MAJOR MESSAGES

Splice site consensus sequences do have sufficient information to uniquely direct spliceosomal transesterification events

Splicing enhancers and silencers appear to provide additional recognition information

“Exon definition” can enable the recognition of introns that are large

Splicing enhancers and silencers enable alternative splicing

Exome sequencing of tumors has revealed frequent mutations in the spliceosome. The mechanisms by which these mutations promote tumor formation is an outstanding question.

Hematopoietic tumors display the highest frequencies of spliceosome mutations with up to 80% of patients with myelodysplastic syndrome harboring mutations.

Mutations occur in specific hotspots in factors involved in 3’ splice site selection and always occur in the heterozygous state.

Spliceosome mutations in tumors alter splicing patterns but the effects are mutation-specific, raising a quandry

Several unifying hypotheses have been proposed for how spliceosome mutation promote tumor formation.

Spliceosome mutations in cancer offer a natural repository of potentially informative mutations that impact spliceosome function in human cells
The question: Why are there no described free-living eukaryotes that lack introns?

Model 1: Intron gain mechanisms outpace exceeds intron loss mechanisms (nonadaptive evolution).
Model 2: Evolution has found uses for splicing (adaptive evolution).

SPlicing REGULATORS, EXON DEFINITION, and SPlicing DECISIONS

Additional information in introns and exons: splicing enhancers and silencers

*S. cerevisiae* 5’ splice site and branchpoint sequences conform closely to the optimal consensus sequence. Beyond its close relatives (the hemiascomycetes), other eukaryotes (from protists to man) do not share this property.

Additional sequences that influences 5’ splice site, branchpoint and 3’ splice site usage exist in premRNAs. These splicing enhancers (ESEs, ISEs) and silencers (ESSs, ISSs) are bound by sequence-specific RNA binding proteins (RBPs). There about 1500 RBPs in humans. Perhaps ~100 are involved in splicing.

ESE, exonic splicing enhancer; ESS, exonic splicing silencer; ISE, intronic splicing enhancer; ISS, intronic splicing silencer.
**The exon definition model**
Berget and colleagues observed that the mutation of a 5' splice site could prevent the splicing of the upstream intron in vitro. Because exons are short (170 nts on average in humans) while introns are longer (median length of human introns is 1334 bp), she proposed that exons were the initial unit of recognition (“exon definition”) and that enabled subsequence recognition of the upstream intron (“intron definition”).

**The SR protein family -- major activators of splicing**
A major class of proteins that recognize ESEs in exons are so called SR proteins, named because they harbor serine-arginine dipeptides repeats in clusters (SRSF1-12). These are phosphorylated and seem to function as protein-protein interaction domains with other SR domains. They bind SR-domains in U2AF and U1 snRNP to enhance their binding to weak consensus sequences. There is no structural information on the mechanism of SR-SR domain interactions. It appears that SR domains are interchangeable suggesting limited specificity of such interactions.
Splicing enhancers and silencers enable alternative splicing regulation

The ability of sequence-specific RNA binding proteins to influence snRNP binding means that tissue-specific proteins in this category can control the outcome of splicing when different splicing events are in competition.

Alternative splicing occurs in several categories with the cassette exon and retained intron being the most common events. Cassette exons often contain stop codons leading to nonsense-mediated decay (NMD), providing a mechanisms by which a splicing switch can lead to RNA turnover. Thus, splicing patterns can control what protein isoform is produced as well as its amount.

![Diagram of alternative splicing events](image-url)
PTPB proteins as a paradigm for splicing regulation

Splicing enhancers and silencers enable tissue-specific alternative splicing. One of many examples involves silencer proteins PTPB1 and PTBP2. These form a circuit that controls alternative in developing central neurons in mammals. PTBPs appear to act through multiple mechanisms. One characterized mechanism by which PTBP1 silences a neuron-specific exon of the Src kinase. In vitro studies show that PTBP1 binds U1 snRNP and prevent exon definition, leading to exon skipping.

There are likely numerous steps at which regulation can occur. For example, silencing of an exon in the lymphocyte transmembrane phosphatase CD45 by hnRNPL occurs by inhibition of the A-->B complex transition (triple snRNP addition). For most alternative splicing regulators, the precise step at which regulation occurs is not well-understood.
THE UNSOLVED MYSTERY OF SPLICEOSOME MUTATIONS IN CANCER

Typical drivers of tumor formation include mutations that activate signaling pathways and cell cycle regulatory pathways. Thus, a major surprise came in 2011 from exome sequencing of hematopoietic tumors (blood cancers) which revealed frequent mutation of components of the spliceosome. Myelodysplastic syndromes, a rare and deadly blood cancer, displayed the highest frequencies, approaching 80% in some clinical subsets. However, more common tumors such as Chronic Lymphocytic Leukemia (CLL) also display frequent driver mutations in components of the spliceosome.
The mutations were heterozygous and often hit specific hotspots within spliceosomal components. The small subunit of U2AF (U2AF35), which recognizes the 3' splice site in the A complex, the SR protein SRSF2, and SF3B1 were the most commonly mutated genes. Mutations in BBP/SF1 were also isolates, suggesting a common connection to U2 snRNP binding branchpoint/3' splice site selection. However, some mutations were identified in ZRSR2, a protein that is a component of the so-called minor spliceosome, which is responsible for the splicing of an unusual subset of introns.

The key question becomes “how do these mutations promote tumor formation”? The simplest possibility would be that they trigger changes in alternative splicing patterns in key oncogenes or tumor suppressor proteins to drive tumor formation.
**SF3B1 mutations trigger cryptic branchpoint utilization**

As described in the previous lecture, the SF3a and SF3b complexes are components of U2 snRNP that persist through the Bact stage of spliceosome assembly and appear to ‘hold’ the branchpoint away from the 5’ splice site. SF3B1 (Hsh155 in *S. cerevisiae*) is the major hotspot of mutations and these cluster in the HEAT repeats that embrace the branchpoint and downstream RNA in the Bact complex. Multiple laboratories have found that SF3B1 mutations result in a shift of 3’ splice site usage towards upstream ‘cryptic’ 3’ splice sites (i.e. sites that are not used in wild-type). This is associated with a shift in the use of branchpoint (again towards cryptic branchpoints not used in wild-type cells). The result is a higher incidence of splicing to 3’ splice sites that produce a frameshift in the coding sequence.
U2AF1 (U2AF35) mutations occur in two hotspot residues in ‘zinc knuckles’

The small subunit of U2AF recognizes the PyAG consensus sequence at the 3' splice site (this occurs in the A complex). The large subunit (U2AF2; U2AF65) recognizes the upstream polypyrimidine tract. The recognition of the PyAG is important for introns with a weak polypyrimidine tract.

U2AF1 recognizes the 3' splice site through two “zinc knuckles.” This motif was first recognized in retroviral capsid proteins which bind viral RNA through a zinc knuckle. The hotspots for mutation lie in the zinc knuckles. The two most common are in serine 34 and glutamine 157. Remarkably, each causes a specific shift in 3' splice site usage. S34F/Y mutations cause a shift towards CAG 3' splice sites and away from UAG 3' splice sites. Q157P/R mutations cause a shift towards 3' splice sites that have a G as the first base of the exon and away from 3' splice sites that have an A as the first base of the exon.

Although there is not a crystal structure of U2AF bound to the 3' splice site, modeling suggests that it is feasible for the two zinc knuckles to recognise the bases upstream and downstream of the conserved AG at the 3' splice site.
NMD hypothesis for tumor promotion by spliceosome mutation

While it is clear that spliceosome mutations in tumors alter splicing, it is far from clear how this promotes tumor formation since the different mutations promote different events. These events appear to each be sufficient as spliceosome mutations are mutually exclusive with each other.

Thus, the field is searching for a unifying mechanism. One involves a process called nonsense-mediated decay. One mechanism of NMD (among three proposed) involves the Exon Junction Complex (EJC). The EJC is deposited upstream of splice junctions by the spliceosome and plays a role in RNA export and several other processes. Among these is the triggering of RNA decay if it is not first displaced by a pioneer round of translation.

Because spliceosome mutations lead to massive amounts of out-of-frame messages, it has been hypothesized that the NMD machinery is “swamped” in tumor cells leading to the production of oncogenic protein fragments. RNA-seq analysis indicates that the NMD machinery itself is induced by spliceosome mutations.
R-loop hypothesis for tumor promotion by spliceosome mutation

Transcription is problematic because RNA-DNA hybrids are more stable than DNA-DNA hybrids. Thus, the nascent transcript has a strong propensity to form an extended structure called an R-loop. In both yeast and in mammalian cells, R-loop formation has shown to result double-stranded breaks. This is thought to be due to the triggering of replication fork stalling.

In both yeast and mammalian cells, assembly of nascent transcripts into spliceosomes has been shown to suppress R-loop formation. Indeed, insertion of an intron into an intron-less gene in *S. cerevisiae* has been shown to reduce R-loop formation increase genetic stability.
Among the many assays to detect R-loops genome-wide is the use of a catalytically-dead RNase H, an enzyme that normally recognizes RNA-DNA hybrids and destroys the RNA strand. Application of this assay to cell lines expressing cancer hotspot spliceosome mutations has revealed an increase in R-loop formation genome-wide. This is associated with a global increase in DNA damage, leading to the proposal that increased mutation is a mechanism by which spliceosome mutations promote tumor formation.