nature reviews genetics

https://doi.org/10.1038/s41576-022-00556-8

Review article

Check for updates

# Regulation of pre-mRNA splicing: roles in physiology and disease, and therapeutic prospects

Malgorzata Ewa Rogalska [1,5], Claudia Vivori [1,2,4,5] & Juan Valcárcel [1,2,3]

## Abstract

The removal of introns from mRNA precursors and its regulation by alternative splicing are key for eukaryotic gene expression and cellular function, as evidenced by the numerous pathologies induced or modified by splicing alterations. Major recent advances have been made in understanding the structures and functions of the splicing machinery, in the description and classification of physiological and pathological isoforms and in the development of the first therapies for genetic diseases based on modulation of splicing. Here, we review this progress and discuss important remaining challenges, including predicting splice sites from genomic sequences, understanding the variety of molecular mechanisms and logic of splicing regulation, and harnessing this knowledge for probing gene function and disease aetiology and for the design of novel therapeutic approaches.

## Sections



[1]Genome Biology Program, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain. [2]Department of Medicine and Life Sciences, Universitat Pompeu Fabra (UPF), Barcelona, Spain. [3]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. [4]Present address: The Francis Crick Institute, London, UK. [5]These authors contributed equally: Malgorzata Ewa Rogalska, Claudia Vivori. e-mail: juan.valcarcel@crg.eu

## Introduction

Most primary RNA transcripts produced by eukaryotic RNA polymerase II (RNA Pol II) contain introns that need to be removed by the process of splicing (Box 1, see the figure), to generate functional mRNAs and long non-coding RNAs. Although many aspects of the origin and function of introns remain uncertain (Box 1), they likely originated from autocatalytic RNAs that spread through the genomes of primitive eukaryotes as retrotranscribed DNA transposable elements[1]. During evolution, the strict sequence and 3D structure required for self-removal of intronic RNAs were progressively relaxed as their excision became dependent upon an increasingly complex cellular machinery known as the spliceosome (Fig. 1; see Supplementary Tables 1–3). The function of the spliceosome depends on the recognition of intronic boundaries by small nuclear ribonucleoprotein (snRNP) complexes, followed by a series of conformational transitions that involve remodelling of numerous RNA–RNA, RNA–protein and protein–protein interactions to enable protein-assisted, RNA-based catalysis of intron removal[2–5]. A general principle of spliceosome assembly is that intron boundaries are recognized multiple times, ensuring accuracy in the splicing process[4,6]. Two classes of introns and spliceosomes coexist in complex organisms, the minor class being spliced with slower kinetics but being nevertheless essential for the expression of genes involved in multiple processes, including early development[7].

Variations in the patterns of intron removal (known as alternative splicing) occur in the majority of genes in multicellular organisms[8–10], contributing to proteome diversification as well as to the regulation of gene expression by the degradation of transcripts containing premature termination codons (Box 2). There are numerous examples of alternative splicing events that are important for cell identity, pluripotency and organismal physiology, or that contribute to various pathologies (reviewed in refs. 11–15). How widespread is the functional relevance of alternative splicing, however, remains an open question (Box 2).

Owing to recent developments in the methods, software and data sets available to study alternative splicing (Supplementary Box 1), now is a particularly exciting time for studies of RNA splicing and its regulation. For the first time, detailed cryogenic electron microscopy structures of the spliceosome at various steps of its assembly and catalysis have been determined at unprecedented resolution (Fig. 1; see Supplementary Table 4), providing a structural framework to interpret decades of previous biochemical and genetic studies[2–5]. There is also an unprecedented wealth of transcriptome data showing the large diversity of transcript isoforms, with profound implications for understanding basic biology and disease outcomes. In parallel, strong evidence has accumulated indicating that mutations in factors involved in post-transcriptional regulation contribute to cancer and neurodegeneration[16]. Finally, the success of splicing-modulating therapies for the treatment of spinal muscular atrophy[17], as well as unexpected applications of splicing inhibitors as cancer therapies or for the maintenance of cell totipotency[18,19], bring hope that our increasing understanding of splicing mechanisms will provide a new generation of therapeutics. Alternatively spliced isoforms should also be given careful consideration in the design of mRNA-based therapeutics[20].

This Review focuses on the most recent advances in our understanding of the molecular mechanisms that help distinguish between introns and exons to enable accurate splicing and to regulate alternative splicing. These include the interplay between components of the splicing machinery that recognize the 5′ splice site (5′ss) and 3′ splice site (3′ss), their coupling with the process of transcription, the role of RNA structures and the contribution of RNA modifications. In higher eukaryotes, accurate splicing is a formidable challenge because, whereas intron removal requires single-nucleotide precision to preserve coding information in exons, the sequences delineating exon–intron boundaries are remarkably diverse, often resembling other sequences that are not splice sites[21] (Figs. 2 and 3). Testifying to the high level of specificity required, mutations in pre-mRNAs or in splicing factors lead to alterations in the splicing process that are associated with human pathologies ranging from genetic diseases to neurodegeneration or cancer[13,14,18]. Here, we provide examples of the clinical impact of splicing alterations, and discuss how a deeper understanding of the regulatory mechanisms of splicing can help design therapies that counteract these detrimental effects of splicing alterations or even exacerbate them to facilitate immune responses against tumour cells[22].

## 5′ splice site recognition

Initial recognition of the 5′ss of an intron is carried out by the U1 snRNP complex of the spliceosome through base-pairing interactions involving the 5′ end of U1 small nuclear RNA (snRNA) (Figs. 1 and 2). Given the variability of 5′ss sequences, base pairing is key to defining the efficiency (or 'strength') with which a particular 5′ss is used[23]. 5′ss mutations that weaken base pairing with U1 snRNA can cause defective splice site recognition and disease. For example, a single mutation at position +6 of intron 20 of the *IKBKAP* gene causes exon skipping, introducing a premature termination codon that reduces expression of functional protein[24] (Fig. 4a). This leads to the autosomal recessive condition familial dysautonomia, a neurodegenerative disorder that often causes premature death owing to cardio-respiratory arrest[25]. Conversely, mutations in the 5′ end sequence of one of the multiple gene copies of U1 snRNA, which are observed in patients with chronic lymphocytic leukaemia and sonic hedgehog medulloblastoma, can result in the recognition of novel 5′ss that have base-pairing complementarity with the mutated U1 snRNA, and more generally alter patterns of splicing of multiple genes. These genes include known cancer drivers – for example, resulting in inactivation of tumour suppressor genes or activation of proto-oncogenes – and correlate with worse disease prognosis[26,27].

However, it is striking that most 5′ss that are efficiently recognized do not form perfect base-pairing interactions with the 5′ end of U1 snRNA, even after considering non-canonical base-pairing schemes (such as bulged or other unpaired nucleotides[28–30]). It is also striking that mutations of the 5′ss that are not predicted to significantly disrupt base pairing with U1 snRNA nevertheless alter 5′ss recognition, being associated with diseases such as Fanconi anaemia, haemophilia, neurofibromatosis and phenylketonuria[29]. Massively parallel splicing assays assessing all of the 32,768 possible 5′ss sequences – NNN/GYNNNN (N = any nucleotide, maintaining at position +1 the G required for catalysis; Y = pyrimidine at position +2, present in 99.6% of 5′ss) – in three different minigene contexts confirmed the relevance of base pairing between U1 snRNA and the 5′ss for splicing efficiency but also revealed marked context-related differences[31]. This suggests that additional nearby sequences, their cognate factors and interactions between these and U1 snRNP components can aid the efficient use of 5′ss that have suboptimal base pairing with U1 snRNA[32–34]; for example, such additional sequences and factors have been shown to have a role in alternative splicing that regulates the physiological shift in energy metabolism from glycolysis to oxidative phosphorylation[35]. Systematic approaches (such as those described in ref. 31) can help predict the pathogenic effects of 5′ss mutations or of natural sequence variation. For example, they verified that nearly 90% of the 5′ss mutations found in *BRCA2* in breast cancer samples do affect splicing, potentially facilitating genetic screening[31].

### Therapeutic targeting

Blocking the inhibitory effect of an intronic splicing silencer (ISS) on the recognition of a 5′ss using an antisense oligonucleotide known as nusinersen (approved for clinical use in 2016) has provided a major therapeutic breakthrough for patients with spinal muscular atrophy, which is a leading genetic cause of infant mortality[17,36–38] (Fig. 4a). Spinal muscular atrophy is caused by mutations that inactivate *SMN1*, which encodes a protein important for snRNP assembly. For reasons that remain unclear, loss of SMN1 function mainly affects the function of motor neurons, leading to progressive muscle weakness and in the most severe cases to death within the first 2 years of life[38]. Quarterly intrathecal injection of nusinersen increases the levels of protein generated from a second gene, *SMN2*, which under normal circumstances fails to produce functional protein owing to limited inclusion of exon 7 (refs. 17,39).

Facilitated by the early success of nusinersen, an orally available, small-molecule, pyrido-pyrimidinone drug known as risdiplam that has similar effects on *SMN2* exon 7 inclusion obtained US Food and Drug Administration (FDA) approval in 2020 and has shown promising clinical results[40,41]. One mode of action proposed for risdiplam analogues is to 'repair' the bulge formed by the lack of base pairing between U1 snRNA and the last nucleotide of *SMN2* exon 7, thus stabilizing U1 snRNP recruitment[42] (Figs. 2b and 4a). This is achieved, at least in part, by facilitating an interaction between the zinc finger of the U1 snRNP protein U1C and the minor groove of the U1–5′ss helix[42], although other mechanisms may also be involved[43]. It is quite remarkable that a compound that modulates structural features of particular U1–5′ss configurations has therapeutic properties, paving the way to a new generation of compounds targeting 5′ss recognition.

Interestingly, engineering U1 snRNA such that its 5′ end can base pair to intronic sequences downstream of a 5′ss activates the use of the bona fide upstream 5′ss (refs. 44,45) (Fig. 4a). This suggests that an increase in the local concentration of and/or cooperativity between U1 snRNP complexes can enhance 5′ss recognition, perhaps by propagating complexes that change the physical behaviour of the exon. This approach has been used in vitro to correct exon skipping events associated with various pathologies including spinal muscular atrophy, cystic fibrosis or neurological disorders such as CDKL5-deficiency disorder[46–48] (Fig. 4a). However, the oncogenic properties of U1 snRNA mutations mentioned above[26,27] bring a note of caution when considering the general applicability of this approach.

Although not directly related to 5′ss recognition, another major function of the U1 snRNP complex is to bind to 3′ untranslated regions of mRNAs and inhibit the use of proximal 3′ end formation sites, which are often used in actively proliferating cells, including cancer cells[49,50]. Consistent with this, inhibition of U1 snRNP using antisense morpholino oligonucleotides increases cancer cell migration and invasion, whereas increased levels of U1 snRNP inhibit these phenotypes[51].

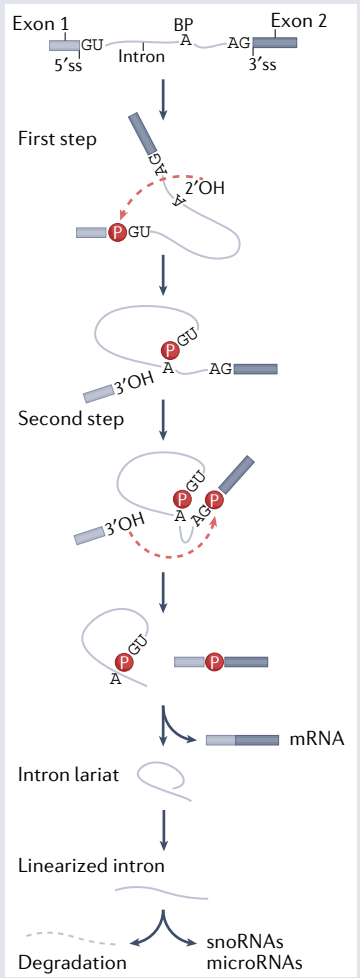
Similar approaches have been proposed for the therapeutic correction of 5′ss recognition in familial dysautonomia (Fig. 4a). These include antisense oligonucleotides targeting ISSs downstream from the 5′ss of IKBKAP exon 20 (refs. 52,53), modified U1 snRNAs[54] and small molecules such as kinetin, a plant cytokinin that enhances the recognition of 5′ss flanked by a particular sequence motif[55,56], or RECTAS, which enhances the phosphorylation of SRSF6, a splicing regulatory factor that functions through an intronic splicing enhancer (ISE) located in intron 20 (refs. 57,58).

Other examples of 5′ss recognition that have significant therapeutic potential for modulation include the induction of a pro-apoptotic

## Box 1

# The function of introns

Introns are internal sequences that are removed from precursor mRNA transcripts by a two-step splicing process (see the figure). The first step involves cleavage of the phosphodiester bond between the upstream exon (exon 1) and the intron, and concomitant formation of a 2′–5′ phosphodiester bond between the 5′ guanosine and an internal adenosine (the branch point (BP)), generating a lariat intermediate. The second step involves cleavage of the phosphodiester bond between the 3′ end of the intron and the downstream exon (exon 2), concomitant with ligation of the two exons and release of the intron in a lariat configuration.




Despite great progress in understanding the splicing process and its regulation, important fundamental questions remain. One such question is whether intronic sequences can have functions of their own or are simply by-products of a process designed to eliminate ancient transposon insertions. Recent work has shown that particular introns in the yeast *Saccharomyces cerevisiae* accumulate as linear RNA species under various stress conditions, with their accumulation contributing to stress responses through a regulatory network that involves the target of rapamycin complex (TORC), a key integrator of growth signalling[295]. Indeed, systematic deletion of introns in *S. cerevisiae* led to an impaired response to starvation, which was linked to the function of intronic sequences as repressors of ribosomal protein synthesis[296]. These observations are consistent with the long-standing proposal that introns might provide, via autonomous functions, an additional layer of genetic information to that provided by mature mRNAs and long non-coding RNAs[297]. It remains unclear, however, whether the functions documented above for certain yeast introns, as well as other examples of non-coding RNAs located within introns (such as small nucleolar RNAs (snoRNAs)[298] and microRNAs), are important exceptions to a general lack of function for introns.

3′ss, 3′ splice site; 5′ss, 5′ splice site.

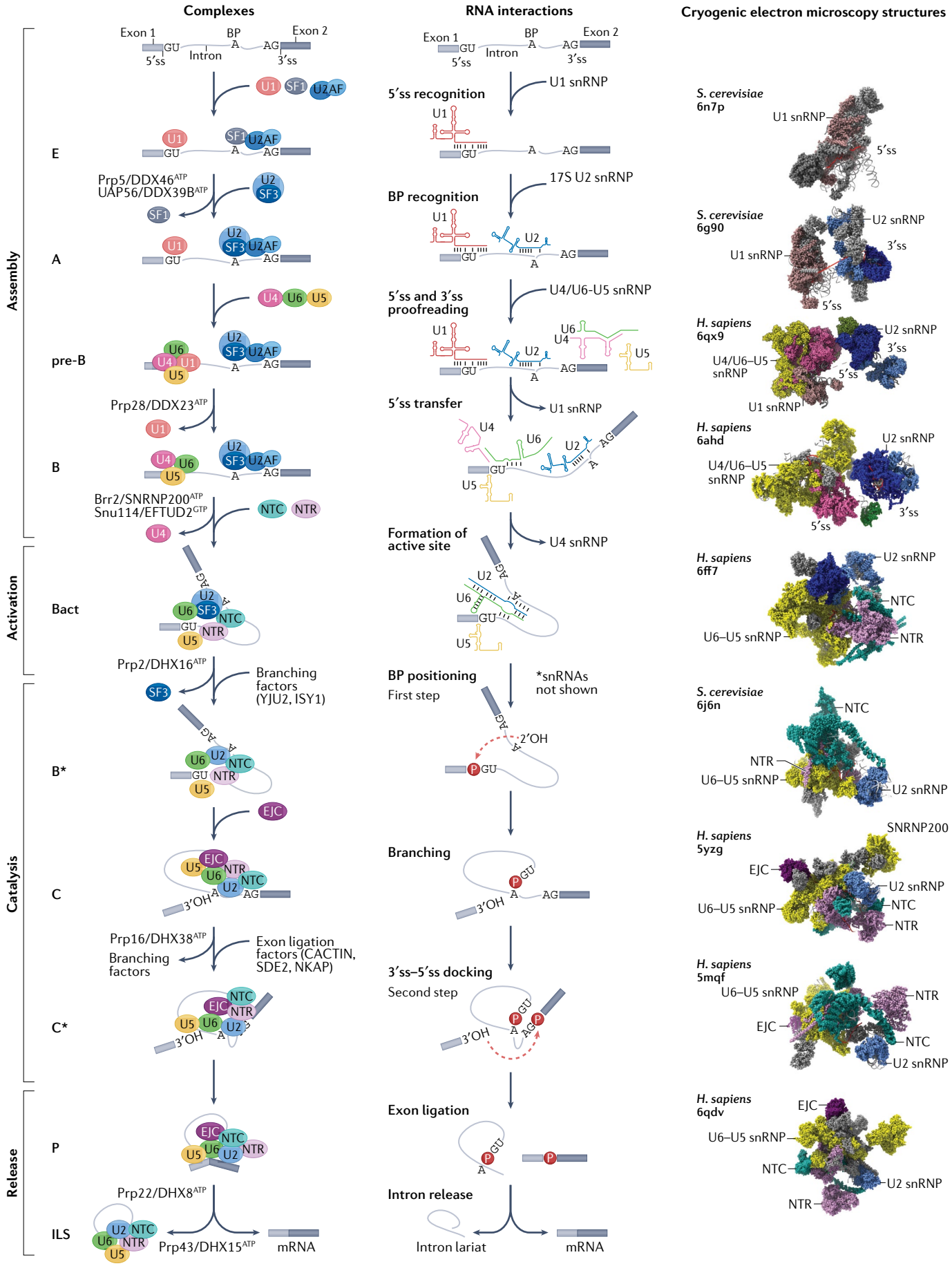

Complexes
RNA interactions
Cryogenic electron microscopy structures
Assembly
Activation
Catalysis
Release
Exon 1
BP
Exon 2
GU
Intron
AG
5′ss
3′ss
E
A
pre-B
B
Bact
B*
C
C*
P
ILS
U1
SF1
U2AF
U2
SF3
U4
U6
U5
NTC
NTR
EJC
Prp5/DDX46ATP
UAP56/DDX39BATP
Prp28/DDX23ATP
Brr2/SNRNP200ATP
Snu114/EFTUD2GTP
Prp2/DHX16ATP
Branching factors (YJU2, ISY1)
Prp16/DHX38ATP
Branching factors
Exon ligation factors (CACTIN, SDE2, NKAP)
Prp22/DHX8ATP
Prp43/DHX15ATP
mRNA
5′ss recognition
U1 snRNP
BP recognition
17S U2 snRNP
5′ss and 3′ss proofreading
U4/U6-U5 snRNP
5′ss transfer
Formation of active site
U4 snRNP
BP positioning
First step
*snRNAs not shown
2′OH
Branching
3′–5′ss docking
Second step
Exon ligation
Intron release
Intron lariat
S. cerevisiae 6n7p
S. cerevisiae 6g90
H. sapiens 6qx9
H. sapiens 6ahd
H. sapiens 6ff7
S. cerevisiae 6j6n
H. sapiens 5yzg
H. sapiens 5mqf
H. sapiens 6qdv
U4/U6–U5 snRNP
U6–U5 snRNP
SNRNP200

**Fig. 1 | The splicing mechanism and the spliceosome.** Pre-mRNA splicing involves the identification of intron–exon boundaries (splice sites) and two successive transesterification reactions (catalytic steps) (see Box 1 for details). The spliceosome comprises 5 small nuclear ribonucleoprotein (snRNP) complexes (U1, U2, U4, U5 and U6) and more than 150 additional proteins, which together recognize the splice sites, bring them together and catalyse intron removal[4,6]. Left-hand column: the dynamics of spliceosome assembly and the exchanges of snRNPs and other factors, which are driven by ATP-consuming RNA helicases — (yeast/human) Prp5/DDX46, Uap56/DDX39B, Prp28/DDX23, Brr2/SNRNP200, Snu114/EFTUD2, Prp2/DHX16, Prp16/DHX38, Prp22/DHX8 and Prp22/DHX15 — that can resolve kinetic traps along the pathway to spliceosome activation, catalysis and product release. Central column: key RNA–RNA interactions that occur during the process. Right-hand column: snapshot cryogenic electron microscopy structures of the different complexes so far available in different organisms[2–4,6,34,95,261–267]. The structures with the highest resolution available are shown (Protein Data Bank (PDB) codes indicated at the top left of each). Structures from *Saccharomyces cerevisiae* are shown for complexes that have not yet been determined for *Homo sapiens* (for example, the E and A complex structures). For simplicity and owing to space constraints, a recently described pre-Bact complex has not been included[268] and additional snapshots are likely to emerge from future work, including those corresponding to spliceosome proofreading mechanisms. See Supplementary Tables 1–3 for a full list of spliceosome components in *H. sapiens*, *S. cerevisiae* and *Schizosaccharomyces pombe*, and see Supplementary Table 4 for details of all published cryogenic electron microscopy structures of spliceosome complexes. BP, branch point; EJC, exon junction complex; NTC, Prp19 (NineTeen) complex; NTR, ntc-related complex; SF1, splicing factor 1; snRNA, small nuclear RNA; 3′ss, 3′ splice site; 5′ss, 5′ splice site; ILS, intron lariat spliceosome.

isoform of BCL-X in cancer cells[59] or the repression of a cryptic 5′ss in lamin A, which becomes activated in Hutchinson–Gilford's progeria[60].

## 3′ splice site recognition

Recognition of the 3′ end of introns in higher eukaryotes is initiated by the cooperative binding of three interacting proteins — splicing factor 1 (SF1; also known as branch point-binding protein (BBP)) and the U2AF heterodimer (U2AF1–U2AF2) — to three adjacent sequence motifs, namely the branch point (BP), polypyrimidine tract (PPT) and 3′ss (Figs. 1 and 2). Although distance constraints determine the use of a particular 3′ss upon recognition of the BP, a given 3′ss is frequently associated with more than one functional BP[61]. Mutations in U2AF1 that have been identified in various types of cancer, including myeloid malignancies and lung adenocarcinomas, alter the specificity of 3′ss recognition such that different mutations enhance binding to and selection of 3′ss that have specific nucleotides flanking the conserved 3′ss AG[62–64]. For example, whereas S34F/Y mutants of U2AF1 favour the inclusion of exons harbouring CAG 3′ss and disfavour the inclusion of exons harbouring UAG 3′ss, Q157P/R mutants of U2AF1 favour the inclusion of exons with AG/G 3′ss and promote skipping of exons with AG/A 3′ss[63]. One study found that U2AF1 mutations directly affect stress granule components and responses[65]. Mutations in U2AF2 have also been found in cancer samples and correlate with reduced binding to PPTs[66]. These results illustrate how modulating the binding of core splicing factors, which are generally required for splicing of most introns, can be rate-limiting for splice site selection.

SF1 is subsequently replaced by U2 snRNP at the BP (Fig. 1) and the U2 snRNP proteins SF3B1 and PHF5A have a key role in BP recognition (Fig. 2a). A major rearrangement of SF3B1, from an open to a closed conformation, is triggered by recognition of the pre-mRNA, with the HEAT repeats domain of SF3B1 establishing specific contact with the adenosine at the BP sequence, which is sandwiched between SF3B1 and PHF5A (refs. 67–69) (Fig. 2b). As occurs for multiple other transitions in the spliceosome cycle, an RNA-dependent helicase (PRP5; associated with U2 snRNP) provides a mechanism for proofreading, ensuring that proper recognition of the BP has been achieved within the closed conformation of SF3B1 (ref. 70).

Mutations in SF3B1 are common in various types of tumour[62,71]. They occur in 81% of patients with a class of myelodysplastic syndrome having perinuclear iron accumulations known as ring sideroblasts[72]; the SF3B1 K700E mutation is among the most common single mutations detected in any gene in patients with chronic lymphocytic leukaemia[73,74]; and SF3B1 mutations are detected in 15–36% of eye melanomas[75,76]. SF3B1 mutations are associated with changes in alternative splicing of numerous genes, and involve a characteristic pattern of activation of cryptic 3′ss 10–30 nucleotides upstream of canonical 3′ss (at least in some cases associated with the use of an alternative BP[77]), intron retention and, intriguingly, enhanced splicing of certain partially retained introns[78]. The activation of cryptic 3′ss has been linked to a reduced interaction of mutant SF3B1 with the splicing factor SUGP1 (ref. 79).

One key question relates to how transcriptome changes induced by mutations in factors that recognize 3′ss can influence tumour progression, particularly considering that the same mutation, for example K700E in SF3B1, correlates with worse prognosis in chronic lymphocytic leukaemia but with better prognosis in myelodysplastic syndrome[72–74], or with shorter or longer overall survival depending on the melanoma class[75,76]. Pan-cancer splicing analysis and positive-enrichment CRISPR screening showed that the effects of various SF3B1 mutations converge on repression of BRD9, a core component of the non-canonical BAF chromatin remodelling complex that is a potent tumour suppressor for uveal melanoma, through activation of a 'poison exon' that introduces a premature termination codon and leads to degradation of *BRD9* mRNA[80] (Fig. 4b). Other splicing alterations associated with SF3B1 mutations are also likely to contribute, including for example increased expression of telomerase RNA (and telomerase activity) or decreased expression of the MAP3K7 kinase (which is related to increased NF-κB signalling) or of the haem transporter ABCB7 (which is relevant to sideroblastic anaemia)[81–83].

## Therapeutic targeting

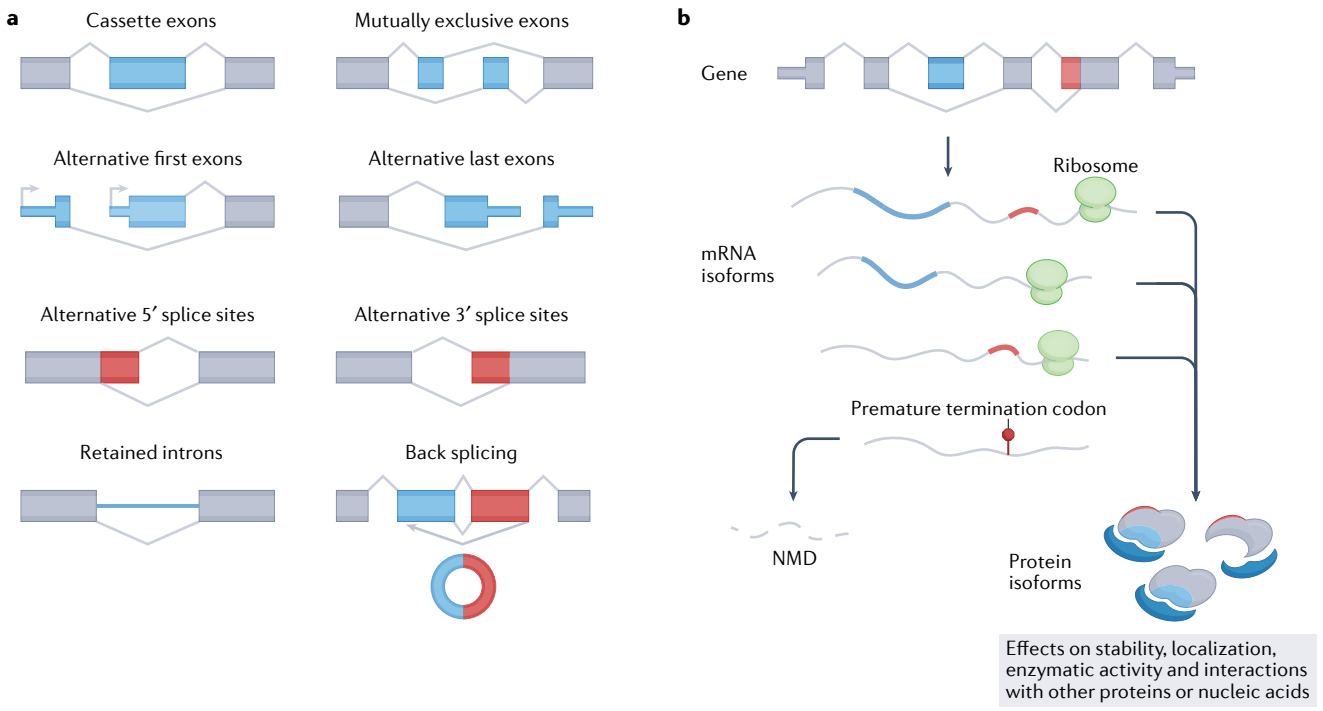
Surprisingly, various families of small molecules that inhibit the conformational change of SF3B1 during BP recognition at the 3′ end of introns have anti-proliferative effects in vitro and inhibit tumour growth in various mouse cancer models[18,67–69,84]. One of these (H3B-8800) is currently in clinical trials for myelodysplastic syndrome[71] (Fig. 4b). Different BP sequences have differential sensitivity to these compounds, with BP sequences that have more extensive base pairing with U2 snRNA being more resistant to their effects. As a result, these compounds can induce changes in splice site selection, rather than general splicing inhibition, at concentrations that have cytostatic rather than cytotoxic effects[85–88]. Treatment with one such compound, pladienolide B, reprogrammes mouse pluripotent cells into totipotent blastomere-like cells that can be cultured stably in vitro[19], suggesting that SF3B1 inhibitors can exert specific effects on various biologically relevant programmes of post-transcriptional regulation. Different structural variants of these small molecules induce alternative splicing changes that are only partially overlapping[89], which suggests that slight modifications to their chemical structures might generate drugs of improved specificity.

Cancer cells, particularly those with mutations in SF3B1, seem to be more sensitive to compounds that inhibit SF3B1 than are non-cancer cells[71]. These observations led to the concept that although cancer cells can tolerate marked alterations in their transcriptomes (for example, induced by mutations in splicing factors) that contribute to tumour progression, as a result they become more susceptible to further perturbations of the splicing process. A similar synergistic effect (known

## Box 2

# Functions of alternative splicing

Different combinations of binary splice site choices produce various classes of alternative splicing events, which are observed at different frequencies in the human transcriptome. These include cassette exons or mutually exclusive exons that can be included or skipped, the use of alternative splice sites associated with the use of alternative promoters or polyadenylation sites (giving rise to alternative first or last exons, respectively), the use of alternative 5′ or 3′ splice sites (5′ss or 3′ss) within exons, the retention of certain introns and reverse splicing reactions (back splicing) that generate circular RNA molecules (see the figure, part **a**). However, the extent to which alternative splicing affects protein and/or cell function or, rather, represents a by-product of transcriptome noise is unclear. There are many examples of alternative splicing generating protein isoforms that are relevant for cellular or organismal phenotypes, disease progression or the ecology of organisms, ranging from apoptotic switches to opioid analgesia, from neural function to sexual behaviour or seed edibility (reviewed in refs. 299,300). The question is whether these functional examples are the norm or, rather, are exceptions among the hundreds of thousands of alternatively spliced transcripts that exist. Whereas some proteomic studies detect mainly a single isoform expressed at the protein level in most tissues[301], or a limited number of tissue-specific isoforms in specific protein families[302], other studies suggest that a major fraction of alternatively spliced mRNAs is translated to different protein isoforms[303] (see the figure, part **b**). Systematic analyses of the effects of alternative splicing on protein–protein interaction networks are also more compatible with alternative splicing having widespread effects on protein and cellular function[304–309]. Alternative splicing tends to affect disordered protein domains, which are often involved in functionally important protein–protein interactions[304,310,311]. Protein isoforms can also have differences in stability, localization, enzymatic activity and protein–nucleic acid interactions. Also of relevance to the functions of alternative splicing, 30% of alternative splicing events introduce premature termination codons that can trigger nonsense-mediated decay (NMD) and other mechanisms of RNA degradation and, therefore, alternative splicing often functions to control mRNA abundance[312–315]. The exon junction complex (EJC), which is deposited on the mRNA on completion of splicing, has a role in translation-coupled nonsense-mediated decay as well as impairing cryptic splice site usage[316]. A pooled CRISPR–Cas9 screen assessing the relevance of such ‘poison exons’ that cause premature termination of translation showed that, for 50% of the tested exons, deletion had effects on cancer cell viability and xenograft growth, with a subset of these exons having tumour suppressor activity[317]. These results suggest the widespread functionality of splicing regulation, even in cases where alternative splicing does not generate alternative protein products.



as synthetic lethality) applies to the cell cycle inhibitor indisulam and related compounds, which enhance degradation of the U2AF2-related factor RBM39, resulting in alternative splicing changes that have cytotoxic effects in haematological malignancies, particularly in those that already have mutations in other splicing factors such as SF3B1 or U2AF (refs. 90,91) (Fig. 4b).

Recent results suggest an interesting additional use of splicing factor inhibitors such as pladienolide B or indisulam in oncology. Previous work has shown that splicing alterations in cancer can lead to the production of tumour-associated neoantigens, for example through the activation of cryptic splice sites that introduce in-frame or out-of-frame novel amino acid sequences in protein-coding genes[92,93]. This, in turn, can enhance the

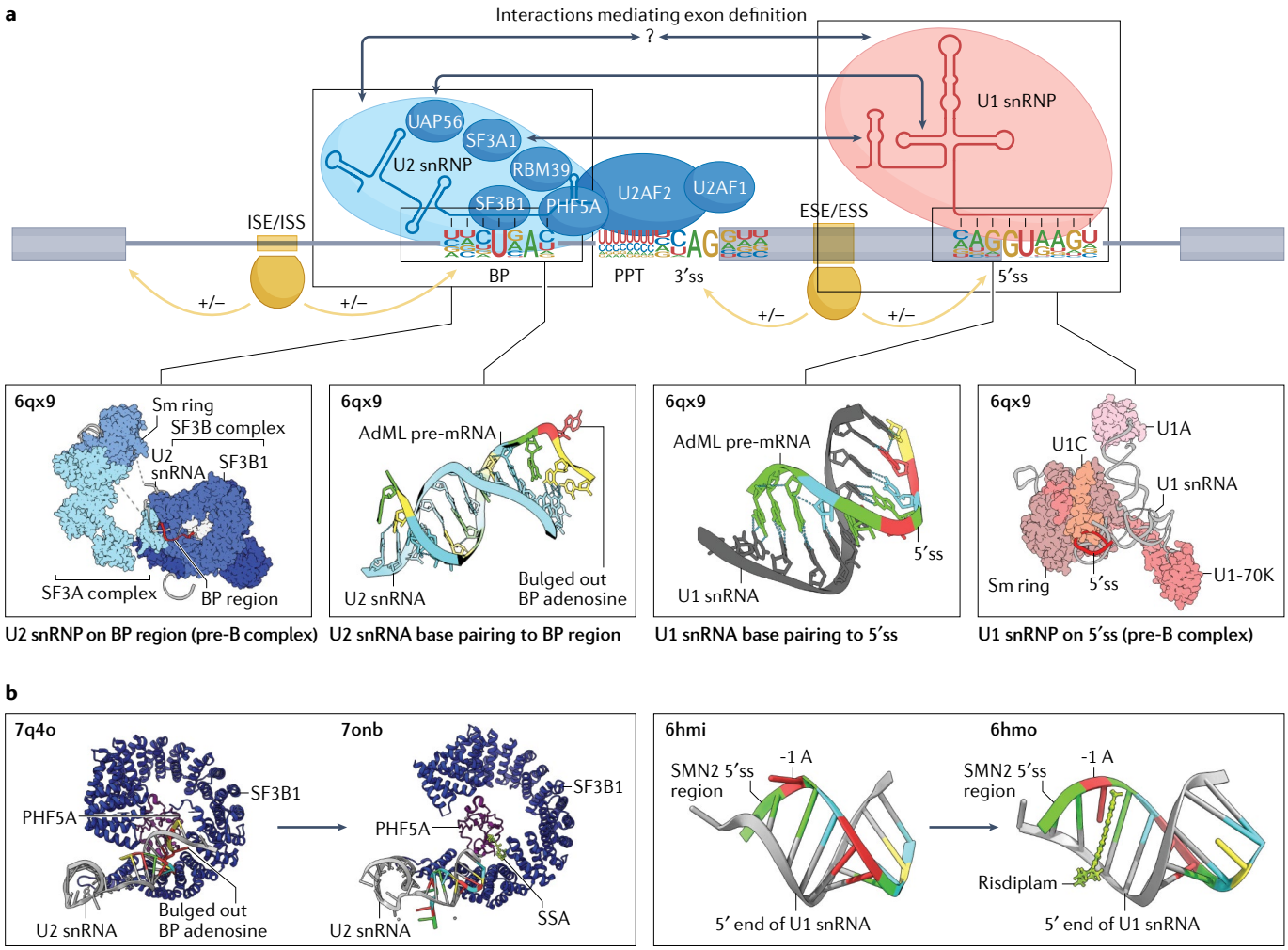
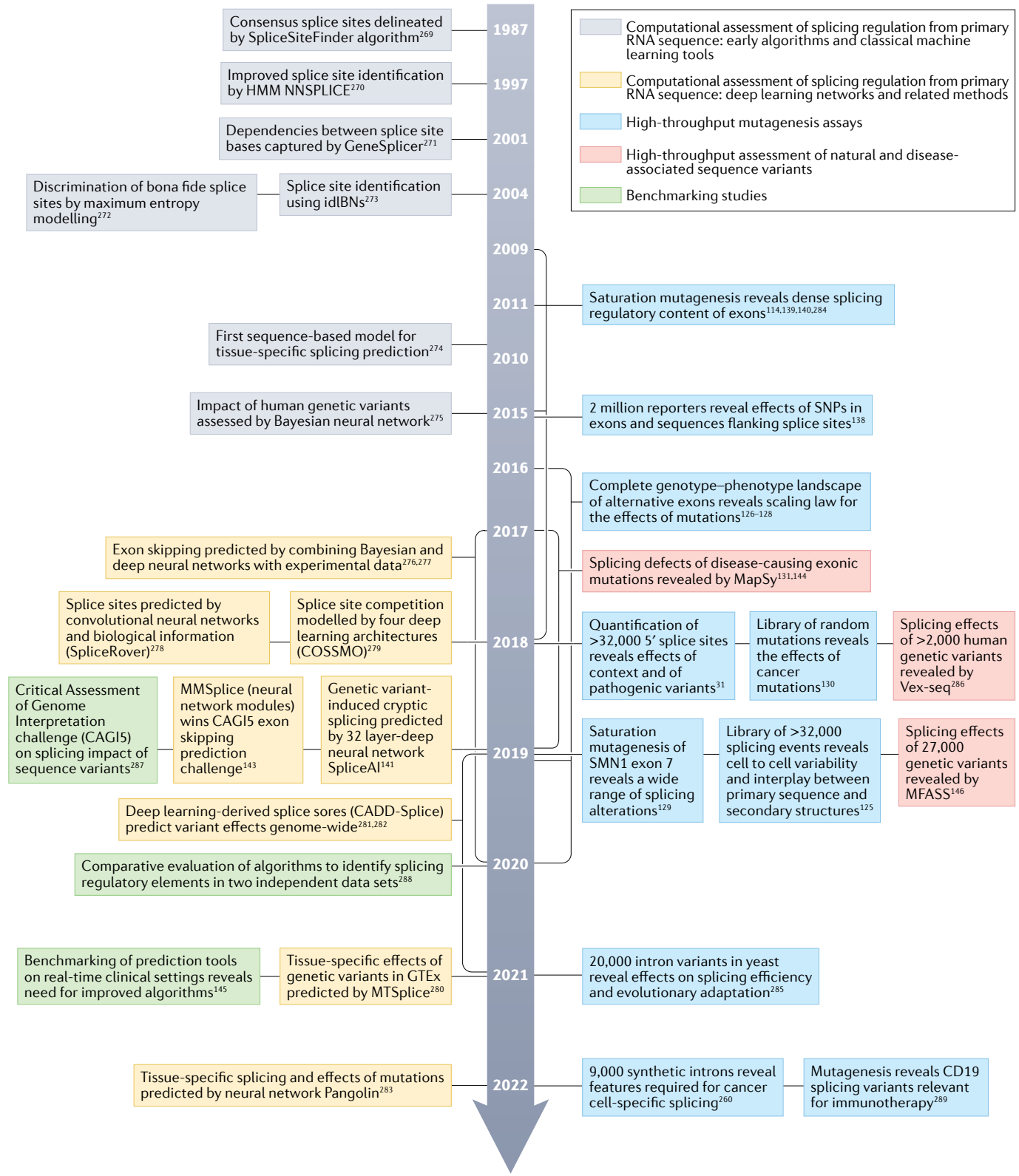



**Fig. 2 | Early splice site recognition and exon definition. a**, Base-pairing interactions between pre-mRNA sequences (the branch point (BP) and 5′ splice site (5′ss)) and small nuclear RNA (snRNA) components of small nuclear ribonucleoprotein (snRNP) complexes (U2 and U1, respectively) are crucial for the definition of intron–exon boundaries. These interactions are assisted by proteins of the snRNP complexes and other auxiliary factors — such as U2AF2, which binds to the polypyrimidine tract (PPT), and U2AF1, which binds to the 3′ splice site (3′ss) adenosine–guanosine (AG)[4,6]. Various sequences can function as splice sites (represented by sequence logos), their strength generally correlating with their potential to base pair with U2 or U1 snRNAs and with the length and uridine-richness of the PPT. Intronic and exonic splicing enhancers (ISEs and ESEs) and intronic and exonic splicing silencers (ISSs and ESSs) are recognized by regulatory factors (shown in orange), such as SR proteins, heterogenous nuclear ribonucleoproteins (hnRNPs), RBM or CELF proteins, that enhance or inhibit the association of U1 and U2 snRNPs with the splice sites. Interactions between U1 and U2 snRNP complex components have been proposed to mediate intron (and possibly also exon) definition and splice site pairing (see main text for details); additional stabilizing interactions across exons and introns are likely. The four structural snapshots illustrate the principles of 3′ss and 5′ss recognition by U2 and U1 snRNPs, respectively, and the role of RNA–RNA interactions in these processes. **b**, Structural snapshots showing the changes in conformation of U2 and U1 snRNP complex components upon binding to spliceostatin A (SSA) and a risdiplam analogue. SSA belongs to a family of splicing inhibitors with antitumour properties that prevent the transition of the SF3B1–PHF5A complex that mediates BP recognition from an open to a closed conformation. Risdiplam is a small molecule that stabilizes the interaction between U1 snRNA and the 5′ss of *SMN2* exon 7, facilitating exon inclusion and the production of functional SMN2 protein as a therapy for spinal muscular atrophy. Protein Data Bank (PDB) codes for each structure are indicated at the top left of each box.




immune surveillance of cancer cells. Thus, small-molecule splicing modulators can enhance the production of tumour neoepitopes that trigger effective antitumour immune responses. Pharmacological modulation of splicing can therefore be combined with therapies that prevent the inhibition of T cell-mediated immune responses – 'checkpoint blockers' such as antibodies to PD1 – to enhance their effects[22] (Fig. 4b).

**Fig. 3 | A timeline of key events in cracking the 'splicing code'.** How does a cell distinguish between exons and introns? How does one cell type decide that a particular sequence should be included in the mature mRNA whereas another cell type decides to skip it? How is this achieved given that the sequences at intron boundaries (5′ splice site (5′ss) and 3′ splice site (3′ss)) are highly diverse in multicellular organisms — with the exception of GU/C at the 5′ end of the intron and adenosine–guanosine (AG) at the 3′ end of the intron (Fig. 2)? In the past four decades, several computational and high-throughput experimental approaches have been developed to crack the 'splicing code' — in other words, to identify from genomic sequences alone bona fide splice sites and their differential use[31,114,125–131,138–141,143–146,260,269–289]. See Supplementary Table 5 for further details of these studies. GTEx, Genotype-Tissue Expression database; idlBNs, inclusion-driven learned Bayesian networks; SNP, single nucleotide polymorphism.

Similarly to the regulation of 5′ss recognition, antisense oligonucleotides targeting 3′ss also have potential therapeutic applications. For example, blocking activation of the poison exon in *BRD9* that is activated by SF3B1 mutations suppresses tumour growth[80], and targeting a 3′ss in the oncogene *ERG* inhibits the proliferation of prostate cancer cells[94] (Fig. 4b).

## Splice site communication

5′ss and 3′ss necessarily need to pair ('commit') to each other for the splicing process to occur. Assembly of the U4/U6–U5 tri-snRNP on pre-mRNAs at positions where 5′ss and 3′ss are recognized by U1 and U2 snRNPs, respectively, establishes such pairing through multiple RNA–RNA, RNA–protein and protein–protein interactions with the pre-mRNA and/or early splicing factors[6,95] (Fig. 1). Connections between 5′ss and 3′ss can, however, also occur before tri-snRNP assembly and such interactions are thought (although not yet fully proven) to stabilize complexes formed on the splice sites across the intron. Examples of such connections involve interactions between U1 snRNP and the 3′ and 5′ domains of U2 snRNP[34]; the interaction of stem–loop IV of U1 snRNA (which is essential for splicing) with a non-canonical RNA binding domain in the U2 snRNP protein SF3A1 (refs. 96,97); the interaction of stem–loop III of U1 snRNA with the U2AF-associated RNA helicase UAP56 (ref. 98); and the interaction of SF1 with the U1 snRNP-associated protein Prp40p in yeast[99,100] (Fig. 2). It is currently unclear whether these are the main contacts for early splice site communication or whether multiple other molecular bridges can be formed on different introns.

An important complementary concept posits that some of these interactions might occur not only between splice sites across introns but also across internal exons through a process known as exon definition[101,102] (Fig. 2). The mutual stabilization of splice site recognition complexes across internal exons can explain various long-standing observations, including the upper (250 nucleotides) and lower (50 nucleotides) length constraints of these exons (which define the boundaries for optimal exon definition interactions) and the fact that genetic mutations in splice sites can induce exon skipping (as a consequence of the failure of exon definition) instead of intron retention (which would be the consequence of a failure of intron definition)[103,104]. It might also explain the results of saturation mutagenesis of 5′ss, whereby the strength of the upstream 3′ss was found to influence the effects of 5′ss mutations[31].

An important exception to the length constraints of internal exons are microexons, which are ~3–27 nucleotides in length but have important regulatory effects on protein functions in nervous system development, synaptic transmission and autism spectrum disorder[105–107]. Microexons have evolved specific mechanisms of recognition coordinated by the neuron-specific regulatory protein SRRM4, which functions as a master regulator for this programme[108–110]. Master regulatory factors have been described to coordinate other programmes of splicing regulation, for example for sex determination in fruit flies, shaping synapses or coordinating epithelial–mesenchymal transition in vertebrates[111–113].

## Role of regulatory sequences

Exon sequences themselves can also contribute to splice site recognition and exon definition through the function of exonic splicing enhancers (ESEs) and exonic splicing silencers (ESSs) (Fig. 2). Together with ISEs and ISSs, these regulatory sequences are thought to nucleate the assembly of complexes of regulatory factors that promote or inhibit splice site recognition by the core splicing machinery[108,114,115] (Fig. 2). This can be achieved through various mechanisms, including the recruitment of core splicing factors through direct interactions with these regulatory complexes[116,117], the establishment of exclusion zones through cooperative coating of the RNA by RNA binding proteins[118,119], interference with specific interactions mediating exon and/or intron definition[120] or the formation of higher-order assemblies (possibly even experiencing local phase transitions) involving tyrosine-rich intrinsically disordered protein domains that are themselves regulated by alternative splicing[121,122]. Classical examples include proteins of the arginine–serine-rich (SR) family, which have positive effects on splicing from exonic enhancers, and proteins of the heterogenous nuclear ribonucleoprotein (hnRNP) family, which inhibit spliceosome assembly from intronic silencers[108,115]. However, often the same sequence motif and cognate factors can have positional effects, for example promoting exon skipping when bound upstream of an alternative exon but promoting exon inclusion when bound downstream[108,115,123–125].

Not unexpectedly, exon mutations have broad effects on alternative splicing. Recent high-throughput, saturation mutagenesis studies have shown that two thirds of all possible mutations in an alternative exon can affect its inclusion, whereas this was not the case for constitutive exons[126–130]. This has obvious implications for the joint evolution of splicing and protein codes[131,132] and for understanding the effects of synonymous mutations in natural genetic variation associated with genetic diseases[133–135] as well as in cancer[136,137].

Several recent studies have aimed to systematically assess the effects of exonic or proximal intronic sequence motifs on splice site selection using high-throughput read-outs (Fig. 3; see Supplementary Table 5). For example, one study assessing and modelling the effects of random libraries of 25 nucleotides flanking 5′ss or 3′ss (involving more than 2 million synthetic minigenes) showed that the vast majority of possible hexamer sequence motifs influence splice site selection, having similar effects in 5′ss or 3′ss competition assays[138]. This large variability of sequence motifs is consistent with results from other reports, although these studies found differential positional effects of ESSs but not ESEs[139,140]. Other studies have also highlighted the importance of genomic context and of starting levels of exon inclusion on the effects of mutations and in the generation of transcriptome complexity during evolution[127,141,142], as well as the contribution of various inputs to splice site selection in most native contexts[125] (Fig. 3; see Supplementary Table 5).

These efforts have obvious relevance for predicting the effects of potentially pathogenic mutations, and various strategies have been envisioned to assist in genetic counselling by modelling and/or

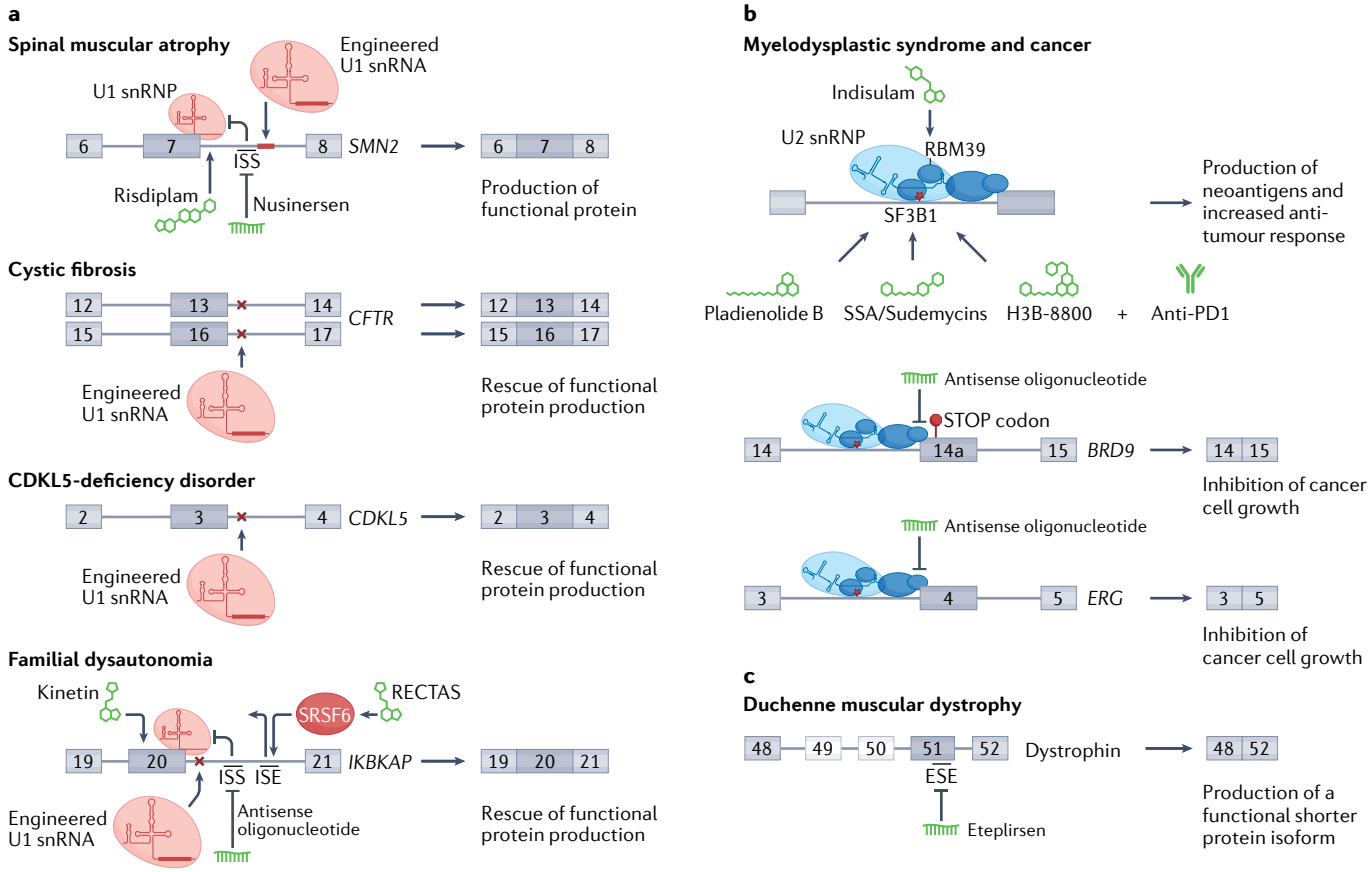



**Fig. 4 | Therapeutic targeting of splice site recognition and exon definition.** Examples of how small-molecule drugs, antisense oligonucleotides or engineered U1 small nuclear ribonucleoprotein (snRNP) complexes can be used as potential or current therapies for genetic diseases or cancer. **a**, Targeting 5′ splice site (5′ss) recognition. Examples are provided of four diseases in which promoting 5′ss recognition can enhance the inclusion of exons and lead to the synthesis of functional proteins whose production was disrupted in the indicated pathologies. This is achieved by using engineered U1 snRNP (in which the 5′ end sequence of U1 small nuclear RNA (snRNA) has been modified to target the complex to specific locations within a transcript), antisense oligonucleotides (including nusinersen) targeting intronic silencers or small molecules (see text for details). **b**, Targeting 3′ splice site (3′ss) recognition. Examples of therapeutic interventions relevant in oncology include the production of neoantigens by inducing inhibition of 3′ss recognition using the indicated SF3B1-targeting or RBM39-targeting drugs; neoantigens in combination with immune checkpoint blockade (such as anti-PD1 therapy) can elicit immunotherapy responses. Other examples include blocking 3′ss recognition by antisense oligonucleotides, leading to exon skipping events that inhibit cancer cell growth. **c**, Targeting regulatory sequences. Antisense oligonucleotides (including eteplirsen) targeting exonic enhancers induce skipping of exons containing inactivating mutations in the Dystrophin gene that cause Duchenne muscular dystrophy, leading to in-frame deletion and production of a shorter protein isoform that rescues function. ESE, exonic splicing enhancer; ISE, intronic splicing enhancer; ISS, intronic splicing silencer; SSA, spliceostatin A.

experimentally assessing splicing perturbations in genes of interest[141,143–145]. For example, a high-throughput system designed to test the effects of more than 27,000 variants annotated in the Genome Aggregation Database found that very rare variants had large effects on splicing, mostly located outside the splice sites themselves[146].

### Therapeutic targeting

The antisense oligonucleotide eteplirsen, which targets ESE sequences, has been approved as a therapy for Duchenne muscular dystrophy. Eteplirsen promotes skipping of an exon in the dystrophin gene that harbours inactivating mutations, leading to in-frame production of a shorter, but still functional, protein that restores muscle function[147] (Fig. 4c). More generally, approaches targeting exon sequences may be used to generate truncated protein variants lacking specific domains, for interrogating gene and protein function without requiring complex genome engineering.

## Co-transcriptional regulation

Long-standing evidence ranging from electron microscopy[148] to transcriptome analyses of RNAs physically associated with transcribing RNA Pol II[149] indicates that the removal of introns from pre-mRNAs can occur co-transcriptionally, even almost immediately after the 3′ss exits from the polymerase tunnel (reviewed in refs. 150,151) (Fig. 5). Although in vitro-transcribed model pre-mRNAs can be spliced in nuclear extracts or upon transfection or injection (for example, in the nucleus of *Xenopus* oocytes)[152], which indicates that splicing can be uncoupled from transcription, co-transcriptional splicing has important mechanistic implications, as functional connections between the transcription and

splicing machineries can enhance splicing efficiency and influence splice site choice[153,154]. Transcription elongation rates determine the time window during which alternative splice sites enter into competition (reviewed in refs. 153,155,156), which can influence splice site choice via a kinetic model that has been shown to operate both in vitro and in vivo, in animals and in plants[157–159] (Fig. 5a). Conversely, splice site recognition can influence promoter choice, transcription elongation and proper 3′ end formation[154,160–162].

However, more recently, the extent to which co-transcriptional splicing occurs and/or contributes to the majority of splicing events has been questioned[163]. Results from long-read sequencing of nascent pre-mRNA transcripts indicate that splicing in human and *Drosophila* cells typically occurs after RNA Pol II has transcribed several kilobases of pre-mRNA, with the order of intron removal not following strictly the order of transcription[164]. Another study combining similar technologies with precision run-on sequencing found that, during mouse erythropoiesis, although introns are often spliced during the time of transcription of the downstream intron, nascent transcripts with a high proportion of unspliced introns are also detected[162]. Along the same lines, a three-pronged methodology to characterize nascent RNAs

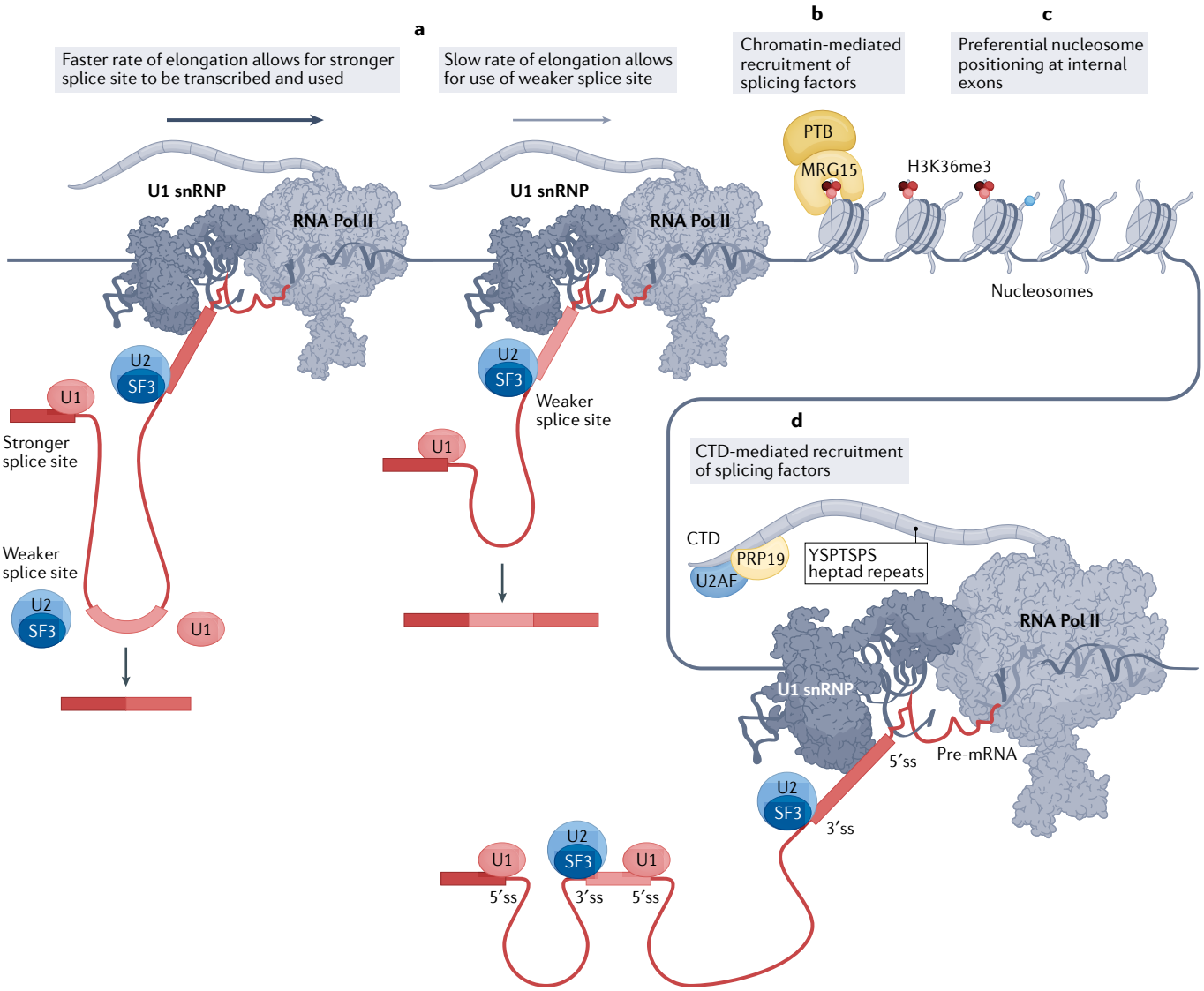



**Fig. 5 | Co-transcriptional splicing.** Intron removal can occur on nascent mRNA transcripts during transcription by RNA polymerase II (RNA Pol II), which has important mechanistic implications for splice site selection. **a**, The elongation rates of RNA Pol II can affect the competition between alternative splice sites. For example, slow rates of elongation facilitate the use of weaker splice sites before stronger splice sites are transcribed[155]. **b**, Chromatin can recruit splicing regulators (such as PTB) through proteins that recognize epigenetic marks, such as MRG15, which binds both PTB and the trimethylation of histone H3 on lysine 36 (H3K36me3)[186,290]. **c**, Preferential nucleosome positioning at internal exons may favour exon definition[291–293]. **d**, There is evidence to suggest that splicing factors (such as U2AF–PRP19)[294] may be recruited through direct interactions with RNA Pol II, including its carboxy-terminal domain (CTD). The CTD is composed of multiple repeats of the heptad amino acid sequence YSPTSPS harbouring potential phosphorylation sites that are linked to different elongation states of the enzyme[156] and to a switch between transcriptional and splicing condensates[178]. Importantly, the interaction between U1 small nuclear ribonucleoprotein (snRNP) and RNA Pol II requires a 5′ splice site (5′ss) being present in the nascent transcript[253]. 3′ss, 3′ splice site.

also documented examples of co-transcriptional splicing occurring immediately after 3′ss transcription, as well as other examples in which splicing was delayed until RNA Pol II had transcribed sequences much further downstream[165]. Nevertheless, clear examples of post-transcriptional splicing do exist. These include examples of coordinately regulated distant alternative splicing events within the same transcript in different cell types[166,167], and the recently reported category of detained introns, excision of which can be induced by signalling cues during development or meiosis, or in certain cancers[168–171]. The otherwise fully processed polyadenylated transcripts containing detained introns are retained at the gene locus until splicing of the detained intron allows for mRNA export to the cytoplasm and translation, thus allowing for rapid changes in protein expression in response to external cues[169–172].

A long-standing issue related to co-transcriptional splicing is whether splicing is subject to some form of compartmentalization in the nucleus. For example, splicing has been proposed to occur in the vicinity of nuclear speckles, regions of the nucleus that have an accumulation of splicing factors, contain active spliceosomes and have phase transition properties (reviewed in refs. 168,173–177). This is particularly relevant considering that different phosphorylation states of the intrinsically disordered carboxy-terminal domain (CTD) of RNA Pol II can drive an exchange from condensates involved in transcription initiation to condensates in nuclear speckles[178]. Two recent studies reported two distinct areas of the nucleus, one associated with nuclear speckles and the other with peripheral nuclear lamina, that had characteristic patterns of splicing regulation[179,180]. Such patterns correlate with differences in the levels of splicing factors and/or genomic guanosine–cytosine (GC) content in these areas, the latter being associated with distinct genomic architectures that have been linked to exon versus intron definition[179–181].

Chromatin organization, including nucleosome positioning and epigenetic marks such as histone modifications and DNA methylation, has also been associated with splice site recognition and regulation (reviewed in refs. 182,183) (Fig. 5b,c). Proposed mechanistic models involve proteins that recognize methylated DNA and slow down RNA Pol II, indirectly affecting splice site selection[184], or adaptor proteins such as MRG15 that recognize both histone tail modifications and splicing factors, thus increasing the local concentration of splicing regulators such as PTB[185] (Fig. 5b). Proving causality, a recent study used genome editing tools to introduce histone modifications that are observed during epithelial–mesenchymal transition at specific genomic locations, which induced corresponding splicing modifications and concomitant biological effects[186].

It is clear that splicing regulation can be influenced by chromatin spatial organization and epigenetic modification, as further illustrated by the recent report of enhanced effects of combining the splicing modulatory drug nusinersen with a histone deacetylase inhibitor[187]. Future work will establish how general and diverse are co-transcriptional mechanisms influencing the splicing code.

## Epitranscriptomic regulation

More than 70 chemical modifications of RNA molecules have been described in eukaryotes and, for some of these, dedicated protein factors involved in their deployment, reading or erasing are known[188]. Such epitranscriptomic modifications might establish a regulatory code on RNA[189], resembling the epigenetic code on DNA and histones that functions to recruit or inhibit enzymatic complexes that modulate transcription, replication or DNA repair[190]. For example, in *Caenorhabditis elegans*, $N^6$-methyladenosine ($m^6A$) modification at the adenosine–guanosine (AG) dinucleotide of 3′ss inhibits its recognition by U2AF1, leading to the retention of an intron in the pre-mRNA encoding *S*-adenosylmethionine (SAM) synthetase[191]. This results in the downregulation of SAM synthetase expression as part of a negative feedback loop by which an excess of methionine triggers accumulation of methylated SAM, which itself functions as the methyl donor for the enzymatic deposition of $m^6A$ on pre-mRNA. A similar mechanism operates in human cells[192]. In this case, under conditions of abundant SAM, the $N^6$-adenosyl-methyl transferase METTL16 methylates a loop structure at the 3′ untranslated region of SAM synthetase pre-mRNA. Under conditions of low levels of SAM, METTL16 binds to the unmethylated loop and enhances splicing of SAM synthetase pre-mRNA, facilitating expression of the enzyme[192]. Thus, in both *C. elegans* and humans, $m^6A$ functions as a sensor of the availability of metabolites to switch off the expression of SAM synthetase via regulation of splicing.

Another mechanism by which $m^6A$ can regulate splicing is via proteins that recognize this modification. For example, hnRNPG (also known as RBMX) recognizes $m^6A$ modifications at exonic positions close to regulated splice sites and, through interactions with RNA polymerase, modulates alternative splicing[193]. Another example is the nuclear $m^6A$ reader YTHDC1, which recruits splicing regulatory factors of the SR protein family (such as SRSF3), but antagonizes the binding of other factors such as SRSF10, with the result of promoting inclusion of alternative exons[194]. The effects of YTHDC1 on splicing and polyadenylation may underlie its essential function in germline development[195]. $m^6A$-mediated regulation of the splicing kinetics of multiple introns, associated with positional effects of $m^6A$ deposition on nascent transcripts[196], is an attractive proposed mechanism for the coordinated control of developmental programmes. However, the extent to which $m^6A$ modifications generally regulate splicing programmes remains unclear, with one study arguing that their major functional effect is on cytoplasmic mRNA stability[197].

Recent work highlights the regulatory potential of pseudouridine modifications at alternatively spliced regions of pre-mRNAs and their regulatory sequences, with direct effects on splicing efficiencies[198]. The tissue-specific expression of pre-mRNA pseudouridine synthases thus offers another potential mechanism for the control of alternative splicing and 3′ end formation[198].

Functionally important RNA modifications also occur in snRNAs, which have 2′-*O*-methyl and pseudouridylated residues at phylogenetically conserved positions, in addition to characteristic cap structures at their 5′ ends (2,2,7-trimethyl-guanosine for U1, U2, U4 and U5 snRNAs and γm-guanosine for U6 snRNAs) (reviewed in ref. 199). Some of these modifications have been shown to be important for snRNP biogenesis and/or for efficient splice site recognition, and an additional $m^6A$ modification in U2 snRNA has been proposed to modulate 3′ss choice[199,200].

## Regulation by RNA structure

One difficulty in assessing the functional effects of RNA structure on alternative splicing is that RNAs exist, almost invariably from birth, as RNP complexes in which the associated proteins strongly influence the conformation(s) that RNAs adopt during or after folding. Although methods for the high-throughput analysis of higher-order transcriptome structure in living cells (reviewed in ref. 201) remain to be fully exploited in investigating splicing regulation, recent studies argue that introns may be more highly structured than exons and that distinct RNA folding around alternative exons, depending on RNA Pol II elongation rates, influences splicing outcomes[202,203]. Additional

## Glossary

**Alternative splicing**
The process by which intron and/or exon sequences are differentially recognized in different cell types or biological conditions to generate distinct mRNAs and long non-coding RNAs from the same primary transcript.

**Back splicing**
The process by which a 5′ splice site is spliced to a 3′ splice site located upstream in the same pre-mRNA molecule, leading to the generation of a circular RNA, typically spanning one or a few exons.

**Branch point**
(BP). An intronic adenosine nucleotide, typically located 15–45 nucleotides 5′ of the 3′ end of introns, which engages in formation of a 2′–5′ phosphodiester bond with the 5′ end of the intron after the first catalytic step of the splicing reaction.

**Exon definition**
A model for the mutual stabilization of splicing factors recognizing splice sites flanking internal exons in multicellular organisms.

**Intron definition**
A model for the mutual stabilization of splicing factors recognizing the splice sites across an intron, which likely has a major role in the efficient co-transcriptional splicing of many introns.

**Introns**
Internal sequences within primary transcripts produced by eukaryotic RNA polymerase II (RNA Pol II) that are removed through the process of pre-mRNA splicing, allowing their flanking sequences (exons) to be spliced together and thus generate functional mRNAs and long non-coding RNAs.

**MicroRNAs**
A class of small regulatory RNAs whose function is to induce the degradation or repress the translation of mRNAs with which they have full or partial complementarity, respectively. They are often transcribed as part of intronic sequences, from which they are released to be assembled with specific proteins on microRNA-induced silencing complexes.

**Premature termination codons**
Translation termination codons in mRNA arising from single-nucleotide mutations or from alternative splicing events that disrupt an open reading frame, often leading to mRNA degradation by the process of nonsense-mediated decay.

**Recursive splicing**
The sequential excision of shorter pieces of a long intron, each piece being separated from the next by a zero-length exon.

**Small nucleolar RNAs**
(snoRNAs). A class of small regulatory RNAs whose function is to guide the addition of chemical modifications at specific residues in other RNAs, including ribosomal, transfer or small nuclear RNAs (snRNAs). They are often transcribed as part of intronic sequences, from which they are released to be assembled with specific proteins on small nucleolar RNP complexes.

**Spliceosome**
The molecular machinery involved in intron removal, composed of 5 small nuclear ribonucleoprotein (snRNP) complexes (U1, U2, U4, U5 and U6) and more than 150 accessory proteins.

evidence has accumulated supporting roles for secondary RNA structures in splice site recognition (reviewed in refs. 204,205), explaining for example temperature-sensitive splice site selection[206], or bringing together distant splice sites through long-range base pairing[207]. In mammalian cells, stem–loop structures involving splice sites or regulatory sequences, chaperoned by various RNA binding proteins, modulate alternative splicing decisions during development and are of relevance for potential therapeutic approaches in spinal muscular atrophy and tauopathies (neurodegenerative disorders characterized by the deposition of abnormal Tau protein in neurons)[208–211]. Links have also been found between the function of RNPs containing structured RNAs and tumour biology, including G-quadruplexes recognized by hnRNPF that enhance exon inclusion events[212] and are relevant for cancer progression[213], and a promestastatic splicing programme that is regulated by interactions between the protein SNRPA1 and structured splicing enhancers[214].

Base-pairing interactions involving sequences flanking internal exons have been shown to contribute to the production of circular RNAs by facilitating back splicing between the 3′ss and 5′ss associated with the looped-out exon[215–220]. Such sequences are often associated with repetitive DNA elements and their limited conservation has been viewed as evidence against the general functionality of at least some families of circular RNAs[221–223]. There is, however, evidence for the functional relevance of specific circular RNAs, for example as 'sponges' for proteins or microRNAs[224–226], and a functional CRISPR screen showed that a group of circular RNAs are important for cell growth, mostly in a cell type-specific manner, or for the preimplantation development of mouse embryos[227]. Furthermore, some circular RNAs can direct the synthesis of peptides or proteins, the general functional relevance of which is under intense debate[228–232].

## Conclusions and future perspectives

The progress reviewed above provides a valuable framework to understand how the spliceosome has evolved various mechanisms to regulate splice site selection. These range from the control of RNA structure or RNA modifications to the spatial organization of genes in the nucleus, from tight coupling between transcription, chromatin and RNA processing to the complex interactions of regulatory sequences and factors that modulate exon and intron definition. However, important challenges to our understanding of pre-mRNA splicing remain.

A comprehensive and quantitative assessment of the isoform structure of individual full transcript molecules using long-read sequencing in single cells remains challenging[233], but this will be necessary to reconstruct the spatial regulation of alternative splicing (spatial transcriptomics) and to understand precisely the contributions of alternative splicing to tissue development and homeostasis. Linked to this is the need for high-resolution methods to reliably assess cell to cell variability in splice site selection and in the levels and/or activity of regulatory factors[234–237] (Supplementary Box 1). If individual cells of the same type diverge markedly in their alternative splicing decisions[125,238–244], this would call for major revision of our understanding of the molecular mechanisms of splicing regulation, which are currently mostly based upon the study of cell populations. For example, a recent study showed that the regulation of intron excision in certain yeast ribosomal protein genes can be used to induce phenotypic heterogeneity that facilitates population adaptation to starvation or high

levels of sugar availability[245]. Similarly important will be to assess the levels, origins, proofreading mechanisms and potential biological functions of 'noise' in the splicing process, for example by observing synthesis and processing kinetics of single nascent RNA molecules in real time[246,247]. In this regard, recursive splicing[246,248,249] and variations in the order of intron removal[164,250] can have important roles in the kinetics of RNA processing and in splice site selection.

Our knowledge of rate-limiting, regulatable steps in spliceosome assembly and catalysis remains mostly based upon (painstaking) efforts to understand the biochemical process as it occurs in a limited number of pre-mRNA substrates that are detectably spliced in cell or nuclear extracts, which is uncoupled from transcription and chromatin, and neither recapitulates the complexity of splicing decisions nor reflects cell type-specific variations in splicing, epigenetic or epitranscriptomic factors. It is conceivable that cell type-specific or even substrate-specific spliceosomes exist, characterized by different composition, stoichiometry and/or modifications of their components, in addition to the modulation of their function by master regulatory factors that respond to environmental cues to shape tissue-specific transcriptomes during development[251]. Recent efforts to develop cryogenic electron microscopy methods to visualize the complexes involved in coupling between transcription and splicing[252,253], in early steps of splice site communication[34,254] and in the function of higher-order supraspliceosomes[255], as well as tomography-based visualization of spliceosomes in situ, should pave the way to a better understanding of the molecular basis of cell type-specific splicing regulation. Such efforts can be complemented by genetic analyses aimed at reconstructing networks of splicing regulation[91,109,256] and by detailed characterization of pathogenic variants in pre-mRNAs and splicing factors. Other important open questions concern the extent to and mechanisms by which long non-coding RNAs can contribute to the regulation of alternative splicing[257] and how alternative splicing of long non-coding RNAs — including the combination of exons across classical transcriptional units — can contribute to generating a large repertoire of RNA modules with possible functions in gene regulation[258].

Ultimately, integrating structural and functional information to predict patterns of alternative splicing is likely to benefit from artificial intelligence methods that can generate models of splice site selection[259] (Fig. 3). Such approaches will have applications for understanding the effects of genetic variation or pathogenic mutations, as well as for the design of novel therapies to correct splicing alterations or to eliminate cells, such as cancer cells, that have pathogenic splicing phenotypes[260].

Published online: 16 December 2022

## References


1. Irimia, M. & Roy, S. W. Origin of spliceosomal introns and alternative splicing. *Cold Spring Harb. Perspect. Biol.* **6**, a016071 (2014).
2. Plaschka, C., Newman, A. J. & Nagai, K. Structural basis of nuclear pre-mRNA splicing: lessons from yeast. *Cold Spring Harb. Perspect. Biol.* **11**, a032391 (2019).
3. Wan, R., Bai, R., Yan, C., Lei, J. & Shi, Y. Structures of the catalytically activated yeast spliceosome reveal the mechanism of branching. *Cell* **177**, 339–351 (2019).
4. Kastner, B., Will, C. L., Stark, H. & Lührmann, R. Structural insights into nuclear pre-mRNA splicing in higher eukaryotes. *Cold Spring Harb. Perspect. Biol.* **11**, a032417 (2019).
5. Tholen, J. & Galej, W. P. Structural studies of the spliceosome: bridging the gaps. *Curr. Opin. Struct. Biol.* **77**, 102461 (2022).
6. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
7. Turunen, J. J., Niemelä, E. H., Verma, B. & Frilander, M. J. The significant other: splicing by the minor spliceosome. *Wiley Interdiscip. Rev. RNA* **4**, 61–76 (2013).
8. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415 (2008).
9. Wang, E. T. et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
10. Nilsen, T. W. & Graveley, B. R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**, 457–463 (2010).
11. Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* **18**, 437–451 (2017).
12. Zavolan, M. & Kanitz, A. RNA splicing and its connection with other regulatory layers in somatic cell reprogramming. *Curr. Opin. Cell Biol.* **52**, 8–13 (2018).
13. Scotti, M. M. & Swanson, M. S. RNA mis-splicing in disease. *Nat. Rev. Genet.* **17**, 19–32 (2016).
14. Manning, K. S. & Cooper, T. A. The roles of RNA processing in translating genotype to phenotype. *Nat. Rev. Mol. Cell Biol.* **18**, 102–114 (2017).
15. Wright, C. J., Smith, C. W. J. & Jiggins, C. D. Alternative splicing as a source of phenotypic diversity. *Nat. Rev. Genet.* **23**, 697–710 (2022).
16. Gebauer, F., Schwarzl, T., Valcárcel, J. & Hentze, M. W. RNA-binding proteins in human genetic disease. *Nat. Rev. Genet.* **22**, 185–198 (2021).
17. Finkel, R. S. et al. Nusinersen versus sham control in infantile-onset spinal muscular atrophy. *N. Engl. J. Med.* **377**, 1723–1732 (2017).
**This study provides evidence for the clinical benefit of splicing modulation in the treatment of a human genetic disorder**.
18. Bonnal, S. C., López-Oreja, I. & Valcárcel, J. Roles and mechanisms of alternative splicing in cancer — implications for care. *Nat. Rev. Clin. Oncol.* **17**, 457–474 (2020).
19. Shen, H. et al. Mouse totipotent stem cells captured and maintained through spliceosomal repression. *Cell* **184**, 2843–2859 (2021).
**This study reveals an unexpected link between splicing activity and cell totipotency, with potential applications in regenerative medicine**.
20. To, K. K. W. & Cho, W. C. S. An overview of rational design of mRNA-based therapeutics and vaccines. *Expert. Opin. Drug. Discov.* **16**, 1307–1317 (2021).
21. Black, D. L. Finding splice sites within a wilderness of RNA. *RNA* **1**, 763–771 (1995).
22. Lu, S. X. et al. Pharmacologic modulation of RNA splicing enhances anti-tumor immunity. *Cell* **184**, 4032–4047 (2021).
**This study illustrates the potential of splicing inhibitors to enhance the generation of neoantigens expressed in cancer cells**.
23. Kondo, Y., Oubridge, C., van Roon, A. M. M. & Nagai, K. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5′ splice site recognition. *eLife* **4**, 1–19 (2015).
24. Slaugenhaupt, S. A. et al. Tissue-specific expression of a splicing mutation in the IKBKAP gene causes familial dysautonomia. *Am. J. Hum. Genet.* **68**, 598–605 (2001).
25. Dietrich, P. & Dragatsis, I. Familial dysautonomia: mechanisms and models. *Genet. Mol. Biol.* **39**, 497–514 (2016).
26. Shuai, S. et al. The U1 spliceosomal RNA is recurrently mutated in multiple cancers. *Nature* **574**, 712–716 (2019).
27. Suzuki, H. et al. Recurrent noncoding U1 snRNA mutations drive cryptic splicing in SHH medulloblastoma. *Nature* **574**, 707–711 (2019).
**Shuai et al. and Suzuki et al. show that mutations in snRNAs can promote cancer progression**.
28. Roca, X. & Krainer, A. R. Recognition of atypical 5′ splice sites by shifted base-pairing to U1 snRNA. *Nat. Struct. Mol. Biol.* **16**, 176–182 (2009).
29. Roca, X. et al. Widespread recognition of 5′ splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes. Dev.* **26**, 1098–1109 (2012).
30. Roca, X., Krainer, A. R. & Eperon, I. C. Pick one, but be quick: 5′ splice sites and the problems of too many choices. *Genes. Dev.* **27**, 129–144 (2013).
31. Wong, M. S., Kinney, J. B. & Krainer, A. R. Quantitative activity profile and context dependence of all human 5′ splice sites. *Mol. Cell* **71**, 1012–1026.e3 (2018).
**This study presents a systematic assessment of the activity and context dependence of sequence variation at 5′ splice sites**.
32. Aznarez, I. et al. A systematic analysis of intronic sequences downstream of 5′ splice sites reveals a widespread role for U-rich motifs and TIA1/TIAL1 proteins in alternative splicing regulation. *Genome Res.* **18**, 1247–1258 (2008).
33. Yu, Y. et al. Dynamic regulation of alternative splicing by silencers that modulate 5′ splice site competition. *Cell* **135**, 1224–1236 (2008).
34. Plaschka, C., Lin, P. C., Charenton, C. & Nagai, K. Prespliceosome structure provides insights into spliceosome assembly and regulation. *Nature* **559**, 419–422 (2018).
35. Jourdain, A. A. et al. Loss of LUC7L2 and U1 snRNP subunits shifts energy metabolism from glycolysis to OXPHOS. *Mol. Cell* **81**, 1905–1919 (2021).
36. Singh, N. N., Singh, R. N. & Androphy, E. J. Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. *Nucleic Acids Res.* **35**, 371–389 (2007).
37. Hua, Y. et al. Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model. *Nature* **478**, 123–126 (2011).
38. Jha, N. N., Kim, J. K. & Monani, U. R. Motor neuron biology and disease: a current perspective on infantile-onset spinal muscular atrophy. *Future Neurol.* **13**, 161–172 (2018).
39. Albrechtsen, S. S., Born, A. P. & Boesen, M. S. Nusinersen treatment of spinal muscular atrophy — a systematic review. *Dan. Med. J.* **67**, 1–12 (2020).
40. Ratni, H. et al. Discovery of risdiplam, a selective survival of motor neuron-2 (SMN2) gene splicing modifier for the treatment of spinal muscular atrophy (SMA). *J. Med. Chem.* **61**, 6501–6517 (2018).
41. Darras, B. T. et al. Risdiplam-treated infants with type 1 spinal muscular atrophy versus historical controls. *N. Engl. J. Med.* **385**, 427–435 (2021).

42. Campagne, S. et al. Structural basis of a small molecule targeting RNA for a specific splicing correction. *Nat. Chem. Biol.* **15**, 1191–1198 (2019).
**This study provides the molecular rationale for the specific effects of a small molecule modulator of 5′ splice recognition**.
43. Singh, R. N., Seo, J. & Singh, N. N. RNA in spinal muscular atrophy: therapeutic implications of targeting. *Expert. Opin. Ther. Targets* **24**, 731–743 (2020).
44. Alanis, E. F. et al. An exon-specific U1 small nuclear RNA (snRNA) strategy to correct splicing defects. *Hum. Mol. Genet.* **21**, 2389–2398 (2012).
45. Rogalska, M. E. et al. Therapeutic activity of modified U1 core spliceosomal particles. *Nat. Commun.* **7**, 11168 (2016).
46. Donegà, S. et al. Rescue of common exon-skipping mutations in cystic fibrosis with modified U1 snRNAs. *Hum. Mutat.* **41**, 2143–2154 (2020).
47. Donadon, I. et al. Rescue of spinal muscular atrophy mouse models with AAV9-exon-specific U1 snRNA. *Nucleic Acids Res.* **47**, 7618–7632 (2019).
48. Balestra, D. et al. Splicing mutations impairing CDKL5 expression and activity can be efficiently rescued by U1snRNA-based therapy. *Int. J. Mol. Sci.* **20**, 4130 (2019).
49. Kaida, D. et al. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**, 664–668 (2010).
50. Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A. & Burge, C. B. Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. *Science* **320**, 1643–1647 (2008).
51. Oh, J. M. et al. U1 snRNP regulates cancer cell migration and invasion in vitro. *Nat. Commun.* **11**, 1–8 (2020).
52. Sinha, R. et al. Antisense oligonucleotides correct the familial dysautonomia splicing defect in IKBKAP transgenic mice. *Nucleic Acids Res.* **46**, 4833–4844 (2018).
53. Bruun, G. H. et al. Blocking of an intronic splicing silencer completely rescues IKBKAP exon 20 splicing in familial dysautonomia patient cells. *Nucleic Acids Res.* **46**, 7938–7952 (2018).
54. Donadon, I. et al. Exon-specific U1 snRNAs improve ELP1 exon 20 definition and rescue ELP1 protein expression in a familial dysautonomia mouse model. *Hum. Mol. Genet.* **27**, 2466–2476 (2018).
55. Axelrod, F. B. et al. Kinetin improves IKBKAP mRNA splicing in patients with familial dysautonomia. *Pediatr. Res.* **70**, 480–483 (2011).
56. Hims, M. M. et al. Therapeutic potential and mechanism of kinetin as a treatment for the human splicing disease familial dysautonomia. *J. Mol. Med.* **85**, 149–161 (2007).
57. Yoshida, M. et al. Rectifier of aberrant mRNA splicing recovers tRNA modification in familial dysautonomia. *Proc. Natl Acad. Sci. USA* **112**, 2764–2769 (2015).
58. Ajiro, M. et al. Therapeutic manipulation of IKBKAP mis-splicing with a small molecule to cure familial dysautonomia. *Nat. Commun.* **12**, 1–12 (2021).
59. Zhang, J. et al. Correction of Bcl-x splicing improves responses to imatinib in chronic myeloid leukaemia cells and mouse models. *Br. J. Haematol.* **189**, 1141–1150 (2020).
60. Osorio, F. G. et al. Hutchinson–Gilford progeria: splicing-directed therapy in a new mouse model of human accelerated aging. *Sci. Transl. Med.* **3**, 1–12 (2011).
61. Pineda, J. M. B. & Bradley, R. K. Most human introns are recognized via multiple and tissue-specific branchpoints. *Genes. Dev.* **32**, 577–591 (2018).
62. Yoshida, K. et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).
63. Shirai, C. L. et al. Mutant U2AF1 expression alters hematopoiesis and pre-mRNA splicing in vivo. *Cancer Cell* **27**, 631–643 (2015).
64. Park, S. M. et al. U2AF35(S34F) promotes transformation by directing aberrant ATG7 pre-mRNA 3′ end formation. *Mol. Cell* **62**, 479–490 (2016).
65. Biancon, G. et al. Precision analysis of mutant U2AF1 activity reveals deployment of stress granules in myeloid malignancies. *Mol. Cell* **82**, 1107–1122.e7 (2022).
66. Maji, D. et al. Representative cancer-associated U2AF2 mutations alter RNA interactions and splicing. *J. Biol. Chem.* **295**, 17148–17157 (2020).
67. Cretu, C. et al. Molecular architecture of SF3b and structural consequences of its cancer-related mutations. *Mol. Cell* **64**, 307–319 (2016).
68. Cretu, C. et al. Structural basis of splicing modulation by antitumor macrolide compounds. *Mol. Cell* **70**, 265–273 (2018).
69. Teng, T. et al. Splicing modulators act at the branch point adenosine binding pocket defined by the PHF5A–SF3b complex. *Nat. Commun.* **8**, 1–16 (2017).
70. Zhang, Z. et al. Molecular architecture of the human 17S U2 snRNP. *Nature* **583**, 310–313 (2020).
71. Seiler, M. et al. H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. *Nat. Med.* **24**, 497–504 (2018).
**This study reports the development of an SF3B1 inhibitor that has greater effects in cancer cells that have mutations in SF3B1 components**.
72. Mangaonkar, A. A. et al. Prognostic interaction between bone marrow morphology and SF3B1 and ASXL1 mutations in myelodysplastic syndromes with ring sideroblasts. *Blood Cancer J.* **8**, 1–4 (2018).
73. Quesada, V. et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat. Genet.* **44**, 47–52 (2011).
74. Wang, L. et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* **365**, 2497–2506 (2011).
75. Furney, S. J. et al. SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov.* **3**, 1122–1129 (2013).
76. Rose, A. M. et al. Detection of mutations in SF3B1, EIF1AX and GNAQ in primary orbital melanoma by candidate gene analysis. *BMC Cancer* **18**, 1–9 (2018).
77. Darman, R. B. et al. Cancer-associated SF3B1 hotspot mutations induce cryptic 3′ splice site selection through use of a different branch point. *Cell Rep.* **13**, 1033–1045 (2015).
78. Shiozawa, Y. et al. Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia. *Nat. Commun.* **9**, 3649 (2018).
79. Zhang, J. et al. Disease-causing mutations in SF3B1 alter splicing by disrupting interaction with SUGP1. *Mol. Cell* **76**, 82–95 (2019).
80. Inoue, D. et al. Spliceosomal disruption of the non-canonical BAF complex in cancer. *Nature* **574**, 432–436 (2019).
**This study shows that mutations in SF3B1 converge on repression of the tumour suppressor BRD9 by activation of a poison exon**.
81. Dolatshad, H. et al. Cryptic splicing events in the iron transporter ABCB7 and other key target genes in SF3B1-mutant myelodysplastic syndromes. *Leukemia* **30**, 2322–2331 (2016).
82. Lee, S. C. W. et al. Synthetic lethal and convergent biological effects of cancer-associated spliceosomal gene mutations. *Cancer Cell* **34**, 225–241 (2018).
**This study explains the synthetic lethality of splicing mutations in myelodydplastic syndrome as a consequence of aberrant splicing and downregulation of regulators of haematopoietic stem cell survival and quiescence**.
83. Wang, L. et al. Transcriptomic characterization of SF3B1 mutation reveals its pleiotropic effects in chronic lymphocytic leukemia. *Cancer Cell* **30**, 750–763 (2016).
84. Desterro, J., Bak-Gordon, P. & Carmo-Fonseca, M. Targeting mRNA processing as an anticancer strategy. *Nat. Rev. Drug. Discov.* **19**, 112–129 (2020).
85. Corrionero, A., Miñana, B. & Valcárcel, J. Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. *Genes. Dev.* **25**, 445–459 (2011).
86. Xargay-Torrent, S. et al. The splicing modulator sudemycin induces a specific antitumor response and cooperates with ibrutinib in chronic lymphocytic leukemia. *Oncotarget* **6**, 22734–22749 (2015).
87. Larrayoz, M. et al. The SF3B1 inhibitor spliceostatin A (SSA) elicits apoptosis in chronic lymphocytic leukaemia cells through downregulation of Mcl-1. *Leukemia* **30**, 351–360 (2016).
88. Gao, Y. & Koide, K. Chemical perturbation of Mcl-1 pre-mRNA splicing to induce apoptosis in cancer cells. *ACS Chem. Biol.* **8**, 895–900 (2013).
89. Vigevani, L., Gohr, A., Webb, T., Irimia, M. & Valcárcel, J. Molecular basis of differential 3′ splice site sensitivity to anti-tumor drugs targeting U2 snRNP. *Nat. Commun.* **8**, 2100 (2017).
90. Han, T. et al. Anticancer sulfonamides target splicing by inducing RBM39 degradation via recruitment to DCAF15. *Science* **356**, eaal3755 (2017).
**This study reports a splicing-based mechanism for the anticancer activity of sulfonamides and highlights the relevance of controlling RBM39 levels in haematopoietic and lymphoid cancer cell lineages**.
91. Wang, E. et al. Targeting an RNA-binding protein network in acute myeloid leukemia. *Cancer Cell* **35**, 369–384 (2019).
92. Jayasinghe, R. G. et al. Systematic analysis of splice-site-creating mutations in cancer. *Cell Rep.* **23**, 270–281 (2018).
93. Kahles, A. et al. Comprehensive analysis of alternative splicing across tumors from 8,705 patients. *Cancer Cell* **34**, 211–224 (2018).
**This study reveals extensive generation of tumour-specific alternative splicing events with the potential to generate MHC class I-binding neoantigen peptides**.
94. Li, L. et al. Targeting the ERG oncogene with splice-switching oligonucleotides as a novel therapeutic strategy in prostate cancer. *Br. J. Cancer* **123**, 1024–1032 (2020).
95. Charenton, C., Wilkinson, M. E. & Nagai, K. Mechanism of 5′ splice site transfer for human spliceosome activation. *Science* **364**, 362–367 (2019).
96. Sharma, S., Wongpalee, S. P., Vashisht, A., Wohlschlegel, J. A. & Black, D. L. Stem–loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. *Genes. Dev.* **28**, 2518–2531 (2014).
97. Martelly, W., Fellows, B., Senior, K., Marlowe, T. & Sharma, S. Identification of a noncanonical RNA binding domain in the U2 snRNP protein SF3A1. *RNA* **25**, 1509–1521 (2019).
98. Martelly, W. et al. Synergistic roles for human U1 snRNA stem–loops in pre-mRNA splicing. *RNA Biol.* **18**, 2576–2593 (2021).
99. Abovich, N. & Rosbash, M. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89**, 403–412 (1997).
100. Becerra, S., Andrés-León, E., Prieto-Sánchez, S., Hernández-Munain, C. & Suñé, C. Prp40 and early events in splice site definition. *Wiley Interdiscip. Rev. RNA* **7**, 17–32 (2016).
101. De Conti, L., Baralle, M. & Buratti, E. Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* **4**, 49–60 (2013).
102. Schneider, M. et al. Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. *Mol. Cell* **38**, 223–235 (2010).
103. Robberson, B. L., Cote, G. J. & Berget, S. M. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **10**, 84–94 (1990).
104. Berget, S. M. Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270**, 2411–2414 (1995).
105. Gonatopoulos-Pournatzis, T. & Blencowe, B. J. Microexons: at the nexus of nervous system development, behaviour and autism spectrum disorder. *Curr. Opin. Genet. Dev.* **65**, 22–33 (2020).
106. Gonatopoulos-Pournatzis, T. et al. Autism-misregulated eIF4G microexons control synaptic translation and higher order cognitive functions. *Mol. Cell* **77**, e16 (2020).
107. Parras, A. et al. Autism-like phenotype and risk gene mRNA deadenylation by CPEB4 mis-splicing. *Nature* **560**, 441–446 (2018).
**Gonatopoulos-Pournatzis et al. and Parras et al. reveal a link between misregulation of microexons in the cap-binding translation factor eIF4G and autism disorder**.

108. Ule, J. & Blencowe, B. J. Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Mol. Cell* **76**, 329–345 (2019).
109. Gonatopoulos-Pournatzis, T. et al. Genome-wide CRISPR–Cas9 interrogation of splicing networks reveals a mechanism for recognition of autism-misregulated neuronal microexons. *Mol. Cell* **72**, 510–524 (2018).
110. Choudhary, B., Marx, O. & Norris, A. D. Spliceosomal component PRP-40 is a central regulator of microexon splicing. *Cell Rep.* **36**, 109464 (2021).
111. Salz, H. K. Sex determination in insects: a binary decision based on alternative splicing. *Curr. Opin. Genet. Dev.* **21**, 395–400 (2011).
112. Ule, J. et al. Nova regulates brain-specific splicing to shape the synapse. *Nat. Genet.* **37**, 844–852 (2005).
113. Roy Burman, D., Das, S., Das, C. & Bhattacharya, R. Alternative splicing modulates cancer aggressiveness: role in EMT/metastasis and chemoresistance. *Mol. Biol. Rep.* **48**, 897–914 (2021).
114. Ke, S. et al. Saturation mutagenesis reveals manifold determinants of exon definition. *Genome Res.* **28**, 11–24 (2018).
115. Fu, X. D. & Ares, M. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* **15**, 689–701 (2014).
116. Wu, J. Y. & Maniatis, T. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**, 1061–1070 (1993).
117. Jobbins, A. M. et al. The mechanisms of a mammalian splicing enhancer. *Nucleic Acids Res.* **46**, 2145–2158 (2018).
118. Clerte, C. & Hall, K. B. Characterization of multimeric complexes formed by the human PTB1 protein on RNA. *RNA* **12**, 457–475 (2006).
119. Braun, J. E., Friedman, L. J., Gelles, J. & Moore, M. J. Synergistic assembly of human pre-spliceosomes across introns and exons. *eLife* **7**, 1–18 (2018).
120. Sharma, S., Maris, C., Allain, F. H. T. & Black, D. L. U1 snRNA directly interacts with polypyrimidine tract-binding protein during splicing repression. *Mol. Cell* **41**, 579–588 (2011).
121. Gueroussov, S. et al. Regulatory expansion in mammals of multivalent hnRNP assemblies that globally control alternative splicing. *Cell* **170**, 324–339 (2017).
122. Ying, Y. et al. Splicing activation by Rbfox requires self-aggregation through its tyrosine-rich domain. *Cell* **170**, 312–323 (2017).
**Gueroussov et al. and Ying et al. illustrate the importance of intrinsically disordered domains mediating higher order multivalent assemblies for the function of various families of splicing regulatory factors**.
123. Witten, J. T. & Ule, J. Understanding splicing regulation through RNA splicing maps. *Trends Genet.* **27**, 89–97 (2011).
124. Yee, B. A., Pratt, G. A., Graveley, B. R., van Nostrand, E. L. & Yeo, G. W. RBP-Maps enables robust generation of splicing regulatory maps. *RNA* **25**, 193–204 (2019).
125. Mikl, M., Hamburg, A., Pilpel, Y. & Segal, E. Dissecting splicing decisions and cell-to-cell variability with designed sequence libraries. *Nat. Commun.* **10**, 1–14 (2019).
126. Julien, P., Miñana, B., Baeza-Centurion, P., Valcárcel, J. & Lehner, B. The complete local genotype–phenotype landscape for the alternative splicing of a human exon. *Nat. Commun.* **7**, 1–8 (2016).
127. Baeza-Centurion, P., Miñana, B., Schmiedel, J. M., Valcárcel, J. & Lehner, B. Combinatorial genetics reveals a scaling law for the effects of mutations on splicing. *Cell* **176**, 549–563 (2019).
**Deep mutagenesis of an alternatively spliced exon reveals that the effect of a mutation is maximal at intermediate initial levels of exon inclusion**.
128. Baeza-Centurion, P., Miñana, B., Valcárcel, J. & Lehner, B. Mutations primarily alter the inclusion of alternatively spliced exons. *eLife* **9**, 1–74 (2020).
129. Souček, P. et al. High-throughput analysis revealed mutations' diverging effects on SMN1 exon 7 splicing. *RNA Biol.* **16**, 1364–1376 (2019).
130. Braun, S. et al. Decoding a cancer-relevant splicing decision in the RON proto-oncogene using high-throughput mutagenesis. *Nat. Commun.* **9**, 3315 (2018).
131. Soemedi, R. et al. Pathogenic variants that alter protein code often disrupt splicing. *Nat. Genet.* **49**, 848–855 (2017).
132. Rong, S. et al. Mutational bias and the protein code shape the evolution of splicing enhancers. *Nat. Commun.* **11**, 1–10 (2020).
133. Mueller, W. F., Larsen, L. S. Z., Garibaldi, A., Hatfield, G. W. & Hertel, K. J. The silent sway of splicing by synonymous substitutions. *J. Biol. Chem.* **290**, 27700–27711 (2015).
134. Dufner-Almeida, L. G., do Carmo, R. T., Masotti, C. & Haddad, L. A. Understanding human DNA variants affecting pre-mRNA splicing in the NGS era. *Adv. Genet.* **103**, 39–90 (2019).
135. Cummings, B. B. et al. Transcript expression-aware annotation improves rare variant interpretation. *Nature* **581**, 452–458 (2020).
136. Supek, F., Miñana, B., Valcárcel, J., Gabaldón, T. & Lehner, B. Synonymous mutations frequently act as driver mutations in human cancers. *Cell* **156**, 1324–1335 (2014).
137. Sharma, Y. et al. A pan-cancer analysis of synonymous mutations. *Nat. Commun.* **10**, 2569 (2019).
138. Rosenberg, A. B., Patwardhan, R. P., Shendure, J. & Seelig, G. Learning the sequence determinants of alternative splicing from millions of random sequences. *Cell* **163**, 698–711 (2015).
139. Ke, S. et al. Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res.* **21**, 1360–1374 (2011).
140. Arias, M. A., Lubkin, A. & Chasin, L. A. Splicing of designer exons informs a biophysical model for exon definition. *RNA* **21**, 213–229 (2015).
141. Jaganathan, K. et al. Predicting splicing from primary sequence with deep learning. *Cell* **176**, 535–548 (2019).
**This study developed a deep neural network to predict splice site utilization, including cryptic splice sites induced by genetic variants associated with autism and intellectual disability**.
142. Bao, S., Moakley, D. F. & Zhang, C. The splicing code goes deep. *Cell* **176**, 414–416 (2019).
143. Cheng, J. et al. MMSplice: modular modeling improves the predictions of genetic variant effects on splicing. *Genome Biol.* **20**, 1–15 (2019).
144. Rhine, C. L. et al. Future directions for high-throughput splicing assays in precision medicine. *Hum. Mutat.* **40**, 1225–1234 (2019).
145. Riepe, T. V., Khan, M., Roosing, S., Cremers, F. P. M. & t Hoen, P. A. C. Benchmarking deep learning splice prediction tools using functional splice assays. *Hum. Mutat.* **42**, 799–810 (2021).
146. Cheung, R. et al. A multiplexed assay for exon recognition reveals that an unappreciated fraction of rare genetic variants cause large-effect splicing disruptions. *Mol. Cell* **73**, 183–194 (2019).
147. Lim, K. R. Q., Maruyama, R. & Yokota, T. Eteplirsen in the treatment of Duchenne muscular dystrophy. *Drug. Des. Devel. Ther.* **11**, 533–545 (2017).
148. Beyer, A. L. & Osheim, Y. N. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes. Dev.* **2**, 754–765 (1988).
149. Carrillo Oesterreich, F. et al. Splicing of nascent RNA coincides with intron exit from RNA polymerase II. *Cell* **165**, 372–381 (2016).
150. Neugebauer, K. M. Nascent RNA and the coordination of splicing with transcription. *Cold Spring Harb. Perspect. Biol.* **11**, a032227 (2019).
151. Custódio, N. & Carmo-Fonseca, M. Co-transcriptional splicing and the CTD code. *Crit. Rev. Biochem. Mol. Biol.* **51**, 395–411 (2016).
152. Moon, K. H., Zhao, X. & Yu, Y. T. Pre-mRNA splicing in the nuclei of *Xenopus* oocytes. *Methods Mol. Biol.* **322**, 149–163 (2006).
153. Herzel, L., Ottoz, D. S. M., Alpert, T. & Neugebauer, K. M. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat. Rev. Mol. Cell Biol.* **18**, 637–650 (2017).
154. Anvar, S. Y. et al. Full-length mRNA sequencing uncovers a widespread coupling between transcription initiation and mRNA processing. *Genome Biol.* **19**, 1–18 (2018).
155. Giono, L. E. & Kornblihtt, A. R. Linking transcription, RNA polymerase II elongation and alternative splicing. *Biochem. J.* **477**, 3091–3104 (2020).
156. Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* **15**, 163–175 (2014).
157. Dujardin, G. et al. How slow RNA polymerase II elongation favors alternative exon skipping. *Mol. Cell* **54**, 683–690 (2014).
158. Godoy Herz, M. A. et al. Light regulates plant alternative splicing through the control of transcriptional elongation. *Mol. Cell* **73**, 1066–1074 (2019).
159. Maslon, M. M. et al. A slow transcription rate causes embryonic lethality and perturbs kinetic coupling of neuronal genes. *EMBO J.* **38**, 1–18 (2019).
160. Chathoth, K. T., Barrass, J. D., Webb, S. & Beggs, J. D. A splicing-dependent transcriptional checkpoint associated with prespliceosome formation. *Mol. Cell* **53**, 779–790 (2014).
161. Fiszbein, A., Krick, K. S., Begg, B. E. & Burge, C. B. Exon-mediated activation of transcription starts. *Cell* **179**, 1551–1565 (2019).
**This study provides evidence for the general mechanisms by which alternative splicing influences alternative transcription initiation**.
162. Reimer, K. A., Mimoso, C. A., Adelman, K. & Neugebauer, K. M. Co-transcriptional splicing regulates 3′ end cleavage during mammalian erythropoiesis. *Mol. Cell* **81**, 998–1012 (2021).
163. Bedi, K. et al. Cotranscriptional splicing efficiencies differ within genes and between cell types. *RNA* **27**, 829–840 (2021).
164. Drexler, H. L., Choquet, K. & Churchman, L. S. Splicing kinetics and coordination revealed by direct nascent RNA sequencing through nanopores. *Mol. Cell* **77**, 985–998 (2020).
165. Sousa-Luís, R. et al. POINT technology illuminates the processing of polymerase-associated intact nascent transcripts. *Mol. Cell* **81**, 1935–1950 (2021).
166. Tilgner, H. et al. Microfluidic isoform sequencing shows widespread splicing coordination in the human transcriptome. *Genome Res.* **28**, 231–242 (2018).
167. Gupta, I. et al. Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat. Biotechnol.* **36**, 1197–1202 (2018).
**This technique allows for high-resolution discovery and analysis of alternative splicing in individual cells from complex tissues**.
168. Gordon, J. M., Phizicky, D. V. & Neugebauer, K. M. Nuclear mechanisms of gene expression control: pre-mRNA splicing as a life or death decision. *Curr. Opin. Genet. Dev.* **67**, 67–76 (2021).
169. Boutz, P. L., Bhutkar, A. & Sharp, P. A. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes. Dev.* **29**, 63–80 (2015).
170. Naro, C. et al. An orchestrated intron retention program in meiosis controls timely usage of transcripts during germ cell differentiation. *Dev. Cell* **41**, 82–93 (2017).
171. Mauger, O., Lemoine, F. & Scheiffele, P. Targeted intron retention and excision for rapid gene regulation in response to neuronal activity. *Neuron* **92**, 1266–1278 (2016).
172. Braun, C. J. et al. Coordinated splicing of regulatory detained introns within oncogenic transcripts creates an exploitable vulnerability in malignant glioma. *Cancer Cell* **32**, 411–426 (2017).
**This study demonstrates the relevance of a splicing regulatory programme affecting detained introns in cancer cells and its potential targeting using inihibitors of argininine methylation enzymes**.

173. Galganski, L., Urbanek, M. O. & Krzyzosiak, W. J. Nuclear speckles: molecular organization, biological function and role in disease. *Nucleic Acids Res.* **45**, 10350–10368 (2017).
174. Ilık, İ. A. et al. SON and SRRM2 are essential for nuclear speckle formation. *eLife* **9**, 1–48 (2020).
175. Ilık, İ. A. & Aktaş, T. Nuclear speckles: dynamic hubs of gene expression regulation. *FEBS J.* https://doi.org/10.1111/febs.16117 (2021).
176. Girard, C. et al. Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat. Commun.* **3**, 994 (2012).
177. Dias, A. P., Dufu, K., Lei, H. & Reed, R. A role for TREX components in the release of spliced mRNA from nuclear speckle domains. *Nat. Commun.* **1**, 97 (2010).
178. Guo, Y. E. et al. Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543–548 (2019).
**This study reveals the existence of functionally distinct nuclear condensates involved in transcription initiation and RNA processing and the role of protein phosphorylation in the switch between them**.
179. Barutcu, A. R. et al. Systematic mapping of nuclear domain-associated transcripts reveals speckles and lamina as hubs of functionally distinct retained introns. *Mol. Cell* **82**, 1035–1052 (2022).
180. Tammer, L. et al. Gene architecture directs splicing outcome in separate nuclear spatial regions. *Mol. Cell* **82**, 1021–1034 (2022).
**Barutcu et al. and Tammer et al. reveal that the nuclear sub-localization of transcription influences alternative splicing decisions**.
181. Amit, M. et al. Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. *Cell Rep.* **1**, 543–556 (2012).
182. Iannone, C. & Valcárcel, J. Chromatin's thread to alternative splicing regulation. *Chromosoma* **122**, 465–474 (2013).
183. de Almeida, S. F. & Carmo-Fonseca, M. Reciprocal regulatory links between cotranscriptional splicing and chromatin. *Semin. Cell Dev. Biol.* **32**, 2–10 (2014).
184. Shukla, S. et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74–79 (2011).
185. Luco, R. F. et al. Regulation of alternative splicing by histone modifications. *Science* **327**, 996–1000 (2010).
186. Segelle, A. et al. Histone marks regulate the epithelial-to-mesenchymal transition via alternative splicing. *Cell Rep.* **38**, 110357 (2022).
**An early demonstration that modification of chromatin epigenetic marks can directly affect splice site choice**.
187. Marasco, L. E. et al. Counteracting chromatin effects of a splicing-correcting antisense oligonucleotide improves its therapeutic efficacy in spinal muscular atrophy. *Cell* **185**, 2057–2070 (2022).
188. Helm, M. & Motorin, Y. Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.* **18**, 275–291 (2017).
189. Zhao, B. S., Roundtree, I. A. & He, C. Post-transcriptional gene regulation by mRNA modifications. *Nat. Rev. Mol. Cell Biol.* **18**, 31–42 (2016).
190. Dawson, M. A. & Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell* **150**, 12–27 (2012).
191. Mendel, M. et al. Splice site $m^6A$ methylation prevents binding of U2AF35 to inhibit RNA splicing. *Cell* **184**, 1–18 (2021).
192. Pendleton, K. E. et al. The U6 snRNA $m^6A$ methyltransferase METTL16 regulates SAM synthetase intron retention. *Cell* **169**, 824–835 (2017).
**Mendel et al. and Pendleton et al. link $m^6A$ modification of pre-mRNAs with splicing regulation events important for *S*-adenosylmethionine homeostasis**.
193. Zhou, K. I. et al. Regulation of co-transcriptional pre-mRNA splicing by $m^6A$ through the low-complexity protein hnRNPG. *Mol. Cell* **76**, 70–81 (2019).
194. Xiao, W. et al. Nuclear $m^6A$ reader YTHDC1 regulates mRNA splicing. *Mol. Cell* **61**, 507–519 (2016).
195. Kasowitz, S. D. et al. Nuclear $m^6A$ reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. *PLoS Genet.* **14**, 1–28 (2018).
196. Louloupi, A., Ntini, E., Conrad, T. & Ørom, U. A. V. Transient *N*-6-methyladenosine transcriptome sequencing reveals a regulatory role of $m^6A$ in splicing efficiency. *Cell Rep.* **23**, 3429–3437 (2018).
197. Ke, S. et al. $m^6A$ mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes. Dev.* **31**, 990–1006 (2017).
198. Martinez, N. M. et al. Pseudouridine synthases modify human pre-mRNA co-transcriptionally and affect pre-mRNA processing. *Mol. Cell* **82**, 645–659 (2022).
199. Morais, P., Adachi, H. & Yu, Y.-T. Spliceosomal snRNA epitranscriptomics. *Front. Genet.* **12**, 652129 (2021).
200. Goh, Y. T., Koh, C. W. Q., Sim, D. Y., Roca, X. & Goh, W. S. S. METTL4 catalyzes $m^6Am$ methylation in U2 snRNA to regulate pre-mRNA splicing. *Nucleic Acids Res.* **48**, 9250–9261 (2020).
201. Lu, Z. & Chang, H. Y. Decoding the RNA structurome. *Curr. Opin. Struct. Biol.* **36**, 142–148 (2016).
202. Sun, L. et al. RNA structure maps across mammalian cellular compartments. *Nat. Struct. Mol. Biol.* **26**, 322–330 (2019).
203. Saldi, T., Riemondy, K., Erickson, B. & Bentley, D. L. Alternative RNA structures formed during transcription depend on elongation rate and modify RNA processing. *Mol. Cell* **81**, 1789–1801 (2021).
204. Bartys, N., Kierzek, R. & Lisowiec-Wachnicka, J. The regulation properties of RNA secondary structure in alternative splicing. *Biochim. Biophys. Acta — Gene Regul. Mech.* **1862**, 194401 (2019).
205. Xu, B., Meng, Y. & Jin, Y. RNA structures in alternative splicing and back-splicing. *Wiley Interdiscip. Rev. RNA* **12**, 1–39 (2021).
206. Meyer, M., Plass, M., Pérez-Valle, J., Eyras, E. & Vilardell, J. Deciphering 3'ss selection in the yeast genome reveals an RNA thermosensor that mediates alternative splicing. *Mol. Cell* **43**, 1033–1039 (2011).
207. Kalmykova, S. et al. Conserved long-range base pairings are associated with pre-mRNA processing of human genes. *Nat. Commun.* **12**, 1–17 (2021).
208. Warf, M. B., Diegel, J. V., Von Hippel, P. H. & Berglund, J. A. The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. *Proc. Natl Acad. Sci. USA* **106**, 9203–9208 (2009).
209. Varani, L. et al. Structure of tau exon 10 splicing regulatory element RNA and destabilization by mutations of frontotemporal dementia and parkinsonism linked to chromosome 17. *Proc. Natl Acad. Sci. USA* **96**, 8229–8234 (1999).
210. Kar, A. et al. RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem–loop structure at the 5' splice site. *Mol. Cell. Biol.* **31**, 1812–1821 (2011).
211. Singh, N. N. & Singh, R. N. How RNA structure dictates the usage of a critical exon of spinal muscular atrophy gene. *Biochim. Biophys. Acta—Gene Regul. Mech.* **1862**, 194403 (2019).
212. Georgakopoulos-Soares, I. et al. Alternative splicing modulation by G-quadruplexes. *Nat. Commun.* **13**, 2404 (2022).
213. Huang, H., Zhang, J., Harvey, S. E., Hu, X. & Cheng, C. RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. *Genes. Dev.* **31**, 2296–2309 (2017).
214. Fish, L. et al. A prometastatic splicing program regulated by SNRPA1 interactions with structured RNA elements. *Science* **372**, eabc7531 (2021).
**Early evidence that a structural splicing enhancer coordinates an exon inclusion programme relevant for highly metastatic cancer cells. This regulatory element is activated by a non-canonical function of the snRNP protein SNRPA1**.
215. Dubin, R. A., Kazmi, M. A. & Ostrer, H. Inverted repeats are necessary for circularization of the mouse testis Sry transcript. *Gene* **167**, 245–248 (1995).
216. Jeck, W. R. et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**, 426 (2013).
217. Ashwal-Fluss, R. et al. CircRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell* **56**, 55–66 (2014).
218. Zhang, X. O. et al. Complementary sequence-mediated exon circularization. *Cell* **159**, 134–147 (2014).
219. Liang, D. & Wilusz, J. E. Short intronic repeat sequences facilitate circular RNA production. *Genes. Dev.* **28**, 2233–2247 (2014).
220. Ivanov, A. et al. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep.* **10**, 170–177 (2015).
221. Gruhl, F., Janich, P., Kaessmann, H. & Gatfield, D. Circular RNA repertoires are associated with evolutionarily young transposable elements. *eLife* **10**, 1–33 (2021).
222. Santos-Rodriguez, G., Voineagu, I. & Weatheritt, R. J. Evolutionary dynamics of circular rnas in primates. *eLife* **10**, 1–22 (2021).
223. Xu, C. & Zhang, J. Mammalian circular RNAs result largely from splicing errors. *Cell Rep.* **36**, 109439 (2021).
224. Du, W. W. et al. Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. *Nucleic Acids Res.* **44**, 2846–2858 (2016).
225. Hansen, T. B. et al. Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–388 (2013).
226. Piwecka, M. et al. Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. *Science* **357**, eaam8526 (2017).
227. Li, S. et al. Screening for functional circular RNAs using the CRISPR–Cas13 system. *Nat. Methods* **18**, 51–59 (2021).
228. Pamudurti, N. R. et al. Translation of CircRNAs. *Mol. Cell* **66**, 9–21 (2017).
229. Legnini, I. et al. Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Mol. Cell* **66**, 22–37 (2017).
230. Weigelt, C. M. et al. An insulin-sensitive circular RNA that regulates lifespan in *Drosophila*. *Mol. Cell* **79**, 268–279 (2020).
231. Hansen, T. B. Signal and noise in circRNA translation. *Methods* **196**, 68–73 (2021).
232. Yang, Y. et al. Extensive translation of circular RNAs driven by $N^6$-methyladenosine. *Cell Res.* **27**, 626–641 (2017).
233. Joglekar, A. et al. A spatially resolved brain region- and cell type-specific isoform atlas of the postnatal mouse brain. *Nat. Commun.* **12**, 1–16 (2021).
234. Dominguez, D. et al. Sequence, structure, and context preferences of human RNA binding proteins. *Mol. Cell* **70**, 854–867 (2018).
235. Van Nostrand, E. L. et al. A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).
**A massive effort to identify binding sites on RNA and chromatin for 356 RNA binding proteins and to determine the functional implications for RNA stability, splicing regulation and RNA localization**.
236. Van Nostrand, E. L. et al. Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biol.* **21**, 1–26 (2020).
237. Feng, H. et al. Complexity and graded regulation of neuronal cell-type-specific alternative splicing revealed by single-cell RNA sequencing. *Proc. Natl Acad. Sci. USA* **118**, 1–12 (2021).
238. Shalek, A. K. et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236–240 (2013).
239. Song, Y. et al. Single-cell alternative splicing analysis with expedition reveals splicing dynamics during neuron differentiation. *Mol. Cell* **67**, 148–161 (2017).

240. Liu, W. & Zhang, X. Single-cell alternative splicing analysis reveals dominance of single transcript variant. *Genomics* **112**, 2418–2425 (2020).
241. Huang, Y. & Sanguinetti, G. BRIE2: computational identification of splicing phenotypes from single-cell transcriptomic experiments. *Genome Biol.* **22**, 1–15 (2021).
242. Linker, S. M. et al. Combined single-cell profiling of expression and DNA methylation reveals splicing regulation and heterogeneity. *Genome Biol.* **20**, 1–14 (2019).
243. Kim, H. S., Grimes, S. M., Hooker, A. C., Lau, B. T. & Ji, H. P. Single-cell characterization of CRISPR-modified transcript isoforms with nanopore sequencing. *Genome Biol.* **22**, 1–16 (2021).
244. Thompson, M. et al. Splicing in a single neuron is coordinately controlled by RNA binding proteins and transcription factors. *eLife* **8**, 1–19 (2019).
245. Lukačišin, M., Espinosa-Cantú, A. & Bollenbach, T. Intron-mediated induction of phenotypic heterogeneity. *Nature* **605**, 113–118 (2022).
246. Wan, Y. et al. Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. *Cell* **184**, 2878–2895 (2021).
**This study reports a quasi-genome-scale platform for observing the synthesis and procesing kinetics of single nascent RNA molecules in real time; it reveals large kinetic variation of single intron removal in single cells and widespread stochastic recursive splicing within introns**.
247. Martin, R. M., Rino, J., Carvalho, C., Kirchhausen, T. & Carmo-Fonseca, M. Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. *Cell Rep.* **4**, 1144–1155 (2013).
248. Burnette, J. M., Miyamoto-Sato, E., Schaub, M. A., Conklin, J. & Lopez, A. J. Subdivision of large introns in *Drosophila* by recursive splicing at nonexonic elements. *Genetics* **170**, 661–674 (2005).
249. Sibley, C. R. et al. Recursive splicing in long vertebrate genes. *Nature* **521**, 371–375 (2015).
250. Kim, S. W. et al. Widespread intra-dependencies in the removal of introns from human transcripts. *Nucleic Acids Res.* **45**, 9503–9513 (2017).
251. Jangi, M. & Sharp, P. A. Building robust transcriptomes with master splicing factors. *Cell* **159**, 487–498 (2014).
252. Kokic, G., Wagner, F. R., Chernev, A., Urlaub, H. & Cramer, P. Structural basis of human transcription–DNA repair coupling. *Nature* **598**, 368–372 (2021).
253. Zhang, S. et al. Structure of a transcribing RNA polymerase II–U1 snRNP complex. *Science* **371**, 305–309 (2021).
254. Haselbach, D. et al. Structure and conformational dynamics of the human spliceosomal bact complex. *Cell* **172**, 454–464 (2018).
255. Sebbag-Sznajder, N. et al. Dynamic supraspliceosomes are assembled on different transcripts regardless of their intron number and splicing state. *Front. Genet.* **11**, 1–14 (2020).
256. Papasaikas, P., Tejedor, J. R., Vigevani, L. & Valcárcel, J. Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. *Mol. Cell* **57**, 7–22 (2015).
257. Statello, L., Guo, C. J., Chen, L. L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118 (2021).
258. Deveson, I. W. et al. Universal alternative splicing of noncoding exons. *Cell Syst.* **6**, 245–255 (2018).
259. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
260. North, K. et al. Synthetic introns enable splicing factor mutation-dependent targeting of cancer cells. *Nat. Biotechnol.* **40**, 1103–1113 (2022).
**In this study, splicing alterations characteristic of cancer cells are leveraged to engineer synthetic introns that are specifically spliced in and induce the death of cancer cells**.
261. Zhang, X. et al. Structures of the human spliceosomes before and after release of the ligated exon. *Cell Res.* **29**, 274–285 (2019).
262. Tholen, J., Razew, M., Weis, F. & Galej, W. P. Structural basis of branch site recognition by the human spliceosome. *Science* **375**, 50–57 (2022).
263. Zhan, X., Yan, C., Zhang, X., Lei, J. & Shi, Y. Structures of the human pre-catalytic spliceosome and its precursor spliceosome. *Cell Res.* **28**, 1129–1140 (2018).
264. Li, X. et al. A unified mechanism for intron and exon definition and back-splicing. *Nature* **573**, 375–380 (2019).
265. Zhan, X., Yan, C., Zhang, X., Lei, J. & Shi, Y. Structure of a human catalytic step I spliceosome. *Science* **359**, 537–545 (2018).
266. Bertram, K. et al. Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* **542**, 318–323 (2017).
267. Fica, S. M., Oubridge, C., Wilkinson, M. E., Newman, A. J. & Nagai, K. A human postcatalytic spliceosome structure reveals essential roles of metazoan factors for exon ligation. *Science* **363**, 710–714 (2019).
268. Townsend, C. et al. Mechanism of protein-guided folding of the active site U2/U6 RNA during spliceosome activation. *Science* **370**, eabc3753 (2020).
269. Shapiro, M. B. & Senapathy, P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**, 7155–7174 (1987).
270. Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. Improved splice site detection in Genie. *J. Comput. Biol.* **4**, 311–323 (1997).
271. Pertea, M., Lin, X. & Salzberg, S. L. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res.* **29**, 1185–1190 (2001).
272. Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377–394 (2004).
273. Castelo, R. & Guigó, R. Splice site identification by idlBNs. *Bioinformatics* **20**, 69–76 (2004).
274. Barash, Y. et al. Deciphering the splicing code. *Nature* **465**, 53–59 (2010).
275. Xiong, H. Y. et al. The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806 (2015).
276. Jha, A., Gazzara, M. R. & Barash, Y. Integrative deep models for alternative splicing. *Bioinformatics* **33**, i274–i282 (2017).
277. Jha, A. et al. Enhanced integrated gradients: improving interpretability of deep learning models using splicing codes as a case study. *Genome Biol.* **21**, 1–22 (2020).
278. Zuallaert, J. et al. Splicerover: interpretable convolutional neural networks for improved splice site prediction. *Bioinformatics* **34**, 4180–4188 (2018).
279. Bretschneider, H., Gandhi, S., Deshwar, A. G., Zuberi, K. & Frey, B. J. COSSMO: predicting competitive alternative splice site selection using deep learning. *Bioinformatics* **34**, i429–i437 (2018).
280. Cheng, J., Çelik, M. H., Kundaje, A. & Gagneur, J. MTSplice predicts effects of genetic variants on tissue-specific splicing. *Genome Biol.* **22**, 1–19 (2021).
281. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
282. Rentzsch, P., Schubach, M., Shendure, J. & Kircher, M. CADD-Splice — improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med.* **13**, 1–12 (2021).
283. Zeng, T. & Li, Y. I. Predicting RNA splicing from DNA sequence using Pangolin. *Genome Biol.* **23**, 103 (2022).
284. Zhang, X. H. F., Arias, M. A., Shengdong, K. E. & Chasin, L. A. Splicing of designer exons reveals unexpected complexity in pre-mRNA splicing. *RNA* **15**, 367–376 (2009).
285. Schirman, D., Yakhini, Z., Pilpel, Y. & Dahan, O. A broad analysis of splicing regulation in yeast using a large library of synthetic introns. *PLOS Genet.* **17**, e1009805 (2021).
286. Adamson, S. I., Zhan, L. & Graveley, B. R. Vex-seq: high-throughput identification of the impact of genetic variation on pre-mRNA splicing efficiency. *Genome Biol.* **19**, 1–12 (2018).
287. Mount, S. M. et al. Assessing predictions of the impact of variants on splicing in CAGI5. *Hum. Mutat.* **40**, 1215–1224 (2019).
288. Tubeuf, H. et al. Large-scale comparative evaluation of user-friendly tools for predicting variant-induced alterations of splicing regulatory elements. *Hum. Mutat.* **41**, 1811–1829 (2020).
289. Cortés-López, M. et al. High-throughput mutagenesis identifies mutations and RNA-binding proteins controlling CD19 splicing and CART-19 therapy resistance. *Nat. Commun.* **13**, 5570 (2022).
290. Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R. & Misteli, T. Epigenetics in alternative pre-mRNA splicing. *Cell* **144**, 16–26 (2011).
291. Schwartz, S., Meshorer, E. & Ast, G. Chromatin organization marks exon–intron structure. *Nat. Struct. Mol. Biol.* **16**, 990–995 (2009).
292. Tilgner, H. et al. Nucleosome positioning as a determinant of exon recognition. *Nat. Struct. Mol. Biol.* **16**, 996–1001 (2009).
293. Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C. & Komorowski, J. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res.* **19**, 1732–1741 (2009).
294. David, C. J., Boyne, A. R., Millhouse, S. R. & Manley, J. L. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65–Prp19 complex. *Genes. Dev.* **25**, 972–982 (2011).
295. Morgan, J. T., Fink, G. R. & Bartel, D. P. Excised linear introns regulate growth in yeast. *Nature* **565**, 606–611 (2019).
296. Parenteau, J. et al. Introns are mediators of cell response to starvation. *Nature* **565**, 612–617 (2019).
297. Mattick, J. S. & Gagen, M. J. The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol. Biol. Evol.* **18**, 1611–1630 (2001).
298. Dieci, G., Preti, M. & Montanini, B. Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics* **94**, 83–88 (2009).
299. Kelemen, O. et al. Function of alternative splicing. *Gene* **514**, 1–30 (2013).
300. Vuong, C. K., Black, D. L. & Zheng, S. The neurogenetics of alternative splicing. *Nat. Rev. Neurosci.* **17**, 265–281 (2016).
301. Tress, M. L., Abascal, F. & Valencia, A. Alternative splicing may not be the key to proteome complexity. *Trends Biochem. Sci.* **42**, 98–110 (2017).
302. Rodriguez, J. M., Pozo, F., Di Domenico, T., Vazquez, J. & Tress, M. L. An analysis of tissue-specific alternative splicing at the protein level. *PLoS Comput. Biol.* **16**, e1008287 (2020).
303. Weatheritt, R. J., Sterne-Weiler, T. & Blencowe, B. J. The ribosome-engaged landscape of alternative splicing. *Nat. Struct. Mol. Biol.* **23**, 1117–1123 (2016).
304. Ellis, J. D. et al. Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol. Cell* **46**, 884–892 (2012).
**High-throughput analyses of the impact of alternative splicing reveal widespread effects on the modulation of protein–protein interactions**.
305. Sinha, A. & Nagarajaram, H. A. Effect of alternative splicing on the degree centrality of nodes in protein–protein interaction networks of *Homo sapiens*. *J. Proteome Res.* **12**, 1980–1988 (2013).
306. Yang, Y. et al. Determination of a comprehensive alternative splicing regulatory network and combinatorial regulation by key factors during the epithelial-to-mesenchymal transition. *Mol. Cell. Biol.* **36**, 1704–1719 (2016).

307. Climente-González, H., Porta-Pardo, E., Godzik, A. & Eyras, E. The functional impact of alternative splicing in cancer. *Cell Rep.* **20**, 2215–2226 (2017).
308. Louadi, Z. et al. Functional enrichment of alternative splicing events with NEASE reveals insights into tissue identity and diseases. *Genome Biol.* **22**, 1–22 (2021).
309. Ezkurdia, I. et al. Comparative proteomics reveals a significant bias toward alternative protein isoforms with conserved structure and function. *Mol. Biol. Evol.* **29**, 2265–2283 (2012).
310. Buljan, M. et al. Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. *Curr. Opin. Struct. Biol.* **23**, 443–450 (2013).
311. Buljan, M. et al. Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol. Cell* **46**, 871–883 (2012).
312. Lewis, B. P., Green, R. E. & Brenner, S. E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl Acad. Sci. USA* **100**, 189–192 (2003).
313. Lindeboom, R. G. H., Supek, F. & Lehner, B. The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat. Genet.* **48**, 1112–1118 (2016).
314. Supek, F., Lehner, B. & Lindeboom, R. G. H. To NMD or not to NMD: nonsense-mediated mRNA decay in cancer and other genetic diseases. *Trends Genet.* **37**, 657–668 (2021).
315. Tapial, J. et al. An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res.* **27**, 1759–1768 (2017).
**This study provides a comprehensive database of alternative splicing in tissues and organisms and facilitates the identification of tissue-specific regulatory programmes**.
316. Schlautmann, L. P. & Gehring, N. H. A day in the life of the exon junction complex. *Biomolecules* **10**, 1–17 (2020).
317. Thomas, J. D. et al. RNA isoform screens uncover the essentiality and tumor-suppressor activity of ultraconserved poison exons. *Nat. Genet.* **52**, 84–94 (2020).
**This study reports a high-throughput functional analysis of poison exons, revealing their functions in cancer biology**.

## Acknowledgements
The authors thank M. Irimia, B. Lehner, members of our group and four reviewers for suggestions on the manuscript. M.E.R. was supported by a MSCA Postdoctoral Fellowship and C.V. by a FPI-Severo Ochoa PhD Fellowship from the Spanish Ministry of Economy and Competitiveness. Work in the authors' laboratory has been supported by the European Research Council, European Innovation Council, LaCaixa Health, Worldwide Cancer Research, AGAUR, Spanish Ministry of Economy and Competitiveness and the Centre of Excellence Severo Ochoa. The authors acknowledge support of the Spanish Ministry of Science and Innovation to the EMBL partnership and the CERCA Programme/Generalitat de Catalunya.

## Author contributions
All authors contributed to all aspects of the article.

## Competing interests
The authors declare no competing interests.

## Additional information
**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41576-022-00556-8.

**Correspondence** should be addressed to Juan Valcárcel.

**Peer review information** *Nature Reviews Genetics* thanks R. Luco, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



## Related links
**Genome Aggregation Database:** https://gnomad.broadinstitute.org/