 mutations are present in the predominant clone and are selected early in the disease process. In contrast, SF3B1 mutations in CLL are most commonly subclonal and enriched in patients with more advanced and aggressive disease (26, 33). SF3B1 mutations are present in 10% to 20% of patients refractory to chemotherapeutic agents such as fludarabine in contrast to <5% of patients with untreated CLL or the CLL precursor monoclonal B-lymphocytosis (reviewed previously in ref. 34). SF3B1 mutations are also more common in patients with CLL with an unmutated immunoglobulin heavy-chain variable region (IGHV) gene (a well-established adverse prognostic factor in this disease), and are enriched in patients with CLL requiring therapeutic intervention over those not requiring therapy. Although SF3B1 mutations have been associated with adverse response to chemotherapeutic agents such as ibritinib, venetoclax, and PI3K inhibitors is not clear.

In the setting of UVM, SF3B1 mutations occur in 10% to 21% of patients and are associated with specific favorable prognostic features including disomy 3 and lower age at diagnosis (27–29). The impact of SF3B1 mutations in this specific subset of patients with UVM has been conflicting, with some reports suggesting SF3B1 mutations are associated with an increased risk of metastasis among disomy 3 patients.

SF3B1 is a member of the U2 snRNP complex, where it physically associates with p14, PHF5A, SF3B3, and the U2 snRNA (Fig. 1B). The U2 snRNP complex is important in recognizing the branchpoint within the intron. Consonant with this, global surveys of bulk RNA-seq data from cancer cells and mouse models bearing mutations in SF3B1 at the most commonly mutated residues in SF3B1, including the SF3B1 K700E and K666N mutations, reveal that SF3B1-mutant cells exhibit use of an aberrant intron-proximal 3′ ss (Fig. 2C; refs. 35–37). Further work by Darman and colleagues, Tang and colleagues, and Carrocci and colleagues has shown that this change in 3′ ss occurs due to reduced fidelity of branchpoint selection in cells expressing mutant SF3B1 (35, 38, 39). Each of the mutations in SF3B1 occurs
Figure 2. Frequency, location, and global impact on RNA splicing of recurrently mutated splicing factors in cancer. **A,** Histogram depicting frequency of mutations in SF3B1, SRSF2, U2AF, and ZRSR2 in hematopoietic malignancies and solid tumors. In addition to these four genes (which are the most frequently mutated in hematopoietic malignancies), a host of additional splicing factors affected by hotspot as well as presumed loss-of-function mutations are also mutated in cancer and not shown here. RARS, refractory anemia with ring sideroblasts; RCMD-RS, refractory cytopenias with multilineage dysplasia and ring sideroblasts; MDS, myelodysplastic syndromes; CMML, chronic myelomonocytic leukemia; AML-MRC, acute myeloid leukemia with myelodysplasia-related changes; CLL, chronic lymphocytic leukemia. **B,** Protein diagrams of the four mutated splicing factors shown in **A** with location of mutant residues. Hotspot mutations are shown in red. HD, HEAT repeat domain; Zn, Zinc finger; RRM, RNA recognition motif; RS, serine/arginine-rich domain; UHM, U2AF homology motif. **C,** Diagram of an intron, two flanking exons, and locations in RNA where the four factors from **A** bind (top left). Mutations in SRSF2 skew the binding avidity of SRSF2 such that mutants bind C-rich sequences more avidly to promote exon splicing while reducing binding affinity for G-rich sequences (middle left). SF3B1 is responsible for recognition of the branchpoint. Mutations in SF3B1 cause recognition of an aberrant branchpoint leading to intron proximal alternative 3′ ss selection (bottom left). Finally, U2AF1 is responsible for recognition of the 3′ yAG|r dinucleotide [where “y” represents the C- or T- pyrimidine nucleotide immediately intronic to the AG and r represents the first nucleotide in the downstream exon (at the +1 position)]. As shown on the right, U2AF1 S34F/Y mutations favor inclusion of cassette exons bearing a 3′ ss containing a C-nucleotide at the −3 position whereas Q157 mutations promote splicing of exons with G nucleotides at the +1 position. **A** and **B** adapted from Dvinge et al. RNA splicing factors as oncoproteins and tumor suppressors. Nat Rev Cancer 2016;16:413–30. Used with permission.
within the fourth to seventh HEAT repeat domains (Fig. 2B). Recent mapping of cancer-associated mutations in SF3B1 onto the crystal structure of human SF3B1 by Cretu and colleagues suggest that SF3B1 mutations alter the conformation of the HEAT repeat domains and/or interactions with U2AF2 or the SF3B complex protein p14 (40). Ultimately these changes would modify how the U2 complex interacts with the branchpoint. As noted above, SF3B1 is affected by a number of distinct hotspot mutations, many of which are associated with specific lineages of cancer. The most common individual hotspot mutation in SF3B1 is the SF3B1 K700E mutation, which is present across myeloid malignancies, CLL, and many solid tumors including breast cancer, pancreatic ductal adenocarcinoma, and others. In addition, there are several mutated residues in SF3B1 that appear to be cancer-specific. These include SF3B1 R625 mutations, enriched in melanomas; SF3B1 E902 mutations, exclusive to bladder cancer; and SF3B1 G742 mutations, most enriched in CLL. The functional basis for the lineage specificity of individual hotspot mutations in SF3B1 is currently unknown. It is possible that each mutation is associated with distinct missplicing events that may be important in transforming specific tissue types, giving rise to histologically distinct cancers. For example, the bladder cancer-specific E902 hotspot mutation is not associated with the same intron-proximal 3' ss usage as other SF3B1 mutational hotspots (31). However, the mechanistic basis for this observation or how it relates to bladder cancer pathogenesis is unclear. Finally, it is important to note that the lineage specificity of mutations could also occur due to the potential for tissue-specific nucleotide mutability, gene mutation rates, and/or expression of interacting proteins.

Although SF3B1 is the most frequently mutated splicing factor across cancers, more has been learned about how hotspot mutations in the splicing factors U2AF1 and SRSF2 affect splicing and disease development. U2AF1 is an RBP and part of a heterodimeric U2AF complex with its partner, U2AF2, that serves to recruit the U2 snRNP to the branchpoint (Fig. 2C). U2AF1 binds the AG dinucleotide at the 3' ss, whereas U2AF2 binds the polypyrimidine tract (Fig. 2C). U2AF1 is affected by hotspot mutations at the S34 and Q157 residues, each of which occurs in one of its two zinc fingers (Fig. 2B). Similar to the allele specificity of mutations in SF3B1, each mutant residue of U2AF1 is associated with specific subtypes of cancer. For example, S34 mutations are recurrent in LUAD whereas Q157 mutations are not present in this disease (41). Interestingly, work by Ilagan and colleagues, and others, showed that aberrant splicing driven by mutant U2AF1 occurs in an allele-specific manner at sequences surrounding the AG dinucleotide at the 3' ss (42). U2AF1 mutants skew normal splicing activity of U2AF such that U2AF1 S34 mutants promote exon inclusion at the 3' ss when a C-nucleotide is located at the -3 position, although Q157 mutants promote exon inclusion when a G-nucleotide is in the +1 position (Fig. 2C).

Unlike SF3B1 and U2AF1, which are core components of the spliceosome, SRSF2 is an auxiliary splicing factor that is a member of the serine/arginine-rich (SR) family of RBPs that binds splicing enhancers to recruit the core spliceosome to promote exon splicing. In its wild-type (WT) state, SRSF2 physically binds CCNG and GGNG sequences within RNA equally well to promote splicing (43). However, mutations in SRSF2, which occur as point mutations or in-frame deletions at P95 (Fig. 2B), skew this binding preference such that the mutant proteins promote RNA binding and splicing at C-rich sequences while having reduced affinity for G-rich sequences (Fig. 2C; refs. 44, 45).

Despite the consistency of these findings that sequence-specific aberrant splicing changes are associated with point mutations in SF3B1, U2AF1, and SRSF2, to date there have been very few examples linking splicing alterations induced by mutant splicing-factor mutations to functional consequences in a tumor model (see EZH2-aberrant splicing described below). These challenges likely relate to the lack of uniformity in methods for the detection of splicing alterations by RNA-seq as well as inherent limitations of certain splicing analysis tools noted earlier. Even once the full spectrum of reliable splicing changes induced by spliceosome gene mutations are identified, requisite functional studies to systematically understand the functional impact of individual splicing changes on protein expression and function, tumorigenesis, and cell differentiation are needed. In addition to evaluating the effect of aberrant splicing on the expression of individual proteins, it will also be critically important to systematically explore the effects of cancer-associated changes in splicing on the proteome. Such efforts would help elucidate the extent to which alterations in splicing result in the downregulation of canonical protein abundance due to either the expression of NMD-inducing transcripts or aberrant protein isoforms from novel splicing events.

In addition to SF3B1, SRSF2, and U2AF1, presumed loss-of-function mutations in a fourth RNA splicing factor, ZRSR2, also occur in myeloid malignancies (Fig. 2B). Interestingly, ZRSR2 is the only protein of the four frequently mutated splicing factors in leukemias that primarily functions in the minor spliceosome. Although most introns are spliced by the major spliceosome (so-called U2-type introns), a small subset (<1%) of introns have distinct 5' and 3' splice sites which are recognized by a separate splicing complex known as the minor spliceosome (46, 47). Mutations in ZRSR2, encoded on the X chromosome, have a male predominance in MDS. ZRSR2 mutations occur throughout its open reading frame, suggesting that mutations in ZRSR2 confer loss-of-function. Consequently, ZRSR2-deficient leukemia cells have been shown to have global increases in minor (or “U12-type”) IR. However, how these splicing changes in ZRSR2-mutant cells relate to those in cells bearing mutations in the other RNA-splicing factors is currently unclear (48). Likewise, which ZRSR2-regulated splicing events are important for myeloid leukemia pathogenesis, if any, are unknown.

How Do Mutations in Splicing Factors Confer a Cellular Advantage?

The fact that the splicing factors mutated in cancer play essential roles in splicing catalysis and the mutations are clearly associated with global alterations in RNA splicing highlight the potential that aberrant splicing of specific transcripts is responsible for disease development. However, it is still unclear whether missplicing of hundreds of distinct
mRNAs drives clonal selection as opposed to one or handfuls of individual missplicing events. Interestingly, several recent studies of subjects with age-related clonal hematopoiesis, a condition associated with a risk of developing MDS and acute myeloid leukemia (AML), have shown that mutations in the RNA-splicing factors SRSF2 and U2AF1 are highly predictive for the eventual development of overt myeloid malignancies (49, 50). These data, in combination with selection for specific mutant residues in these factors, further underscore the concept that mutant splicing factors confer clonal advantage to cells. Despite the strength of these human genetic data, currently there is very little evidence from experimental models that spliceosomal gene mutations confer a cellular benefit or are required for disease maintenance. For example, several studies in human cancer cell lines bearing naturally occurring mutations in RNA-splicing factors have demonstrated that eradication of mutant U2AF1 or SF3B1 has no impact on cell growth (51, 52). These data suggest the possibility that RNA-splicing factor mutations may be required for cancer initiation while being dispensable for tumor maintenance. At the same time, expression of mutations in RNA-splicing factors in conditional knockin mice has very consistently been associated with impaired self-renewal (37, 44, 53). However, these data do not exclude the possibility that expression of mutant splicing factors may confer an advantage in specific genetic contexts or in response to certain non-cell-autonomous stressors or microenvironmental stimuli that are not yet appreciated. To this end, it is important to note that mutations in RNA-splicing factors have nonrandom mutational co-occurrences that may be important in understanding the role of mutant splicing factors in disease pathogenesis. There is also the distinct possibility that nonhuman experimental models of splicing factor mutations may not accurately capture the biological impact of these mutations owing to species-specific changes in splicing.

Currently, very few mRNAs misspliced by mutant splicing factors have been rigorously functionally linked to disease development or dissected in molecular detail. One notable exception is the missplicing of EZH2 driven by mutant SRSF2. Mutations in SRSF2 promote expression of an unannotated isoform of EZH2 due to the presence of C-rich exon splicing enhancers within an exon that is typically skipped. This so-called poison exon contains a premature termination codon; thus, mutant SRSF2 promotes expression of a form of EZH2 mRNA that undergoes NMD (44). Thus, these findings explain the paradoxical observations that loss-of-function mutations in EZH2 are enriched in the same constellation of myeloid neoplasms as mutations in SRSF2; however, EZH2 and SRSF2 mutations are significantly mutually exclusive with each other (54). Moreover, EZH2 loss functionally promotes myeloid malignancy development in vivo whereas restoration of EZH2 expression rescues the impaired hematopoiesis characteristic of mutant SRSF2. Despite this exquisite link between EZH2 aberrant splicing and mutant SRSF2, there are still hundreds of additional mRNAs misspliced by mutant SRSF2, and the relative importance of these additional events to the pathogenesis of SRSF2-mutant cancers is not clear.

In addition to altered splicing, mutant RNA-splicing factors may play a role in disease development through effects on gene expression or pathways not strictly related to splicing. For example, there is increasing evidence that transcriptional alterations unrelated to splicing of individual mRNAs are perturbed in spliceosomal-mutant cells. Transcription and RNA splicing are intimately coupled in vivo. In fact, RNA splicing occurs cotranscriptionally with recruitment of splicing factors to splicing signals in nascent RNA following RNA polymerase II (RNAPII) elongation. Interestingly, it is known that WT SRSF2 is important in RNAPII transcriptional elongation, and recent work has demonstrated that mutant SRSF2 is associated with impaired RNAPII pause release (55, 56). As a result, SRSF2-mutant cells actually have increased DNA/RNA hybrids (so-called R loops) at gene promoters and associated activation of the ATR DNA-damage response signaling pathway.

Recent work by Chen and colleagues and Nguyen and colleagues demonstrated that augmented R-loop formation and ATR activation is also seen in U2AF1-mutant cells (56, 57). However, it is not precisely clear how mutant U2AF1 alters RNAPII pause release. R-loop accumulation is also seen in the presence of compounds that impede splicing catalysis, suggesting a more general contribution of splicing fidelity to RNAPII elongation (58). Although R-loops have been associated with increased genomic instability, myeloid neoplasms bearing mutations in splicing factors are not known to harbor high mutation burdens or aneuploidy. Although the exact contribution of R-loops to leukemia pathogenesis is not yet clear, ATR activation by R-loops may have therapeutic relevance in spliceosomal-mutant cells. Recent studies have demonstrated that U2AF1-mutant cells are preferentially sensitive to ATR inhibition (58). This may have implications for ongoing clinical trials of ATR inhibitors for patients with cancer.

It is also important to note that the earliest reports of RNA-splicing factor mutations suggested a link between aberrant RNA splicing and NMD. Overexpression of mutant U2AF1 in cell lines was associated with increased expression of mRNAs encoding NMD components (23). Subsequent evaluation of SF3B1-mutant cancer cell lines identified an 8- to 10-fold decrease in canonical protein expression of genes with transcripts misspliced by mutant SF3B1 and predicted to undergo NMD (35). However, it has yet to be demonstrated whether the actual enzymatic process of NMD is altered in any way in spliceosomal-mutant cells or if these cells are preferentially reliant on NMD to process aberrant transcripts. Finally, there have been recent intriguing reports of roles for WT and mutant U2AF1 on translation of cytosolic mRNAs and alternative polyadenylation (APA; refs. 59, 60). How these potential roles of U2AF1 relate to its well-established function in nuclear RNA splicing is not yet clear, nor is it clear if other recurrently mutated splicing factors can affect these processes as well.

In addition to mutations in genes encoding RNA-splicing factors, there are numerous examples of alterations in expression of splicing factors that play a causative role in tumorogenesis (reviewed recently in ref. 61). For example, several studies have identified that overexpression of SRSF1, which is upregulated in cancers with amplification of chromosome 17q23, can transform a variety of cell types (62, 63). Although SRSF1 is involved in NMD, RNA export, and translation in
In addition to splicing, it has previously been identified that the oncogenic activity of SRSF1 depends on its regulation of splice isoforms of critical regulators of apoptosis and cell survival. In addition to this well-studied example, there is an interesting recent description of a specific form of kidney cancer marked by fusions involving splicing factors. This condition, known as translocation renal cell carcinoma, contains fusions of the transcription factor MITF to a variety of partners, many of which are splicing regulatory factors including SFPQ, LUC7L3, KHSRP, and KHDRBS2 (64). Further efforts to determine whether and how RNA splicing is altered in this condition may be very informative.

**ALTERATIONS IN mRNA PROCESSING BEYOND SPlicing: ALTEREd POLYadenylATion IN CANCER**

In addition to RNA splicing, processing of 3′ ends of mRNA is also critical in the regulation of gene expression and function. As nascent mRNA is generated by RNApolI, the 3′ end of the RNA must undergo endonucleolytic cleavage followed by synthesis of a poly(A) tail. These two coupled reactions are referred to as cleavage and polyadenylation (or simply polyadenylation). It has been known for several decades that through APA, a single gene can give rise to multiple transcripts that differ only in the sequence at their 3′ termini through cleavage at distinct polyadenylation sites (PAS). It is currently thought that at least 70% of mammalian mRNAs express APA isoforms (reviewed recently in ref. 65). APA has many parallels to AS in that it is an enzymatic process that (i) exhibits tissue specificity (66), (ii) is regulated by cis elements embedded in mRNA sequences and trans-acting proteins recognizing these sequences, (iii) can result in alternative mRNA isoforms, and (iv) occurs mostly cotranscriptionally. Pre-mRNA is cleaved 10 to 30 nucleotides downstream of a polyadenylation (pA) signal, a hexameric sequence located 10 to 35 nucleotides upstream of the PAS. The canonical pA signal is AAUAAA, but this sequence can adopt variants more than 10 times weaker. Importantly, the splicing machinery is also critically important in promoting the use of PASs at the 3′ ends of genes. It is now understood that, in addition to helping to identify the 5′ ss, the U1 snRNA is also critically important in suppressing premature 3′ end cleavage and polyadenylation from cryptic PASs within introns (67). As such, blocking the binding of the U1 snRNA to the 5′ ss induces premature cleavage of transcripts within introns. This role of U1 in suppressing transcriptional termination increases the likelihood of the spliceosome identifying a downstream 3′ ss and is termed “telescoping.”

It is currently thought that most APA sites are located in 3′ untranslated regions (3′-UTR; refs. 66, 68). In such cases, the choice of one APA site over another does not change the protein-coding mRNA sequence, but results in mRNA species that differ in the sequences within the 3′-UTR (Fig. 3A). Given that the 3′-UTR contains cis elements involved in mRNA stability, translation, nuclear export, and mRNA localization, altering 3′-UTR content can have strong effects on mRNA expression and function. For example, more than half of microRNA (miRNA) target sites are located in alternative UTRs (69). Shortening of 3′-UTRs would therefore remove miRNA-mediated repression. There are several well-characterized examples of oncogenes upregulated by removal of miRNA repression sites through APA. For example, Cyclin D1 is a well-known proto-oncogene in many cancers and undergoes ectopic expression in mantle cell lymphoma. In most cases, a chromosomal translocation brings the coding region of CCND1 (encoding Cyclin D1) under the control of the IGHV promoter [the t(11;14) (q13;q32) translocation]. However, in some patients with mantle cell lymphoma who lack CCND1 translocations, Cyclin D1 overexpression is due to truncation of its 3′-UTR with loss of miRNA repression sites (Fig. 3B; ref. 70). In some cases, 3′-UTR shortening is due to mutations creating a novel pA signal. Moreover, mutation of the miRNA seed regions in a normal-length UTR revealed the same effects on upregulation of Cyclin D1 expression. Similar examples of upregulation of cancer-specific isoforms through loss of 3′-UTR regulatory regions have been described for Cyclin D2 (71) and CDC6 (72). In addition to shortening of 3′-UTRs to release mRNAs from miRNA-mediated repression, recent data also suggest that shortening of the 3′-UTR of an mRNA may allow release of miRNAs to suppress the expression of other mRNAs in trans (Fig. 3C; ref. 73). This idea is based on the finding that cancer cells with shortened 3′-UTRs are enriched in mRNAs proposed to act as competing endogenous RNAs (ceRNA).

In addition to these specific examples of alternative 3′-UTRs regulating mRNA expression, several recent studies have provided surprising evidence that 3′-UTRs also mediate mRNA–protein interactions and protein localization. Given that RBPs interact with 3′-UTRs, Berkovits and Mayr hypothesized that altering 3′-UTR length could affect a “scaffold” function of 3′-UTRs and change nascent protein interactions.

Figure 3. Altered gene regulation in cancer through alternative cleavage and polyadenylation of mRNAs. A, Schematic of how APA of mRNAs in the 3′-UTR of mRNAs results in two distinct isoforms that differ only in their 3′-UTRs. The 3′-UTR can contain multiple potential PASs and additional sequences that may be recognized by RBPs and/or mRNAs. Altering 3′-UTR length may affect mRNA-mediated gene repression, protein-protein interactions, mRNA stability, translation, export, and localization. “CDS” represents the coding sequence. B, Cyclin D1 upregulation is a well-studied example of how altering 3′-UTR length results in proto-oncogene activation. In a proportion of patients with mantle cell lymphoma, Cyclin D1 is upregulated through polymorphisms and mutations in the 3′-UTR that result in the use of a proximal PAS and a shortened 3′-UTR that lacks a miRNA seed region. C, Schematic illustrating how shortening of the 3′-UTR of an mRNA may allow release of miRNAs to suppress the expression of other mRNAs in trans. D, APA may also utilize PAS upstream of the normal stop codon and thereby alter the coding sequence (CDS) of mRNA. In this example, a proximal PAS site is located within an intron, and use of this PAS site results in production of a shorter protein with a novel 3′ amino acid sequence. E, In addition to alterations in 3′-UTR length, Cyclin D1 is also subject to use of upstream PASs. Polymorphism at the end of exon 4 (for example, the G870A polymorphism) may promote use of an intronic polyadenylation signal within intron 4. This cleaves both miRNA binding sites as well as sequences encoding the normal nuclear export signal (NES) from Cyclin D1. As a result, this Cyclin D1 protein isoform is restricted to the nucleus.
A. Alternative polyadenylation in 3′-UTRs

CDS

3′-UTR

AAAAA

Proximal PAS

Distal PAS

B. Mutations/polymorphisms in 3′-UTR

11q13

Cyclin D1

Exon 1

Exon 5

3′-UTR

C. Mutations/polymorphisms in 3′-UTR

Cyclin D1

Exon 1

Exon 5

3′-UTR

Cyclin D1 protein

Proximal PAS

Distal PAS

D. Upstream alternative polyadenylation

CDS

Genomic DNA

Last exon

Exon 1

Exon 5

3′-UTR

Cyclin D1 retained in nucleus

Proximal PAS

Distal PAS

E. Upstream alternative polyadenylation

Cyclin D1

Exon 1

Exon 5

3′-UTR

Cyclin D1 in nucleus and cytoplasm

Intronic PAS

G870A

Proximal PAS

Distal PAS

New protein-coding sequence
with proteins that affect cellular trafficking. In one such example, APA of the 3'-UTR of the mRNA encoding the transmembrane protein CD47 resulted in short or long isoforms of CD47 mRNA that were translated in association with different protein complexes, leading to CD47 localization at either the plasma membrane (long 3'UTR) or the endoplasmic reticulum (short 3'UTR; ref. 74).

Despite individual examples of altered APA events in cancer-associated mRNAs, there have been only a handful of surveys of global APA usage in cancer. Global APA evaluation has been limited by the fact that standard RNA-seq does not effectively capture 3' ends of RNA, and few informatic methods identify APA sites and polyA tail lengths from standard RNA-seq. One key study utilized a novel bioinformatics algorithm for de novo identification of APAs from standard RNA-seq and applied this to 358 TCGA tumor/normal pairs from seven cancer types (75). This identified many tumor-specific APA events in cancer cells, most of which utilize short 3'-UTRs and are generally associated with increased gene expression, likely due to escape from miRNA-mediated gene repression. At the same time, this study also revealed that APA appears to be highly tumor-specific. For example, lung, uterus, breast, and bladder cancers have significantly more APAs than head/neck or kidney cancers. APA has also been shown to be tissue-specific in normal tissues (76). These observations raise the question of what regulates APA.

Recent genetic alterations in genes encoding the cleavage and polyadenylation machinery have not been identified, and recurrent mutations affecting pA sites in cis have not been described in cancer, outside of the few examples noted above. Thus, it is currently believed that APA is regulated by changes in expression of the machinery responsible for APA. The core polyadenylation trans-factors include four multi-subunit protein complexes (CPSF, CSTF, CFI, and CFII). CPSF recognizes the pA signal and CSTF binds downstream U/GU-rich elements (mostly through the CSTF64 subunit) which help mark the pA signal. Interestingly, Xia and colleagues found increased expression of CSTF64 mRNA in most tumors, which is hypothesized to promote usage of proximal, weaker pA sites and prevent usage of stronger, distal pA sites (75).

Although most APA sites occur in 3'-UTRs, there is increasing recognition that PASs may be located upstream of the last exon, usually within introns. Use of such upstream APA sites [also referred to as intronic polyadenylation (IpA) sites when occurring within introns] would be expected to alter the coding region of a gene, similar to how AS would alter the protein-coding mRNA sequence (Fig. 3D). There is increasing recognition that cancer cells may utilize aberrant IpA sites more frequently than normal cells and that IpA may result in truncated, aberrant mRNAs with important functional impact (77). It is important to note that IpA also occurs during normal physiologic regulation of gene expression. For example, as B cells terminally differentiate into plasma cells, the mRNA encoding immunoglobulin M heavy-chain (IgHM) undergoes an IpA event to skip exons encoding the transmembrane domains of IgHM (78). This results in plasma cells expressing a form of IgHM that is soluble and excreted. Several recent studies have applied 3' RNA-seq to normal immune cells in addition to multiple myeloma (76) and CLL (77). Although CLL was found to have increased IpA compared with normal B-cell subsets, myeloma cells had reduced IpA usage. In either instance, aberrant IpA events could be speculated to promote tumorigenesis. In addition to undergoing 3' UTR shortening, CCND1 is also subject to IpA in a manner that promotes tumorigenesis (Fig. 3E; refs. 79, 80). A protein isoform of Cyclin D1 (known as Cyclin D1b) is produced when CCND1 mRNA is cleaved at an APA site within an intron. This removes both the 3' UTR mRNA repression sites as well as sequences encoding the protein's normal nuclear export signal. Generation of Cyclin D1b may occur due to polymorphisms at the end of exon 4, which promote use of a polyA signal within an intron just downstream of exon 4.

As mentioned above, polyadenylation of mRNAs is coupled with mRNA cleavage. Polyadenylation, the template-independent addition of adenosine homopolymers to the 3' end of mRNAs, is critical in regulating mRNA stability. The initial addition of the poly(A) tail occurs in the nucleus and is a fairly well understood process. However, further additions to poly(A) tail length may occur in the cytoplasm by cytoplasmic poly(A) polymerases. This process and the RNA species regulated by it are less well understood. Interestingly, recent data suggest that one of the most commonly mutated genes in multiple myeloma, FAM46C, encodes a cytoplasmic poly(A) polymerase (81). FAM46C, encoded at the 1p12 locus, is affected by deletions as well as homozygous and hemizygous mutations in approximately 10% of patients with myeloma. Work from two groups suggests that FAM46C loss regulates the growth of myeloma cells across a variety of myeloma cell lines (81, 82). However, the RNA substrates of FAM46C and why it is so uniquely mutated in myeloma, as opposed to other cancers, are not understood.

**THERAPEUTIC TARGETING OF RNA PROCESSING IN CANCER**

**Rationale for Targeting Splicing for Cancer Therapy**

As described earlier, it has been documented that cancer cells harbor widespread changes in RNA splicing compared with normal cells, including increased expression of known pro-oncogenic and antiapoptotic isoforms of genes such as MDM2, BCL(X), and VEGF (83). These findings suggested that modulating RNA splicing of specific transcripts, or even globally, might have therapeutic benefit. The discovery of heterozygous hotspot mutations in spliceosome components further highlighted RNA splicing as a potential therapeutic vulnerability for cells bearing these mutations. Initial clues to this possibility came from human genetic data. For example, although hotspot mutations in BRAF, RAS, PIK3CA, and other signaling proteins commonly undergo allele imbalance in a manner that increases the dosage of the mutant allele, this conspicuously does not occur with hotspot mutations in SRSF2, SF3B1, or U2AF1 (84). Instead, cells bearing mutations in these genes consistently retain expression of the WT allele and never undergo loss of heterozygosity or become hemizygous. These data indicate haploinsufficiency of mutations in SF3B1, SRSF2, and U2AF1, and suggest that increased expression of the mutant allele is negatively selected in cells.
Consistent with this genetic requirement for the WT allele in spliceosome–mutant cells, spliceosomal gene mutations are also highly mutually exclusive with one another (23, 85). In MDS, where mutations in the three most commonly mutated splicing factors, SF3B1, SRSF2, or U2AF1, are present in >50% of patients, fewer than 1% of patients harbor a mutation in >1 of these genes simultaneously. These data again argue for the strong requirement of a certain level of normal splicing catalysis in splicing factor–mutant cells.

Each of the above genetic observations has now been evaluated in functional studies of the cellular and phenotypic effects of changing allelic ratios of splicing factor mutations. For example, expression of Srsf2P95H in a homozygous (Srsf2P95H/Srsf2P95H) or hemizygous (Srsf2P95H/null) state led to immediate elimination of Srsf2-mutant hematopoietic cells (86). Very similar findings have been seen in the setting of U2AF1S34F-mutant lung cancer cell lines and SF3B1-mutant breast cancer cells (51, 52). Moreover, induced simultaneous expression of SF3B1P700E/WT and Srsf2P95H/WT mutations in hematopoietic cells in vivo resulted in synthetic lethality (85). The precise mechanistic basis for why mutations in SRSF2, U2AF1, or SF3B1 are intolerable when expressed without the WT allele or combined in the same cell is still not fully understood.

In addition to a requirement for WT spliceosome function in splicing factor–mutant cells, several other genetic subtypes of cancer have been suggested as preferentially sensitive to alterations in splicing. These include several reports of a requirement for normal splicing in MYC-amplified cancers, including glioblastoma, breast cancer, and lymphomas (87–90). Studies have shown that MYC upregulates transcription of the splicing regulatory proteins PTB and HNRNPA1/2, which in turn promote use of the pro-oncogenic isoform of pyruvate kinase (known as PKM2) while suppressing expression of the alternative, mutually exclusive isoform PKM1 (87, 90). More recently, it has been appreciated that the effects of MYC on splicing extend beyond this individual splicing event in PKM1/2 to create global alterations in splicing of many mRNAs (88, 89). Consistent with this, expression of MYC has been identified as conferring sensitivity to genetic or pharmacologic inhibition of splicing (89).

In addition to splicing factor–mutant cells and MYC-driven cancers, given that RNA splicing is essential for cell survival, there have been several studies identifying splicing as a therapeutic vulnerability for cells with “bystander” genetic deletions of splicing factors. For example, breast cancer cell lines containing partial deletion of SF3B1 are highly sensitive to further downregulation of SF3B1, whereas SF3B1 copy-neutral cells are more tolerant to downregulation of SF3B1 (91). Similarly, cells with deletion of the exon–junction complex core member MAGOH, which commonly occurs in cancer cells containing chromosome 1p deletion, are sensitive to loss of the MAGOH paralog MAGOHB (92).

Discovery of Chemical Modulators of Splicing: SF3B-Binding Compounds

In parallel to the above data presenting a rationale for inhibition of splicing in cancer, a series of chemical biology studies uncovered diverse compounds that target the spliceosome. The initial discovery of compounds that selectively inhibit splicing came in 2007 when investigators from Riken and Eisai Co. demonstrated that the natural products FR901464 and pladienolides, and their derivatives including spliceostatin A and E7107, respectively, physically bind to the SF3B complex and inhibit pre-mRNA splicing at an early step in spliceosome assembly (Fig. 4A; refs. 93, 94). These compounds were originally isolated from bacteria (Pseudomonas sp. and Streptomyces sp.) and were shown to exhibit potent cytotoxic effects against various solid tumor cell lines in vitro and in xenograft models. More recently, forward genetic experiments performed in established resistant cell lines identified specific residues in SF3B1 and PHF5A, both components of the SF3B complex that were associated with drug resistance (95, 96). These findings affirmed the on-target specificity of SF3B-binding agents for the spliceosome and provided further clues to the exact structural basis for their mechanism of action. This culminated in publication of the crystal structure of SF3B in complex with pladienolide B and the cryo-EM structure of SF3B bound to E7107 (97, 98). These structures definitively show that SF3B inhibitors act by binding to the SF3B complex and interfering with the ability of the complex to recognize branchpoint nucleotides within introns. Extensive RNA-seq studies of cells treated with these compounds further revealed widespread inhibition of splicing efficiency, with dose-dependent increases in cassette exon skipping and IR (86, 99, 100).

Consistent with the studies above, which identified recurrent genetic subtypes of solid tumors that are particularly vulnerable to inhibition of splicing, hematopoietic cells bearing hotspot mutations in SRSF2, SF3B1, and U2AF1 have been shown to be preferentially sensitive to the SF3B-inhibitory compounds E7107, H3B-8800 (an orally bioavailable analogue of E7107), and sudemycins (analogues of the natural compound FR901464; refs. 37, 86, 99, 101). This has been demonstrated across numerous model systems including in vivo studies in splicing factor–mutant conditional knockin mice, acute and chronic myeloid leukemia patient-derived xenografts, and a variety of human splicing factor–mutant cancer cell lines. Owing to the large number of individual splicing events affected by chemical inhibition of SF3B, in addition to the aberrant splicing caused by each individual mutation, the precise mechanistic basis for preferential sensitivity of splicing factor–mutant cells to these compounds is not entirely clear. However, these agents induce aberrant splicing of mRNAs encoding numerous RNA-splicing factors. Given the preferential dependence of splicing factor–mutant cells on otherwise normal splicing catalysis, aberrant splicing of mRNAs encoding RNA-splicing factor proteins by splicing modulatory drugs may partially explain the preferential effects of these compounds on splicing factor–mutant cells (99). In addition, a number of studies have identified that the mRNAs encoding the expression of BCL2 family proteins are altered by SF3B inhibition. For example, SF3B inhibition promotes expression of proapoptotic isoforms of MCL1 as well as induction of nonfunctional alternatively spliced transcripts of BCL2 and BCL2A1 (102). As such, E7107 has been shown to induce apoptosis in cell lines normally dependent on these proteins for survival and has synergistic effects with BCL2 inhibition in some settings (103).
Although SF3B-inhibitory molecules provide a useful tool to investigate spliceosome assembly and function, the SF3B complex is essential in all cells, and the therapeutic index of these compounds has been a major question. The first SF3B binding agent to be tested in patients was E7107, where 66 patients with refractory metastatic solid tumors were treated in two prior phase I clinical trials (104, 105). Although some promising clinical responses were observed, clinical development of E7107 was halted due to the development of ocular toxicity in two patients, the cause of which was unclear. It is important to note that these trials were performed before the discovery that components of the spliceosome are recurrently mutated in various cancers, and there are no clinical reports to date of the effects of SF3B inhibition on the cancer types most enriched in these mutations. To this end, a phase I clinical trial of H3B-8800 is now ongoing in patients with AML, MDS, and chronic myelomonocytic leukemia that is relapsed/refractory to conventional therapy (clinicaltrials.gov identifier NCT02841540). This trial will present an important opportunity to evaluate the safety of H3B-8800 in patients and to determine the degree of splicing modulation achievable in patients.

**Splicing Inhibitor Sulfonamides: A New Class of Targeted Protein Degraders**

Until recently, few chemical means to perturb splicing were known outside of SF3B inhibitors. In 2017, Han and colleagues and investigators from Eisai discovered that a series of sulfonamide-containing compounds induce proteasomal degradation of the accessory RNA-splicing factor RBM39 (106, 107). These compounds, which include the molecules indisulam, E7820, and chloroquinoxaline sulfonamide, had been known for decades to have anticancer properties in vitro, but the cellular mechanism of action was not fully understood. However, a chemical genetic approach to identify indisulam-resistant cells (by Han and colleagues) and an expression proteomics effort (led by Uehara and colleagues) identified RBM39 as the cellular target of these compounds. This led to the discovery that the anticancer sulfonamides bind a substrate receptor of the CRL4 E3 ubiquitin ligase complex known as DCAF15.
and direct ubiquitin-mediated degradation of RBM39 (Fig. 4B). As such, deletion of DCAF15 renders cells resistant to these compounds, as does the introduction of RBM39 drug-resistant mutations (106, 107).

The discovery of molecules that direct the CRL4–DCAF15 E3 ubiquitin ligase complex to degrade RBM39 was surprising for several reasons. Before this discovery, the only known compounds known to induce targeted protein degradation were the IMiD molecules including lenalidomide (108–110). Although IMiDs normally engage an E3 ubiquitin ligase to target the substrates CRBN, IKZF1, and IKZF3 for proteasomal degradation, there are intense ongoing efforts to restructure these molecules in order to target additional proteins [so-called PROteolysis-TARGETing Chimera (PROTAC) strategies]. Thus, the identification of a novel set of chemically diverse compounds that act via an analogous mechanism opens up new possibilities to design small molecules that can target “undruggable” disease-causing proteins for proteasomal degradation.

In addition to the broad implications of hijacking anticancer sulfonamides to degrade diverse substrates, this discovery has also raised interest in the known substrate of these compounds, RBM39. RBM39 (also referred to as CAPERβ) is an RBP that contains classic RNA recognition motifs in addition to a unique class of protein recognition motifs known as U2AF homology motifs. These domains are also shared by the paralogous proteins RBM23 (CAPERα), PUF60, SPF45, and the well-characterized splicing factors U2AF1 and U2AF2. Although prior studies have suggested RBM39 physically interacts with the estrogen receptor AP1 and other transcription factors, the only rigorous structural evidence regarding RBM39 interactors supports an interaction of RBM39 with SF3B1 and U2AF2 (111, 112).

Consistent with the structural data above, RBM39 degradation in vitro results in clear inhibition of splicing catalysis, analogous to that seen with direct SF3B inhibition. As such, splicing factor–mutant leukemias have recently been shown to be as sensitive to RBM39 degradation as they are to SF3B inhibition (100). However, unlike SF3B inhibitors, whose clinical safety and potential are unclear, numerous anticancer sulfonamides have already completed phase I and II clinical trials. Indisulam, for example, has been used in phase II clinical trials in patients with melanoma, non–small cell lung cancer, and AML with favorable safety profiles (113–115). At the same time, each of these studies was carried out before the mechanism of action of these compounds was known. Thus, it is not known whether the doses of drug utilized in the clinical studies actually caused RBM39 degradation. Likewise, there was no effort in prior studies to stratify patient enrollment based on predictors of sensitivity to these compounds, which include DCAF15 expression levels and the presence of hotspot mutations in splicing factors (100, 106, 107). It will therefore be critical to revisit the clinical utility of these compounds with appropriate pharmacodynamic monitoring for RBM39 degradation and splicing inhibition, as well as stratification based on the aforementioned predictive biomarkers.

**Targeting Post-translational Modifications of Splicing Proteins**

In addition to inhibiting the interaction of the spliceosome with mRNA and degrading splicing proteins, there has been considerable interest in modulating splicing by modifying post-translational modifications (PTM) required for splicing function. It is known that PTMs of a variety of splicing factors regulate spliceosome formation and splicing catalysis. For example, phosphorylation of SR proteins regulates the shuttling of SR proteins in and out of the nucleus to modulate splicing (116–118). Phosphorylation of SR proteins is required for spliceosome complex formation, but dephosphorylation of SR proteins allows splicing catalysis to occur and initiates nuclear export of SR proteins. In another example, symmetric arginine dimethylation (SDMA), catalyzed by PRMT5, is required for snRNP assembly (119, 120). Consequently, a variety of chemical inhibitors of kinases and methyltransferases which act on splicing proteins, such as CDC-like kinases, SR protein kinases, and protein arginine methyltransferases (PRMT), among others, are being evaluated (Fig. 4C).

Of the above approaches, the class of compounds furthest in clinical development and most heavily studied is the PRMT inhibitors. Genetic ablation or chemical inhibition of PRMT5 results in splicing inhibition and anticancer effects across a number of cancer types (88, 121). Given these results, it will be important to determine whether malignancies with splicing-factor mutations are preferentially sensitive to PRMT5 inhibition, and whether inhibition of type I PRMT enzymes [which catalyze asymmetric arginine dimethylation (ADMA)] cause splicing changes similar to PRMT5 inhibition. Of note, consistent with the numerous cellular substrates for PRMT enzymes, many different mechanisms have been proposed to underlie the sensitivity of cancer cells to PRMT inhibitors beyond splicing. Moreover, given the number of splicing proteins and individual protein residues that undergo arginine methylation, discerning which individual arginine methylation events, if any, are responsible for the cellular effects of PRMT inhibitors has been a challenge.

The above mechanistic questions notwithstanding, at least three PRMT5 inhibitors are now in phase I clinical trials for patients with relapsed/refractory solid tumors. These include the compounds GSK3326595 (clinicaltrials.gov NCT02783300 for solid tumors and B-cell non-Hodgkin lymphomas; clinicaltrials.gov NCT03614728 for MDS/AML), PF 06939999 (clinicaltrials.gov NCT03854227), and JNJ-64619178 (clinicaltrials.gov NCT03573310). In addition, a first-in-human trial of the type I PRMT inhibitor GSK3368715 has been initiated for patients with relapsed/refractory B-cell non-Hodgkin lymphomas (clinicaltrials.gov NCT03666988). These studies will provide an important opportunity to evaluate the safety of PRMT inhibition in vivo at levels that affect SDMA/ADMA levels.

**Other Means to Target Splicing in Cancer**

Beyond the therapeutic approaches noted above, inhibition of other enzymatic steps in RNA processing is being pursued. For instance, the observation that N6-adenosine methylation of RNA is required for the survival of certain cancer types but dispensable for normal counterpart cells has led to great interest in identifying chemical inhibitors of the N6-adenosine methyltransferase METTL3 (reviewed recently in ref. 3). In addition, a recent functional genomic
screen by Yamauchi and colleagues showed that the decapping enzyme scavenger DCPS, which catalyzes the final step of 3' to 5' mRNA decay, is required for the survival of AML cells but dispensable for normal hematopoietic cells (122). Evidence for this preferentialdependency was bolstered by the description of families with germline DCPS biallelic loss-of-function mutations who have normal hematopoiesis. Interestingly, several inhibitors of DCPS have been described. One such compound, RG3039 (123), has already completed a phase I clinical trial for spinal muscular atrophy (SMA) and is known to be safe in patients. Given these results, evaluating the clinical potential of DCPS inhibition in patients with cancer will be an important next step.

In addition to chemical modulators of splicing, it is important to note the potential for oligonucleotide-based approaches to modify individual splicing events (Fig. 4D). These approaches consist of modified nucleic acids that base pair with pre-mRNA and modify splicing by blocking RNA–RNA base-pairing or splicing factor–RNA binding interactions. Antisense oligonucleotide (ASO)–based approaches have met clinical success, resulting in FDA approval of the ASOs nusinersen and eteplirsen for SMA and Duchenne muscular dystrophy, respectively. However, whether there will be therapeutic benefit in targeting individual splicing events in disorders such as cancer, which are marked by many simultaneous genetic and splicing alterations, remains to be determined.

The Potential for Splicing-Derived Neoepitopes

Checkpoint inhibitor–based immunotherapies have transformed clinical care for a number of malignancies. However, the majority of patients receiving checkpoint inhibitors do not derive benefit. This observation has spurred intense efforts to identify biomarkers of response to checkpoint blockade as well as pharmacologic approaches to increase response to these therapies. In this regard, a number of studies have shown that the burden of mutations occurring at the level of DNA increases the chances of generating immunogen peptides that can be presented on MHC class I. Interestingly, a number of studies have also suggested that cancer-specific changes to RNA splicing may be an additional source of neoepitopes. For example, the recent pan-cancer TCGA splicing analysis by Kahles and colleagues suggested that predicted neoepitopes generated by tumor-specific AS events are far more abundant than those generated by somatic SNVs (12).

Although this is an intriguing idea, to date there is little functional evidence that cancer-associated aberrant splicing actually elicits tumor immunogenicity. There has been little experimental validation of the immunogenicity of neoantigens derived from altered splicing. Evaluation of the immunogenicity of tumors based on RNA expression of immune markers has differed in their conclusions about whether cancer-associated splicing changes are associated with reduced or increased immune-cell infiltration (15, 31). Moreover, one recent analysis focused on neoepitopes generated from IR failed to identify any association between retained intron neoepitope load and clinical benefit from checkpoint inhibitors in patients with melanoma (124). One important point to note in studies of intron-retained mRNAs is that the fate of mRNAs with IR may not be immediately clear from RNA-seq alone. First, if the IR mRNA is exported to the cytoplasm and contains a premature termination codon, it may undergo NMD instead of being translated into a novel protein isoform. Alternatively, it has recently been recognized that some RNAs with retained introns may remain in the nucleus and be subjected to degradation by the nuclear exosome or undergo splicing post-transcriptionally and later exported to the cytoplasm for translation (125). mRNAs subject to splicing following transcription are referred to as “detained introns.” The mechanisms that detain incompletely spliced transcripts in the nucleus versus allowing their export to the cytoplasm are not entirely clear.

Given these results, further efforts to determine whether pharmacologic means to perturb splicing can induce the generation of neoepitopes and boost the immunogenicity of tumors would be innovative. Moreover, it is still unclear whether introduction of mutations in RNA-splicing factors might increase neoepitope load and/or responsiveness to immune-checkpoint blockade.

CONCLUSIONS

Despite the major advances in our understanding of the genomics, molecular biology, and therapeutic implications of altered RNA processing in cancer, the full contribution of aberrant RNA splicing and polyadenylation to cancer pathogenesis has not been fully elucidated. Further efforts to comprehensively catalog genetic alterations in the vast noncoding regions within genes (both within introns and at 3' UTRs) that have a regulatory impact on RNA splicing and PAS choice are still needed.

Moreover, greater systematic evaluation of the potential functional roles of distinct RNA isoforms, particularly aberrant, unannotated RNA isoforms produced in cancer cells, will be greatly informative for disease biology. Understanding the role of these isoforms could also be incredibly important for therapy. For example, it is currently recognized that acquired pathologic splice variants can confer therapeutic resistance to therapies as diverse as RAF inhibitors (126) and anti-CD19 chimeric antigen receptor T cells (127). These findings underscore the need to understand the dynamic role of RNA processing in untreated cancers as well as serially with therapy.

The identification of functionally important pathologic RNA isoforms opens up the possibility of oligonucleotide-based therapies for splicing alterations in cancer. In parallel, it is hopeful that the ongoing efforts to modulate splicing clinically may address the question of whether globally modifying RNA splicing is safe and therapeutically efficacious in patients. These include ongoing clinical trials to target the SF3B complex, degrade RBM39, and prevent protein arginine methylation of splicing proteins. Finally, the future application of novel methods of assessing splicing, including the use of splicing analysis within single cells and single-molecule RNA-seq technologies, will hopefully provide new insights into splicing dysregulation in cancer.

Disclosure of Potential Conflicts of Interest

O. Abdel-Wahab reports receiving a commercial research grant from H3 Biomedicine Inc., has received honoraria from the speakers bureau of Foundation Medicine Inc., and is a consultant/advisory
board member for Foundation Medicine Inc., H3 Biomedicine Inc., Janssen, and Merck. No potential conflicts of interest were disclosed by the other authors.

**Acknowledgments**

We thank members of the Obeng and Abdel-Wahab laboratories for their reading and input on this manuscript. E.A. Obeng is supported by grants from the EvansMDS Foundation, the Leukemia Research Foundation, and the American Society of Hematology. O. Abdel-Wahab is supported by grants from the EvansMDS Foundation, the Leukemia Research Foundation, the Henry & Marilyn Taub Foundation, NIH/NHLBI (R01 HL128239) and NIH/NCI (1 R01CA201247-01A1), and the Department of Defense Bone Marrow Failure Research Program (BM150092 and W81XWH-12-1-0041), an MSK Steven Greenberg Lymphoma Research Award, the Geoffrey Beene Research Center of MSK, and the Pershing Square Sohn Cancer Research Alliance.

Received April 4, 2019; revised June 21, 2019; accepted August 8, 2019; published first October 14, 2019.

**REFERENCES**


Splicing in Cancer


Altered RNA Processing in Cancer Pathogenesis and Therapy

Esther A. Obeng, Connor Stewart and Omar Abdel-Wahab

Cancer Discov 2019;9:1493-1510. Published OnlineFirst October 14, 2019.

Updated version
Access the most recent version of this article at:
doi:10.1158/2159-8290.CD-19-0399

Cited articles
This article cites 122 articles, 36 of which you can access for free at:
http://cancerdiscovery.aacrjournals.org/content/9/11/1493.full#ref-list-1

Citing articles
This article has been cited by 14 HighWire-hosted articles. Access the articles at:
http://cancerdiscovery.aacrjournals.org/content/9/11/1493.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://cancerdiscovery.aacrjournals.org/content/9/11/1493.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.