Alternative splicing, RNA-seq and drug discovery

Shanrong Zhao

Pfizer Worldwide Research and Development, Cambridge, MA 02139, USA

Alternative splicing, hereafter referred to as AS, is an essential component of gene expression regulation that contributes to the diversity of proteomes. Recent developments in RNA sequencing (RNA-seq) technologies, combined with the advent of computational tools, have enabled transcriptome-wide studies of AS at an unprecedented scale and resolution. RNA mis-splicing can cause human disease, and to target alternative splicing has led to the development of novel therapeutics. Splice variants diversify the repertoire of biomarkers and functionally contribute to drug resistance. Our expanding knowledge of AS variation in human populations holds great promise for improving disease diagnoses and ultimately patient care in the era of sequencing and precision medicine.

Alternative splicing and its role in biological diversity

Pre-mRNA splicing is a fundamental step in mRNA maturation, and its discovery in 1977 revolutionized our understanding of gene expression [1]. Not long after, scientists discovered that splicing patterns of a gene were not fixed; rather, alternative patterns of pre-mRNA splicing produced different mature mRNAs that contained various combinations of exons from a single precursor mRNA. This alternative splicing (AS) increases the diversity and functional capacity of a gene at the post-transcriptional level [2]. AS is widespread in higher eukaryotes. For instance, there are 58 037 annotated genes and 198 093 corresponding isoforms in Gencode Release 25 [3]. On average there are 3.4 annotated transcripts per human gene, a ratio that increases to 7:1 for protein-coding genes.

AS can have a profound effect on gene functions, and different isoforms from the same gene can be involved in distinct processes or even have directly opposing functions [4]. The p53 tumor suppressor gene, also known as tumor protein p53 (*TP53*), is an example of this. *TP53* has a central role in the regulation of DNA-damaged cells and is a well-studied gene for which multiple isoforms exist (Fig. 1). However, not all *TP53* isoforms have the same roles in tumor development; for instance, the Δ133β isoform can inhibit apoptosis of tumor cells that would normally be induced by the full-length p53β isoform [5,6]. Another example is CD45, which consists of multiple members that are all products of AS of a single complex gene [7]. CD45 isoforms are distinct in B and T lymphocytes and vary with the differentiation and activation state of a cell [8]. This gene contains 33 exons, and three exons of the primary transcripts can be alternatively spliced to generate up to eight different protein products [i.e., CD45RA, CD45RB, CD45RC, CD45RAB, CD45RAC, CD45RBC, CD45R0, CD45R (ABC)] (Fig. 2). Although naive T lymphocytes express large CD45 isoforms and are usually positive for CD45R, activated and memory T lymphocytes express the shortest CD45 isoform, CD45R0, which facilitates T cell activation. *TP53* and CD45 are two examples that illustrate how AS can have a profound effect on gene functions and cellular processes.

Common modes of AS and the mechanisms governing AS

Given the importance of AS, not surprisingly it is a highly regulated process [4]. The most common modes of AS are exon skipping, the use of mutually exclusive exons [9], alternative donor (5′ splice) sites, alternative acceptor (3′ splice) sites or intron retention (Fig. 3). In addition, the use of alternative promoters or polyadenylation sites can also further increase mRNA diversity [10–14]. Regulation of alternative splicing is a complicated process in which numerous interacting components are at work, including

E-mail address: Shanrong.Zhao@pfizer.com.

https://doi.org/10.1016/j.drudis.2019.03.030

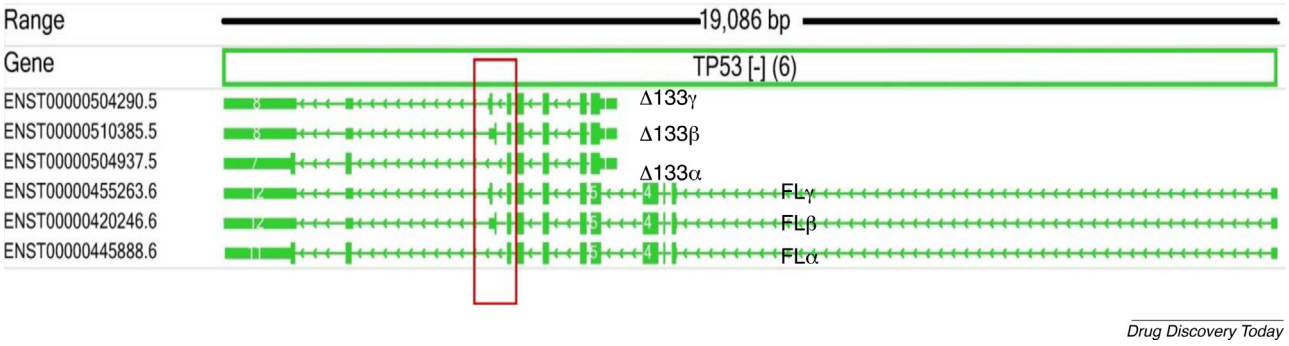

FIGURE 1

Representative isoforms of *TP53*. Different isoforms can be involved in distinct processes or even have directly opposing functions. For instance, the Δ133β isoform inhibits apoptosis of tumor cells that would normally be induced by the full-length FLβ isoform.

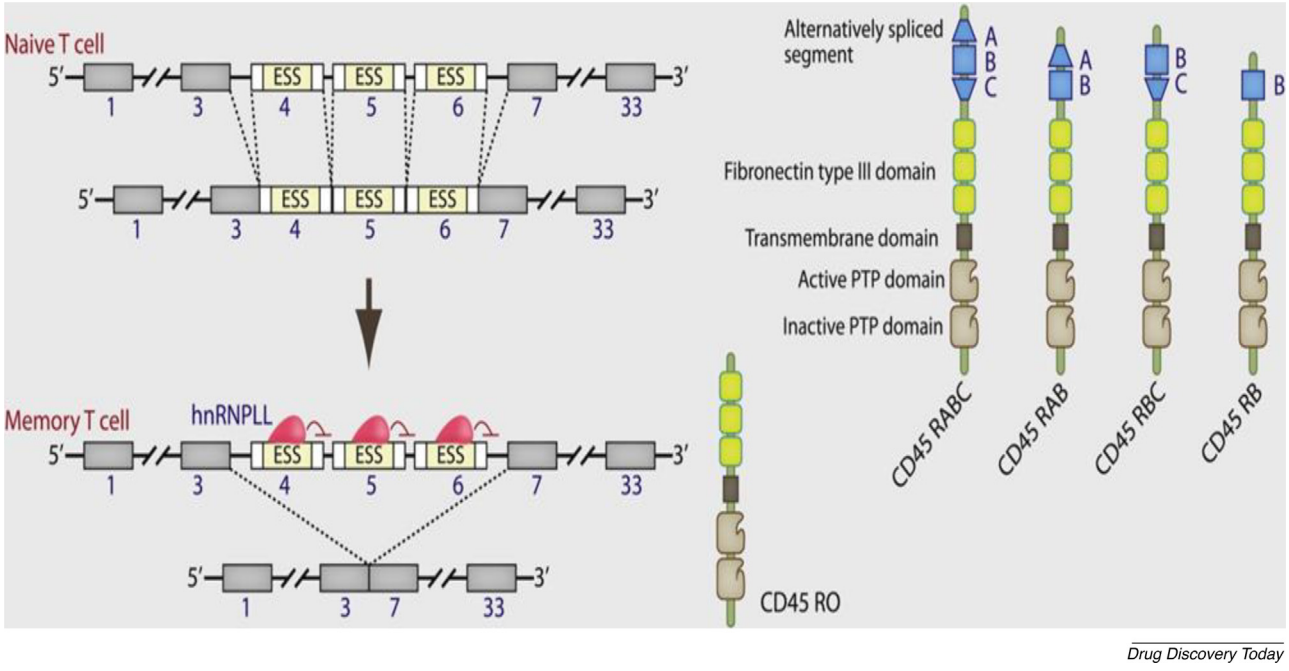

FIGURE 2

Alternative splicing (AS) isoforms of CD45. Adapted, with permission from [7]. AS dictates the inclusion or exclusion pattern of three exons in CD45 (exons 4, 5 and 6), corresponding to segments A, B and C. Differentiation of naive T cells into memory T cells is characterized by the exclusion of all three exons, leading to the short CD45RO isoform by virtue of hnRNPLL binding to exonic splicing silencer (ESS) sequences. Naive T cells express various amounts of different high molecular weight isoforms, such as RABC, RAB, RBC and RB, depending on AS patterns of these exons.

cis-acting elements and *trans*-acting factors, and it is further guided by the functional coupling between transcription and splicing [15–19] (Fig. 4). The most essential *cis*-splicing signals within the pre-mRNA are the 5′ splice site (5′SS), 3′ splice site (3′SS) and branch site (A). The 5′ and 3′ splice sites have highly conserved GU and AG dinucleotides as the first and last two nucleotides of the intron, respectively. Alternative RNA splicing is catalyzed by a large and highly dynamic protein complex called the spliceosome [20]. The spliceosome complex is composed of five small nuclear ribonucleic acids (snRNA U1, U2, U5, U5 and U6) and ~200 protein components. Among the associated 200 proteins, there are two well-studied RNA-splicing-factor families: the heterogeneous nuclear ribonucleoprotein (hnRNP) proteins and the SR (serine–arginine rich) splicing factors [21]. However, it is worth mentioning that not all splicing regulators are hnRNP or SR proteins, for example NOVA, TDP-43, RBFOX1 and RBFOX2 [22]. The U1 small nuclear ribonucleoprotein particle (snRNP) complex recognizes the 5′SS, and the U2 snRNP complex recognizes the branch site. The U2AF (U2 snRNP auxiliary factor) proteins recognize the 3′SS. Splice-site utilization is further regulated by exonic splicing enhancers (ESEs) or silencers (ESSs), or intronic splicing enhancers (ISEs) or silencers (ISSs) within the pre-mRNA. In general, silencers are usually recognized by hnRNP proteins, whereas enhancers recruit SR proteins. However, the function of the *cis* and *trans* regulators of AS could change depending on the particular gene, exon and intron sizes, cellular context, and the developmental or physiological state of the cell. The final decision to include or splice an alternative exon is thus determined by combinatorial effects including competitive binding between hnRNP inhibitors and SR activators.

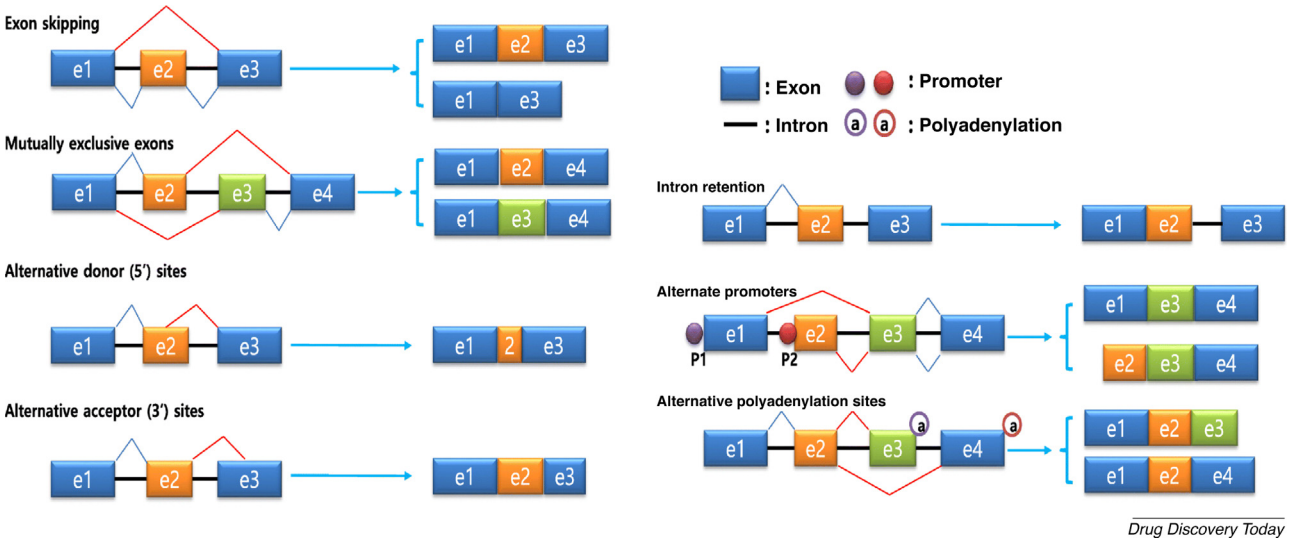

FIGURE 3

Common modes of alternative splicing (AS) of pre-mRNA [4]. The most common modes of AS are exon skipping or the use of mutually exclusive exons, alternative donor (5′ splice) sites or alternative acceptor (3′ splice) sites, or intron retention. In addition, the use of alternative promoters or polyadenylation sites further increases mRNA diversity.

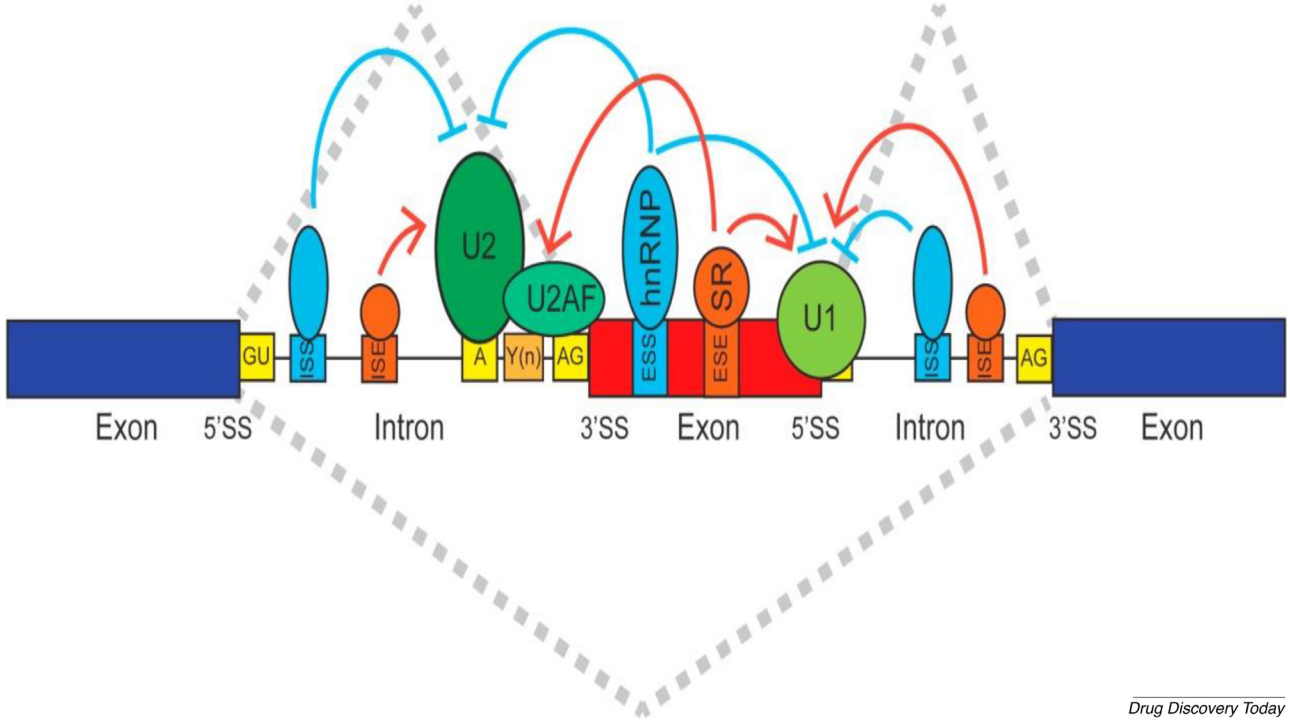

FIGURE 4

Complex interplay between *cis*- and *trans*-acting factors regulate alternative splicing (AS) [80]. The 5′SS is recognized by the U1 snRNP complex, the branch-point is recognized by U2 snRNP complex and the 3′SS (another feature, the polypyrimidine tract) is recognized by U2AF proteins. This constitutes the 'core' splicing machinery. Splice site utilization is further regulated by exonic splicing enhancers (ESEs) or silencers (ESSs), or intronic splicing enhancers (ISEs) or silencers (ISSs) within the pre-mRNA.

RNA-seq and computational algorithms for characterization of AS

Until recently, the conventional molecular biology approach to the characterization of AS was reverse transcription polymerase chain reaction (RT-PCR). Then, in the late 1990s, sequencing of expressed sequence tags (ESTs), which are fragments of full-length mRNAs, revealed widespread AS in eukaryotic organisms [23]. In the mid-2000s, microarrays were designed to sample AS events on a genome scale and successfully used to examine AS across tissues, cellular states and species [24,25]. However, these technologies have low-throughput (RT-PCR and ESTs), high noise (ESTs and splicing microarray) or are limited in their ability to characterize only known splicing events (RT-PCR and splicing microarray).

Recent developments in HTS technologies and computational tools have revolutionized our capacity to investigate AS at the genome-wide level [26–31]. Four landmark papers published in 2008 demonstrated the power of RNA-seq for characterizing AS in mammalian tissues [2,16,30,32]. Computational tools to analyze alternatively spliced transcripts can be divided into two categories based on their functionality. Tools in the first category include RSEM [33], Kallisto [34] and Salmon [35], which can be applied to analyze known or annotated transcript isoforms. Tools in the second category include MISO [17], MAJIQ [19], rMATs [36] and LeafCut [37], which can be used to analyze RNA-seq data at the exon level to detect known and novel splicing events. Owing to different definitions on AS events, different software applied to the same datasets often come with different predictions; sometimes the overlap of the software predictions can be very small [38]. A widely used metric for AS is percent spliced [37] (PSI or ψ), which represents the percentage of a gene's mRNA transcripts that include a specific exon or splice site. For a given AS event, the PSI value can be calculated from the number of RNA-seq reads supporting specific exons or splice junctions. However, if the sequencing depth is not deep enough or the expression levels for a particular gene and its isoforms are very low, the calculated ψ is less reliable. Although RNA-seq is indeed now the preferred technology to monitor genome-wide AS, RT-PCR remains the gold standard to validate novel splicing variants and confirm changes in PSI. This is especially true when the depth of reads is not sufficient for high-confidence assessment.

Although RNA-seq has become the *de facto* standard in transcriptome profiling [26,29], despite the advent of computational tools to monitor AS, most RNA-seq studies to date have focused on only gene-level analyses. As a result, subtle differences in alternative isoform usage and exon inclusion and/or exclusion are missed in these gene-level analyses. To demonstrate how additional scientific insights gained from the isoform- and exon-level analysis of RNA-seq data, Zhang *et al.* recently explored AS patterns in kidney tissue in a preclinical model of progressive diabetic nephropathy [27]. Our analyses identified AS patterns in genes that can be implicated in disease pathogenesis, such as *Shc1*, *Serpinc1*, *Epb4.1l5* and *Il-33*, which would have been overlooked by standard gene-level analysis [27]. Their findings indicate that additional mechanistic insights can be gained through interrogation of AS in addition to conventional gene-level analysis of RNA-seq data.

AS isoforms in health and disease

Indeed, because >90% of human protein-coding genes are alternatively spliced [2], mis-splicing can modify or abrogate important physiological protein function in homeostasis [4]. Increasing evidence has linked AS to human health and disease. Many factors can cause AS disruptions, and thus contribute to disease. These include documented mutations that affect hnRNPs, SR proteins, splice sites, secondary structures of introns or pre-mRNAs, or extended-repeat elements that guide regulatory protein interactions [4,39–41]. Any mutation affecting canonical splice sites in a gene can clearly lead to gene dysfunction and potentially disease. However, often overlooked are mutations that affect splicing enhancers, silencers or *trans*-acting factors, because they can appear to be silent synonymous changes with no effect on amino acid sequence, or to be otherwise innocuous owing to their intronic or distal location. Ever-increasing numbers of these mutations are now being identified in patients with genetic diseases. Indeed, according to some estimates, 50–60% of disease-causing mutations in the human gene mutation database affect splicing [42].

Accumulated data highlight the importance of AS in disparate human diseases, including neurodegenerative disorders, cancer, immune and infectious diseases, cardiovascular and metabolic diseases [4,39,40,42]. For example, AS caused by altered expression levels of the splicing factor serine–arginine splicing factor 1 (SRSF1) has been shown to contribute to tumor progression [43]. Using RNA-seq, Zhou *et al.* [43] comprehensively screened and identified multiple SRSF1-affected AS events. Functionally, they verified that SRSF1 promoted cell proliferation, survival and invasion by specifically switching the AS of myosin IB (*MYO1B*) gene toward the expression of a membrane-localized and oncogenic isoform: MYO1B-fl. Other examples of how AS can affect normal cellular functions and cause abnormal cell growth further highlight the close relationship between AS and cancer biology [44–47]. Aberrant splicing events in cancer-associated genes, namely *BCL2L1, FAS, HRAS, Cyclin D1, CASP2, TMPRSS2-ERG, FGFR2, VEGF, AR* and *KLF6*, were summarized in [47] and are depicted in Fig. 5. The pre-mRNA of B cell lymphoma 2-like 1 (*BCL2L1*) gene can be alternatively spliced to encode two protein isoforms, antiapoptotic Bcl-xL (long isoform) and proapoptotic Bcl-xS (short isoform) [48]. High Bcl-xL:Bcl-xS mRNA ratios, associated with greater tumor cell survival, can be found in several cancer types, such as breast cancer [49]. The caspase-2 (*CASP-2*) gene produces multiple mRNA splice variants. *CASP-2L* encodes a full-length caspase-2 protein, which is expressed in most tissues and promotes apoptosis. However, splice variant *CASP-2S*, resulting from an exon 9 retention event that leads to a premature termination, encodes a truncated protein lacking the active domain and thereby inhibits cell apoptosis [50]. Previous studies have shown that CASP-2S overexpression promotes antiapoptotic activities and protects cells from cytotoxicity by chemotherapeutic agents, such as etoposide in leukemic and lymphoma cells [51]. For other hallmarks of alternative splicing in cancer, please refer to [47,52,53]. *KLF6* and *VEGF*, two genes emerging with potential druggability, are discussed below.

AS in drug development

Because abnormal AS can be associated with disease specificity and severity, strategies to modulate and monitor this process hold promise in therapeutic and biomarker settings [40,54,55]. This section provides some highlights of how AS is emerging as an important aspect in the field of drug discovery.

Modulating AS of a target gene

Over the past two decades, molecular tools have been developed to correct or redirect AS events [54,55]. A typical target-specific approach can be achieved through the use of antisense oligonucleotides (ASO) designed as a complementary molecule that targets a specific mRNA to regulate AS. By designing the ASO to bind and mask *cis*-regulatory splicing sequences, the splicing machinery can be redirected in a precise and often reproducible way. This strategy has been used with notable success with the recent FDA approval of nusinersen [56,57], an ASO to correct

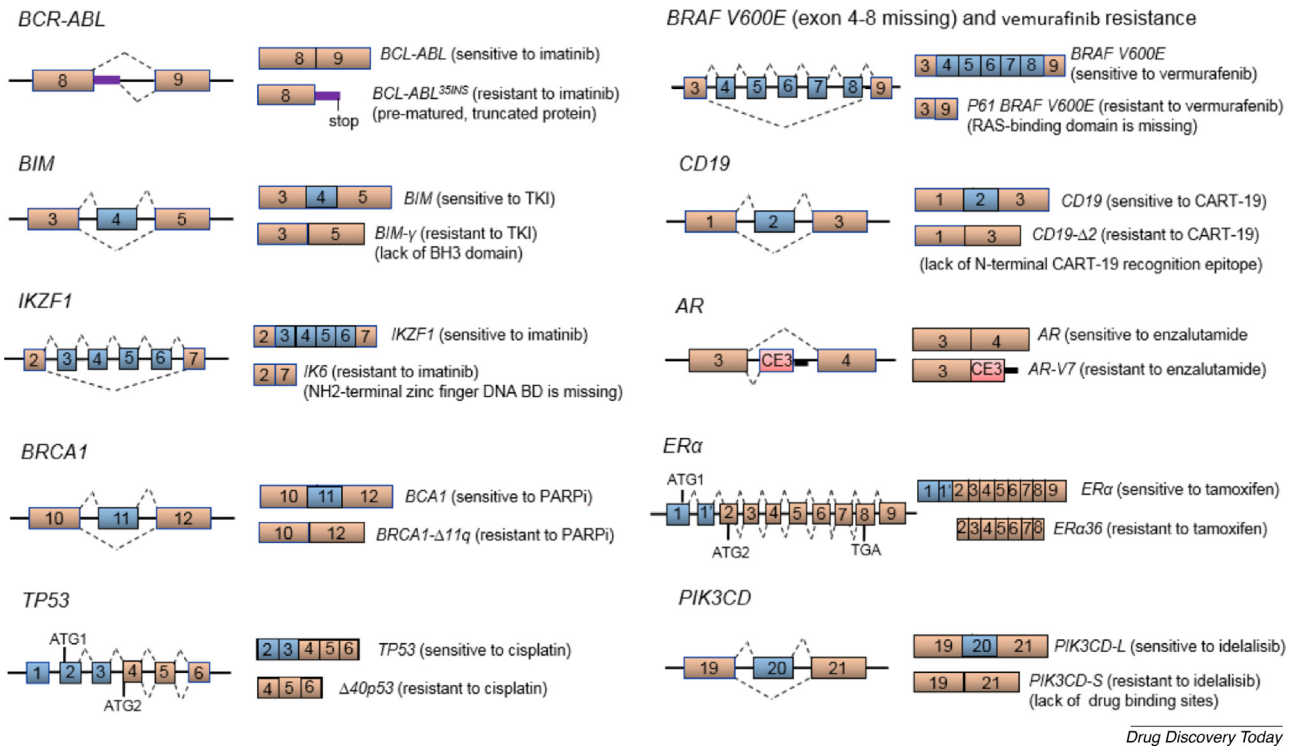

FIGURE 5

Examples of aberrant splice variants and oncogenic consequences for *Bcl-xL, FAS-S, p21H-RAS, Cyclin D1b, CASP-2S, TMPRSS2-ERG+72bp, FGFR2-IIIc, VEGF-165, AR-V7* and *KLF6-SV1*. Abbreviations: *TMPRSS2–ERG+72bp*, TMPRSS2–ERG fusion transcript with inclusion of a 72-bp exon; BCa, breast cancer; PCa, prostate cancer; CE3, cryptic exon 3. The pre-mRNAs are shown on the left panel, and the mature mRNA variants (only the exons surrounding the differential splicing event are shown) following AS are shown on the right panel. Reproduced, with permission, from [47].

splicing in spinal muscular atrophy (SMA). SMA is the second-most-common recessive disorder in humans, and the most common inherited cause of infant mortality [57]. In humans, a gene duplication event of survival of motor neuron 1 (*SMN1*) has given rise to an almost identical gene called *SMN2*. In SMA, *SMN1*, which produces the majority of full-length SMN protein, is either deleted or inactivated by mutations. However, *SMN2* contains a silent C→T substitution in the sixth nucleotide of exon 7 (Fig. 6). This substitution destroys an ESE by abolishing a binding site for the SR protein SF2/ASF, and also creates an exonic splicing silencer through creation of a binding site for hnRNPA1. As a result, this substitution causes skipping of exon 7 and ineffective protein production, with the result that *SMN2* is unable to compensate for any loss in expression of *SMN1*. To prevent *SMN2* exon 7 skipping, and restore functional protein production, the ASO nusinersen blocks an intronic splicing silencer element for the treatment of type 1 (severe) SMA infants and children (Fig. 6).

Another example is ataluren, an ASO that has been approved for treatment of Duchenne muscular dystrophy (DMD) [58]. DMD is a progressive and life-threatening X-linked recessive disorder caused by mutations in the *DMD* gene that results in reduced or absent production of dystrophin – a protein involved in muscle fiber formation. Dystrophin is the largest known human gene, spanning 2.3 Mb and containing 79 exons. DMD in males is typically caused by a nonsense mutation in an exon that introduces a premature termination codon (PTC). The resulting loss of functional dystrophin causes progressive and irreversible muscle wasting in these individuals, leading to disability and premature death. Approximately 250 000 people suffer with this disorder worldwide [59]. Several recent clinical trials have demonstrated that ASO-based drugs that promote the skipping of the PTC-containing exon can yield an mRNA that is resistant to nonsense-mediated mRNA decay and encodes a dystrophin protein that is not fully normal but which retains important functions [58]. Such exon skipping therapy has the potential to convert the lethal Duchenne phenotype into the less severe Becker phenotype [60]. Recently, ataluren was marketed under the trade name Translarna® by PTC Therapeutics.

Targeting the AS machinery with small molecules

The realization that AS isoforms and altered AS machinery in cancer cells are not merely byproducts of the oncogenic state but rather potential drivers of oncogenesis raises the possibility of targeting AS for cancer therapy. Indeed, one of the earliest small-molecule modulators of splicing to be identified was FR901464 and its acetylated derivate spliceostatin A (SSA) [61]. This molecule was shown to have antitumor effects *in vitro* and *in vivo*; and was later discovered to bind and inhibit the U2-snRNP component SF3B1 [62]. Finally, small-molecule inhibitors of histone acetyltransferases (HATs) and histone deacetylases (HDACs) could influence AS via stalling of spliceosome assembly, in addition to their more traditional effects on transcriptional regulation [63].

Nusinersen is the first FDA-approved ASO drug for all types of SMA but efficient delivery remains a major challenge. Orally bioavailable small molecules that specifically target the *SMN2* gene and modulate its splicing are under development for the treatment of SMA, and are currently entering early-phase clinical trials [64]. Oral administration makes these drugs more tolerable than ASOs,

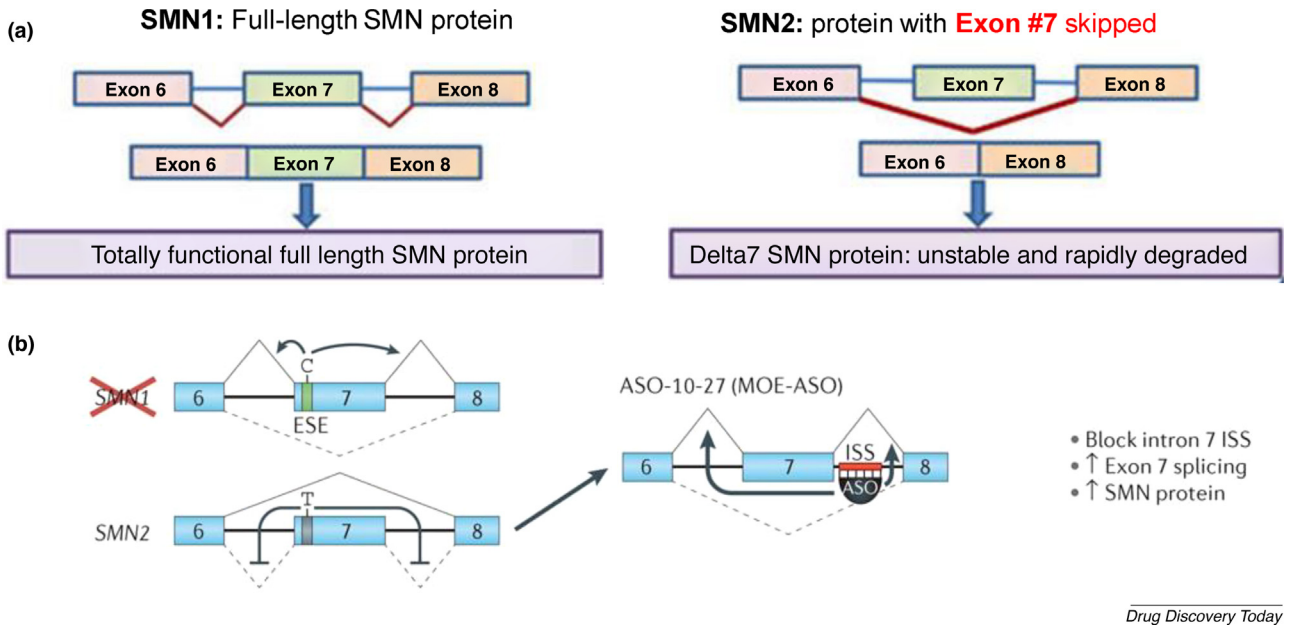

FIGURE 6

Schematic representation of the action mechanism of nusinersen (Spinraza®) in treatment of spinal muscular atrophy (SMA). Survival of motor neuron 1 (SMN1), which produces the majority of full-length SMN protein, is either deleted or inactivated by mutations in SMA patients, whereas the paralogous *SMN2* expresses low levels of SMN protein with exon #7 skipped owing to a C→T transition (gray box) in the exonic splicing enhancer (ESE) region. Nusinersen targets an intronic splicing silencer (ISS; red bar) region and enhances exon 7 splicing to produce full-length SMN protein.

because it overcomes the need for repeated lumbar puncture. The major concern related to this approach is the risk of off-target effects given that small molecules, unlike ASOs, are not completely specific and can, in principle, affect the expression of other genes. To overcome these issues, Roche used a high-throughput *in vitro* screening to identify small molecules RG7800 and RG7916 as potential candidates [64]. For more examples of targeting AS with small molecules, the reader is referred to two excellent, recent reviews on the topic [55,65].

Targeting specific AS protein isoforms

Targeting an individual protein isoform can be the basis of novel therapeutic strategies based on AS biology. The clearest examples of this come from the field of biotherapeutics. For example, the fibronectin (FN) gene can undergo AS that results in the inclusion or exclusion of extra domain A (EDA) exon [66]. Because extracellular EDA-FN is enriched in areas of tissue damage and angiogenesis, monoclonal antibodies that target the EDA sequence are currently a pharmaco-delivery strategy to deliver anti-inflammatory payloads to sites of disease [67]. In another example, several companies have attempted to target a cancer-associated AS form of CD44, CD44v6, to deliver chemotherapeutic agents to tumors, albeit with limited success owing to adverse events in late-stage clinical trials [68].

At time of writing and to the best of my knowledge, there is currently no example of an FDA-approved splicing-isoform-selective small-molecule inhibitor. There is clearly, however, future potential in this regard. The idea of development of drugs to target specific protein isoforms is supported by the validation of the isoform-specific oncogenic role of some cancer genes such as Kruppel-like factor 6 (KLF6) and vascular endothelial growth factor (VEGF). KLF6 encodes a member of the Kruppel-like family of transcription factors. Multiple transcript variants encoding different proteins have been found for this gene, some of which are implicated in carcinogenesis. Full-length KLF6 is a tumor suppressor that displays loss of expression, whereas the splice variant 1: KLF6-SV1 (Fig. 5), plays a crucial part in promoting cell proliferation, survival, migration and angiogenesis of prostate cancer [69,70]. Moreover, KLF6-SV1 is overexpressed in metastatic prostate cancer and associated with increased metastasis, and thus targeted inhibition of KLF6-SV1 can suppress prostate cancer cell growth and spread [71]. VEGF is associated with angiogenesis that plays a key part in cancer progression and development of metastases. Typical splicing of VEGF results in an antiangiogenic splice variant VEGF165b (Fig. 5), which is widely expressed in normal cells and tissues but is downregulated in prostate cancer. This suggests that targeting specific VEGF isoforms could be an important regulatory mechanism of angiogenesis in cancer [72].

Utilizing AS for predictive and prognostic biomarkers

AS, and changes in AS, can occur in a disease-state- and stage-specific manner, indicating that AS patterns could serve as a diagnostic or prognostic biomarker [73–75]. Traditional genetic biomarkers present a great challenge for the diagnosis of complex diseases such as cancer, because many of these diseases arise from multiple distinct molecular mechanisms. Most cancer-specific changes in splicing disrupt protein domains and increase cell proliferation or cell survival, consistent with a functional role for AS in cancer. The exon-specific detection of AS might serve as a reliable biomarker and provide a novel approach to diagnose and monitor cancer progression [76,77]. For instance, several AS variants and cancer-specific splicing events in colorectal cancer have been shown to serve as biomarkers of disease [73]. Given the high frequency of mRNA splicing in cancers, linking this feature to drug response opens new avenues of research in biomarker discovery [75]. To identify robust transcriptomic biomarkers for

drug response across studies, Safikhani *et al.* developed a meta-analytical framework combining the pharmacological data from two large-scale drug-screening datasets. They used an independent pan-cancer pharmacogenomic dataset to test the robustness of their candidate biomarkers across multiple cancer types. They further analyzed two independent breast cancer datasets and found that specific isoforms of *IGF2BP2*, *NECTIN4*, *ITGB6* and *KLHDC9* were significantly associated with AZD6244, lapatinib, erlotinib and paclitaxel, respectively [75]. Their results support isoform expression as a rich resource for biomarkers predictive of drug response.

In addition, two recent studies demonstrated that AS profiles can predict cancer patient survival comparable to, and often with better accuracy than, total gene expression levels [78,79]. Given that AS is quantified as the relative ratio of multiple isoforms from a single gene, AS data are self-normalized on a per-gene basis and can therefore be viewed as having an 'internal control'. By this measure, AS can potentially provide a more intrinsic molecular signature of molecular alterations than total gene expression levels. This is especially true for large-scale RNA-seq studies in clinical trials, where the variability of samples is large because they are usually collected from multiple patients and sites and across many different time points [80].

AS considerations in drug resistance

One of the major challenges in cancer treatment is that tumors that initially respond well to therapy eventually develop resistance to drugs, resulting in treatment failure. Development of resistance and relapse of tumors are major challenges in cancer therapy. Therefore, understanding resistance mechanisms is key to improving therapeutic outcomes. One mechanism of drug resistance can occur through AS [81]. Because AS takes place post-transcriptionally, changes in isoform ratios or generation of novel isoforms that can escape drug action can occur rapidly. In addition, exon inclusion or skipping can alter a complete protein domain, rather than a single amino acid change that commonly occurs owing to genomic mutations. Thus, cancer cells can use this intrinsic mechanism to gain drug resistance without altering the genome. AS is an important determinant of clinical response, and one well-studied gene with drug resistance is androgen receptor (AR) [47,82,83]. AR plays an important part in cancer pathogenesis and might prove to be a relevant therapeutic target for prostate cancer patients [84]. The full-length AR (AR-FL) is composed of an N-terminal domain, a central DNA-binding domain, a hinge region and a C-terminal ligand-binding domain (LBD) (Fig. 7a); *AR-V7*, one of the best characterized AR splice variants, is missing exons 4–7 and is prominently expressed in hormone-refractory prostate cancer. The *AR-V7* variant encodes a functional protein lacking the LBD, a crucial target for androgen ablation therapy. Consequently, patients with *AR-V7*-expressing prostate cancer exhibit resistance to the antiandrogen drug enzalutamide (AR antagonist).

Another example of drug resistance is provided by the breast cancer type 1 susceptibility gene (*BRCA1*) (Fig. 7b). *BRCA1* encodes tumor suppressor proteins that are required for homologous recombination (HR)-mediated repair of double-strand (ds)DNA breaks. In the absence of either BRCA1 or BRCA2 activity, poly (ADP-ribose) polymerase (PARP), which functions as an enzyme to repair single-strand (ss)DNA breaks through base-excision repair, is thought to be an essential safeguard for maintaining genome integrity. Consequently, PARP inhibitors (PARPi) could serve as an effective agent to induce 'synthetic lethality' of tumor cells with dysfunctional BRCA1 and/or BRCA2. Although PARPi therapy has been demonstrated to efficiently sensitize BRCA1/2 mutation-associated cancers and improve survival in patients, not all patients respond to therapy and some patients developed drug resistance after an initial favorable response [85]. Cancer cell lines and tumors harboring mutations in exon 11 of BRCA1 express a *BRCA1-Δ11q*

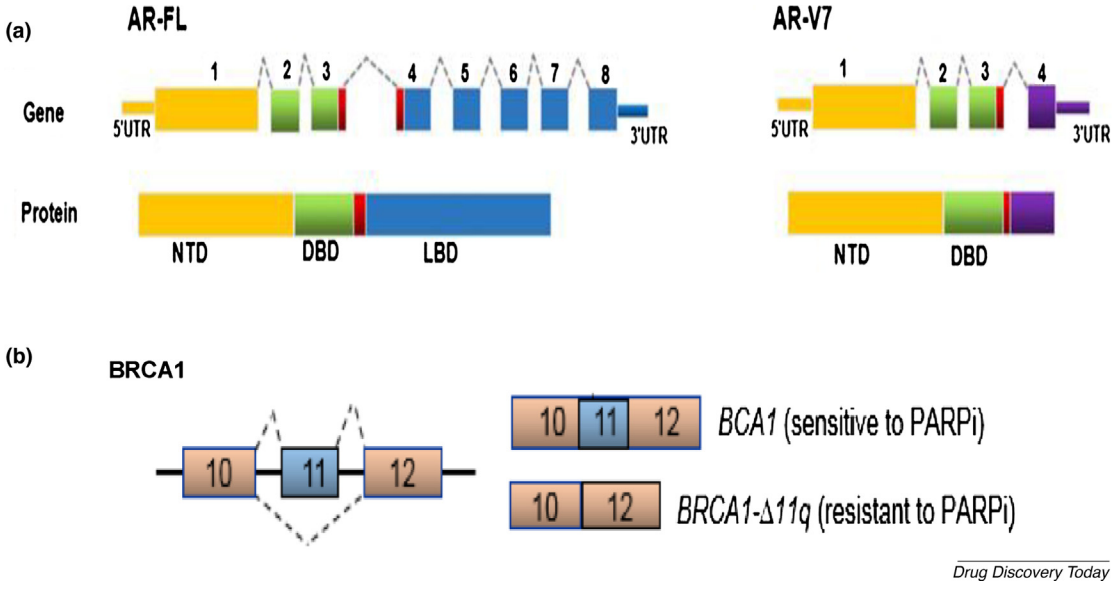

FIGURE 7

Aberrant mRNA splicing variants in drug resistance. **(a)** Androgen receptor (AR). The full-length AR (AR-FL) has a C-terminal ligand-binding domain (LBD), whereas AR-V7 is missing exons 4–7 and encodes a functional protein lacking the LBD, a crucial target for androgen ablation therapy. **(b)** The *BRCA1-Δ11q* splice variant skips exon 11 and consequently bypasses any inactivating mutations in this region, and thus promotes partial resistance to PARPi therapy.

splice variant lacking the majority of exon 11. A recent study has revealed that the *BRCA1-Δ11q* splice variant bypasses any inactivating mutations in this region, and thus promotes partial resistance to PARPi therapy [86]. Breast cancer cells employ a strategy to remove deleterious germline *BRCA1* mutations through alternative mRNA splicing, giving rise to isoforms that retain residual activity and contribute to therapeutic resistance.

Poulikos *et al.* [87] identified a new resistance mechanism in which RAF inhibitor resistance was mediated by dimerization of aberrantly spliced BRAF(V600E). They found that a subset of cells resistant to vemurafenib (PLX4032, RG7204) expressed a 61 kDa variant form of BRAF(V600E): p61BRAF(V600E), which lacks exons 4–8, a region that encompasses the RAS-binding domain. Compared to full-length BRAF(V600E), p61BRAF(V600E) shows enhanced dimerization. Poulikos *et al.* [87] identified BRAF(V600E) splicing variants lacking the RAS-binding domain in the tumors of six of 19 patients with acquired resistance to vemurafenib. Moreover, a mutation that abolishes the dimerization of p61BRAF (V600E) restores its sensitivity to vemurafenib. These data support a novel mechanism of acquired resistance in patients: expression of splicing isoforms of BRAF(V600E) that dimerize in a RAS-independent manner.

AS can also impact cell therapy approaches. For example, the CD19 antigen, expressed on most B cell acute lymphoblastic leukemias (B-ALL), can be targeted with chimeric antigen-receptor-armed T cells (CART-19) but relapses occur in 10–20% of pediatric responders [88]. One mechanism by which this occurs is where CD19 is still present but is no longer recognized by anti-CD19 CAR-T cells because the cell-surface fragment containing the cognate epitope is absent because of AS. Sotillo *et al.* discovered an alternatively spliced CD19 isoform that skipped exon 2 that interferes with CART-19 recognition of the antigen, preventing killing [88]. This mechanism of drug resistance suggests a possibility of targeting alternative CD19 ectodomains to improve survival of patients with B cell neoplasms.

Utilizing AS in precision medicine

The prevalent role of AS in Mendelian and complex diseases suggests that evaluating the impact of genomic variants on splicing should be adopted as an integral part of clinical variant prioritization. Population-scale transcriptome studies have revealed many naturally occurring genetic variants that modulate AS, and consequently influence phenotypic variability and disease susceptibility in human populations [80]. To identify biomarkers that predict drug response, Safikhani *et al.* combined pharmacological data from two large-scale drug screening datasets with genome-wide transcriptomics to identify isoform-level expression patterns that were predictive of drug response *in vitro* [75]. They validated four isoform-based biomarkers (IGF2BP2, NECTIN4, ITGB6 and KLHDC9) that were predictive of drug response to lapatinib, erlotinib, AZD6244 (MEK inhibitor) and paclitaxel, indicating that isoform analysis constitutes a promising new class of biomarkers for patient stratification for cytotoxic and targeted anticancer therapies.

In another example, Wang *et al.* explored the RNA splicing landscape of prostate cancer (PCa) across racial populations to define a molecular mechanism that could contribute to race-related tumor aggressiveness and drug resistance [89]. They identified novel race-specific RNA splicing events as crucial drivers of PCa aggressiveness and therapeutic resistance in African American (AA) men, and identified AA-enriched splice variants of *PIK3CD*, *FGFR3*, *TSC2* and *RASGRP2* with greater oncogenic potential than corresponding European American (EA)-expressed variants. Ectopic overexpression of the newly cloned AA-enriched variant *PIK3CD-S* (*PIK3CD* short, missing exon 20) in EA PCa cell lines increased proliferative and invasive capacity *in vitro* and conferred resistance to the sPI3Kδ inhibitor CAL-101 (idelalisib) *in vivo*. In addition, high *PIK3CD-S* expression in PCa patient specimens was associated with poor survival outcomes. A comprehensive catalogue of AS variation in human populations, along with the ability to discover and characterize splicing-altering variants in specific individuals, holds great promise for improving disease diagnoses and, ultimately, patient care in the era of precision medicine.

Concluding remarks and future perspectives

In recent decades our knowledge of the mechanisms and dynamics of AS has grown exponentially, and it is increasingly appreciated that AS plays a key part in human health and disease. With the emergence of RNA-seq and computational tools that enable examination of AS at a genome-wide level, there will be increasing insights and opportunities to leverage AS to deliver effective medicines to patients. Traditional functional annotations for proteins are recorded at the gene level but, more and more, isoform-centric functional annotations should become the mainstream. Furthermore, functional network analysis of the spliceosomal machinery using knockdown and other approaches will lead to new insights into how variations in core spliceosomal components influence differentiation, development and disease. The heterogeneity of AS in individual cells, and novel RNA splicing errors, should emerge as a result of an enhanced understanding of the human transcriptome, in particular with advances in single-cell technologies [90,91]. On the horizon, we anticipate that there will be breakthroughs in a long-standing interest in developing *in silico* methods to predict AS and its effects. Complementary machine-learning approaches that focus on analyzing sequence variants in disease will accelerate our understanding of the 'splicing code' [92,93]. Finally, clinical trials using splicing modulatory strategies and targeting specific isoforms have produced some encouraging results for several diseases, and these approaches should be applicable to additional disorders caused by mis-splicing.

Acknowledgments

I am grateful for Ken Dower for his critical reading and editing of the manuscript, and to Laird Bloom for helpful suggestions.

References

1 Chow, L.T. *et al.* (1977) An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* 12, 1–8

2 Pan, Q. *et al.* (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415

3 Harrow, J. *et al.* (2012) GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res.* 22, 1760–1774
4 Kim, H.K. *et al.* (2018) Alternative splicing isoforms in health and disease. *Pflugers Arch.* 470, 995–1016
5 Aoubala, M. *et al.* (2011) p53 directly transactivates Delta133p53alpha, regulating cell fate outcome in response to DNA damage. *Cell Death Differ.* 18, 248–258
6 Kim, S. and An, S.S. (2016) Role of p53 isoforms and aggregations in cancer. *Medicine* 95, e3993
7 Zikherman, J. and Weiss, A. (2008) Alternative splicing of CD45: the tip of the iceberg. *Immunity* 29, 839–841
8 Oberdoerffer, S. *et al.* (2008) Regulation of CD45 alternative splicing by heterogeneous ribonucleoprotein, hnRNPLL. *Science* 321, 686–691
9 Hatje, K. *et al.* (2017) The landscape of human mutually exclusive splicing. *Mol. Syst. Biol.* 13, 959
10 Elkon, R. *et al.* (2013) Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.* 14, 496–506
11 Tian, B. and Manley, J.L. (2017) Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.* 18, 18–30
12 Sanfilippo, P. *et al.* (2017) Genome-wide profiling of the 3′ ends of polyadenylated RNAs. *Methods* 126, 86–94
13 Ayoubi, T.A. and Van De Ven, W.J. (1996) Regulation of gene expression by alternative promoters. *FASEB J.* 10, 453–460
14 Reyes, A. and Huber, W. (2018) Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Res.* 46, 582–592
15 Chen, M. and Manley, J.L. (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol* 10, 741–754
16 Wang, E.T. *et al.* (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476
17 Katz, Y. *et al.* (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* 7, 1009–1015
18 Wang, Y. *et al.* (2015) Mechanism of alternative splicing and its regulation. *Biomed. Rep.* 3, 152–158
19 Vaquero-Garcia, J. *et al.* (2016) A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife* 5, e11752
20 van der Feltz, C. *et al.* (2012) Architecture of the spliceosome. *Biochemistry* 51, 3321–3333
21 Patel, S.B. and Bellini, M. (2008) The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res.* 36, 6482–6493
22 Zhou, H. *et al.* (2014) RNA-binding proteins in neurological diseases. *Sci. China Life Sci.* 57, 432–444
23 Modrek, B. and Lee, C. (2002) A genomic view of alternative splicing. *Nat. Genet.* 30, 13–19
24 Lee, C. and Roy, M. (2004) Analysis of alternative splicing with microarrays: successes and challenges. *Genome Biol.* 5, 231
25 Castle, J.C. *et al.* (2008) Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat. Genet.* 40, 1416–1425
26 Zhao, S. *et al.* (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* 9, e78644
27 Zhang, C. *et al.* (2018) Computational identification and validation of alternative splicing in ZSF1 rat RNA-seq data, a preclinical model for type 2 diabetic nephropathy. *Sci. Rep.* 8, 7624
28 Zhang, C. *et al.* (2017) Evaluation and comparison of computational tools for RNA-seq isoform quantification. *BMC Genomics* 18, 583
29 Wang, Z. *et al.* (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63
30 Mortazavi, A. *et al.* (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628
31 Byron, S.A. *et al.* (2016) Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat. Rev. Genet.* 17, 257–271
32 Sultan, M. *et al.* (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321, 956–960
33 Li, B. and Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* 12, 323
34 Bray, N.L. *et al.* (2016) Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527
35 Patro, R. *et al.* (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419
36 Shen, S. *et al.* (2014) rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U. S. A.* 111, E5593–5601
37 Li, Y.I. *et al.* (2018) Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* 50, 151–158
38 Liu, R. *et al.* (2014) Comparisons of computational methods for differential alternative splicing detection using RNA-seq in plant systems. *BMC Bioinf.* 15, 364
39 Tazi, J. *et al.* (2009) Alternative splicing and disease. *Biochim. Biophys. Acta* 1792, 14–26
40 Scotti, M.M. and Swanson, M.S. (2016) RNA mis-splicing in disease. *Nat. Rev. Genet.* 17, 19–32
41 Dlamini, Z. *et al.* (2017) Abnormalities in alternative splicing in diabetes: therapeutic targets. *J. Mol. Endocrinol.* 59, R93–107
42 Wang, G.S. and Cooper, T.A. (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.* 8, 749–761
43 Zhou, X. *et al.* (2019) Splicing factor SRSF1 promotes gliomagenesis via oncogenic splice-switching of MYO1B. *J. Clin. Invest.* 129, 676–693
44 Sveen, A. *et al.* (2016) Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene* 35, 2413–2427
45 Pal, S. *et al.* (2012) Alternative transcription and alternative splicing in cancer. *Pharmacol. Ther.* 136, 283–294
46 Wiesner, T. *et al.* (2015) Alternative transcription initiation leads to expression of a novel ALK isoform in cancer. *Nature* 526, 453–457
47 Wang, B.D. and Lee, N.H. (2018) Aberrant RNA splicing in cancer and drug resistance. *Cancers (Basel)* 10, E458
48 Akgul, C. *et al.* (2004) Alternative splicing of Bcl-2-related genes: functional consequences and potential therapeutic applications. *Cell Mol. Life Sci.* 61, 2189–2199
49 Olopade, O.I. *et al.* (1997) Overexpression of BCL-x protein in primary breast cancer is associated with high tumor grade and nodal metastases. *Cancer J. Sci. Am.* 3, 230–237
50 Solier, S. *et al.* (2005) Nonsense-mediated mRNA decay among human caspases: the caspase-2S putative protein is encoded by an extremely short-lived mRNA. *Cell Death Differ.* 12, 687–689
51 Parent, N. *et al.* (2005) Procaspase-2S inhibits procaspase-3 processing and activation, preventing ROCK-1-mediated apoptotic blebbing and body formation in human B lymphoma Namalwa cells. *Apoptosis* 10, 313–322
52 Oltean, S. and Bates, D.O. (2014) Hallmarks of alternative splicing in cancer. *Oncogene* 33, 5311–5318
53 Ladomery, M. (2013) Aberrant alternative splicing is another hallmark of cancer. *Int. J. Cell Biol.* 2013, 463786
54 Sune-Pou, M. *et al.* (2017) Targeting splicing in the treatment of human disease. *Genes* 8, 87
55 Martinez-Montiel, N. *et al.* (2018) Alternative splicing as a target for cancer treatment. *Int. J. Mol. Sci.* 19, 545
56 Hoy, S.M. (2017) Nusinersen: first global approval. *Drugs* 77, 473–479
57 Talbot, K. and Tizzano, E.F. (2017) The clinical landscape for SMA in a new therapeutic era. *Gene Ther.* 24, 529–533
58 Kole, R. and Leppert, B.J. (2012) Targeting mRNA splicing as a potential treatment for Duchenne muscular dystrophy. *Discov. Med.* 14, 59–69
59 Wood, C.L. and Cheetham, T. (2017) Treatment of Duchenne muscular dystrophy: first small steps. *Lancet* 390, 1467–1468
60 Nakamura, A. (2017) Moving towards successful exon-skipping therapy for Duchenne muscular dystrophy. *J. Hum. Genet.* 62, 871–876
61 Nakajima, H. *et al.* (1996) New antitumor substances, FR901463, FR901464 and FR901465. II. Activities against experimental tumors in mice and mechanism of action. *J. Antibiot.* 49, 1204–1211
62 Kaida, D. *et al.* (2007) Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat. Chem. Biol.* 3, 576–583
63 Kuhn, A.N. *et al.* (2009) Stalling of spliceosome assembly at distinct stages by small-molecule inhibitors of protein acetylation and deacetylation. *RNA* 15, 153–175
64 Parente, V. and Corti, S. (2018) Advances in spinal muscular atrophy therapeutics. *Ther. Adv. Neurol. Disord.* 11, 1756285618754501
65 Salton, M. and Misteli, T. (2016) Small molecule modulators of pre-mRNA splicing in cancer therapy. *Trends Mol. Med.* 22, 28–37
66 White, E.S. *et al.* (2008) New insights into form and function of fibronectin splice variants. *J. Pathol.* 216, 1–14
67 Schwager, K. *et al.* (2009) Preclinical characterization of DEKAVIL (F8-IL10), a novel clinical-stage immunocytokine which inhibits the progression of collagen-induced arthritis. *Arthritis Res. Ther.* 11 R142
68 Sahin, I.H. and Klostergaard, J. (2015) CD44 as a drug delivery target in human cancers: where are we now? *Expert Opin. Ther. Targets* 19, 1587–1591
69 DiFeo, A. *et al.* (2009) The role of KLF6 and its splice variants in cancer therapy. *Drug Resist. Updat.* 12, 1–7
70 DiFeo, A. *et al.* (2006) Roles of KLF6 and KLF6-SV1 in ovarian cancer progression and intraperitoneal dissemination. *Clin. Cancer Res.* 12, 3730–3739

71 Narla, G. *et al.* (2005) Targeted inhibition of the KLF6 splice variant, KLF6 SV1, suppresses prostate cancer cell growth and spread. *Cancer Res.* 65, 5761–5768
72 Biselli-Chicote, P.M. *et al.* (2012) VEGF gene alternative splicing: pro- and anti-angiogenic isoforms in cancer. *J. Cancer Res. Clin. Oncol.* 138, 363–370
73 Yi, Q. and Tang, L. (2011) Alternative spliced variants as biomarkers of colorectal cancer. *Curr. Drug Metab.* 12, 966–974
74 Le, K.Q. *et al.* (2015) Alternative splicing as a biomarker and potential target for drug discovery. *Acta Pharmacol. Sin.* 36, 1212–1218
75 Safikhani, Z. *et al.* (2017) Gene isoforms as expression-based biomarkers predictive of drug response *in vitro*. *Nat. Commun.* 8, 1126
76 Klinck, R. *et al.* (2008) Multiple alternative splicing markers for ovarian cancer. *Cancer Res.* 68, 657–663
77 Venables, J.P. *et al.* (2008) Identification of alternative splicing markers for breast cancer. *Cancer Res.* 68, 9525–9531
78 Trincado, J.L. *et al.* (2016) The prognostic potential of alternative transcript isoforms across human tumors. *Genome Med.* 8, 85
79 Shen, S. *et al.* (2016) SURVIV for survival analysis of mRNA isoform variation. *Nat. Commun.* 7, 11548
80 Park, E. *et al.* (2018) The expanding landscape of alternative splicing variation in human populations. *Am. J. Hum. Genet.* 102, 11–26
81 Siegfried, Z. and Karni, R. (2018) The role of alternative splicing in cancer drug resistance. *Curr. Opin. Genet. Dev.* 48, 16–21
82 Gupta, K. *et al.* (2016) The utility of chemotherapy in the treatment of metastatic prostate cancer. *Anticancer Agents Med. Chem.* 16, 1166–1171
83 Di, C. *et al.* (2018) Function, clinical application, and strategies of pre-mRNA splicing in cancer. *Cell Death Differ.* . http://dx.doi.org/10.1038/s41418-018-0231-3
84 Wadosky, K.M. and Koochekpour, S. (2017) Androgen receptor splice variants and prostate cancer: from bench to bedside. *Oncotarget* 8, 18550–18576
85 Kim, Y. *et al.* (2017) Reverse the resistance to PARP inhibitors. *Int. J. Biol. Sci.* 13, 198–208
86 Wang, Y. *et al.* (2016) The BRCA1-Delta11q alternative splice isoform bypasses germline mutations and promotes therapeutic resistance to PARP inhibition and cisplatin. *Cancer Res.* 76, 2778–2790
87 Poulikakos, P.I. *et al.* (2011) RAF inhibitor resistance is mediated by dimerization of aberrantly spliced BRAF(V600E). *Nature* 480, 387–390
88 Sotillo, E. *et al.* (2015) Convergence of acquired mutations and alternative splicing of CD19 enables resistance to CART-19 immunotherapy. *Cancer Discov.* 5, 1282–1295
89 Wang, B.D. *et al.* (2017) Alternative splicing promotes tumour aggressiveness and drug resistance in African American prostate cancer. *Nat. Commun.* 8, 15921
90 Huang, Y. and Sanguinetti, G. (2017) BRIE: transcriptome-wide splicing quantification in single cells. *Genome Biol.* 18, 123
91 Hayashi, T. *et al.* (2018) Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. *Nat. Commun.* 9, 619
92 Jha, A. *et al.* (2017) Integrative deep models for alternative splicing. *Bioinformatics* 33, i274–282
93 Leung, M.K. *et al.* (2014) Deep learning of the tissue-regulated splicing code. *Bioinformatics* 30, i121–129