



OPEN ACCESS

EDITED BY

Manuela Montanaro,
Università Link Campus, Italy

REVIEWED BY

Samer M. Hussein,
Laval University, Canada
Mariia Vlasenok,
Skolkovo Institute of Science and
Technology, Russia

*CORRESPONDENCE

Jiushi Guo,
✉ jguo22@student.gsu.edu

RECEIVED 07 October 2025

REVISED 26 March 2026

ACCEPTED 26 March 2026

PUBLISHED 23 April 2026

CITATION

Guo J (2026) The “cutting edge” of
non-canonical RNA splicing.
Front. Mol. Biosci. 13:1719817.
doi: 10.3389/fmolb.2026.1719817

COPYRIGHT

© 2026 Guo. This is an open-access
article distributed under the terms of
the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

The “cutting edge” of non-canonical RNA splicing

Jiushi Guo*

Department of Biology, Georgia State University, Atlanta, GA, United States

Splicing, including alternative splicing, is a fundamental post-transcriptional mechanism in eukaryotes that generates functional proteins and transcript diversity. Canonical splicing follows well-defined rules, such as sufficient intron and exon lengths, specific splice junction orientations, consensus dinucleotides at donor and acceptor sites, and mediation by the spliceosome. However, certain splicing events deviate from these canonical rules. This review synthesizes multiple forms of non-canonical splicing within a framework that reflects their increasing deviation from canonical mechanisms, including non-canonical splice sites, non-canonical splicing in lncRNAs, microexons, recursive splicing, trans-splicing, and spliceosome-independent splicing. In addition, this review provides a critical analysis of the current state of research for each form of non-canonical splicing and outlines key directions for future investigation. As a case study, we reanalyzed RNA-seq data from mouse neuronal cells to further examine non-canonical splice sites. These analyses show that non-canonical introns tend to be shorter and that many non-canonical junctions retain at least one canonical donor or acceptor dinucleotide, supporting the view that a substantial subset remains compatible with spliceosome-mediated recognition. Together, this review provides a structured perspective on how canonical splicing rules can be relaxed, repurposed, or bypassed across distinct biological contexts.

KEYWORDS

alternative splicing, non-canonical splicing, non-canonical splice sites, microexons, recursive splicing, trans-splicing, spliceosome-independent splicing, circular RNA

1 Introduction to RNA splicing

Splicing is a post-transcriptional RNA processing mechanism widely conserved among eukaryotes, in which non-coding introns are excised and coding exons are ligated, generally for translation into functional proteins. Across eukaryotes, splicing occurs in multiple forms, ranging from canonical constitutive and alternative splicing to diverse non-canonical mechanisms (Figure 1, Table 1). Under normal circumstances, the spliceosome recognizes a canonical 5' splice donor, typically a GU dinucleotide at the 5' end of the intron, and a canonical 3' splice acceptor, typically an AG dinucleotide at the 3' end, to define each splicing reaction. Both canonical splice sites and the mediation of the spliceosome are essential components of canonical splicing. Canonical splicing is sometimes completed through a strict adherence to a fixed exon-intron structure, yielding a single, obligatory mature transcript; this process is termed constitutive splicing. Conversely, through alternative combinations of exons within a single gene, known as alternative splicing, this process generates multiple mRNA isoforms that expand proteomic diversity and contribute to gene regulation (Modrek and Lee, 2002; McManus and Graveley, 2011; Kelemen et al., 2013; Marasco and Kornblihtt, 2023) (Figure 2).

In eukaryotes, pre-mRNA splicing is carried out by two related ribonucleoprotein machineries: the major (U2-type) spliceosome, which mediates more than 99% of

splicing events, and the minor (U12-type) spliceosome, which accounts for less than 1% (Lin et al., 2010; Kretova et al., 2023). The U2-type spliceosome is typically composed of the small nuclear ribonucleoproteins (snRNPs) U1, U2, U4, U5, and U6, whereas the U12-type spliceosome generally consists of U11, U12, U4atac, U5, and U6atac (Bradley and Anczukow, 2023; Kretova et al., 2023). The intron classes processed by these two spliceosomes differ in the conservation and architecture of their recognition signals. U2-type introns generally exhibit less constrained splice-site consensus sequences, which contributes to greater flexibility in splice-site selection and facilitates alternative splicing; they also typically contain a more prominent 3' polypyrimidine tract (PPT, also known as Py-tract). In contrast, U12-type introns are characterized by more highly conserved recognition elements. Their 3'-end recognition depends more strongly on a highly conserved branch point sequence, and they often lack the canonical 3' PPT typical of U2-type introns (Wu and Fu, 2015; Akinyi and Frilander, 2021). With respect to terminal splice-site dinucleotides, canonical U2-type splicing predominantly follows the GT-AG rule, whereas U12-type splicing is most associated with the AT-AC motif, although a small subset of U12-type introns also uses GT-AG motif (Turunen et al., 2013; El Marabti et al., 2021).

A canonical splicing event typically follows these principles: introns and exons of sufficient lengths, major canonical donor and acceptor sites (GT-AG) contributed sequentially by the preceding exon and the following exon, and mediation by the spliceosome. However, exceptions to these canonical rules have been increasingly documented. Collectively termed non-canonical splicing, these events do not strictly follow this paradigm. Some remain partially recognized and processed by the spliceosome, whereas others proceed through distinct mechanisms and yield diverse molecular outcomes.

2 Splicing with non-canonical splice sites

While canonical splicing relies on well-defined splice site signals and a conserved exon-intron architecture, accumulating evidence indicates that splice site recognition is not an all-or-nothing process. Instead, the spliceosome exhibits a measurable degree of tolerance to deviations from the major canonical GT-AG consensus, giving rise to splicing events that operate at the boundary of canonical rules. These deviations are most prominently manifested at the

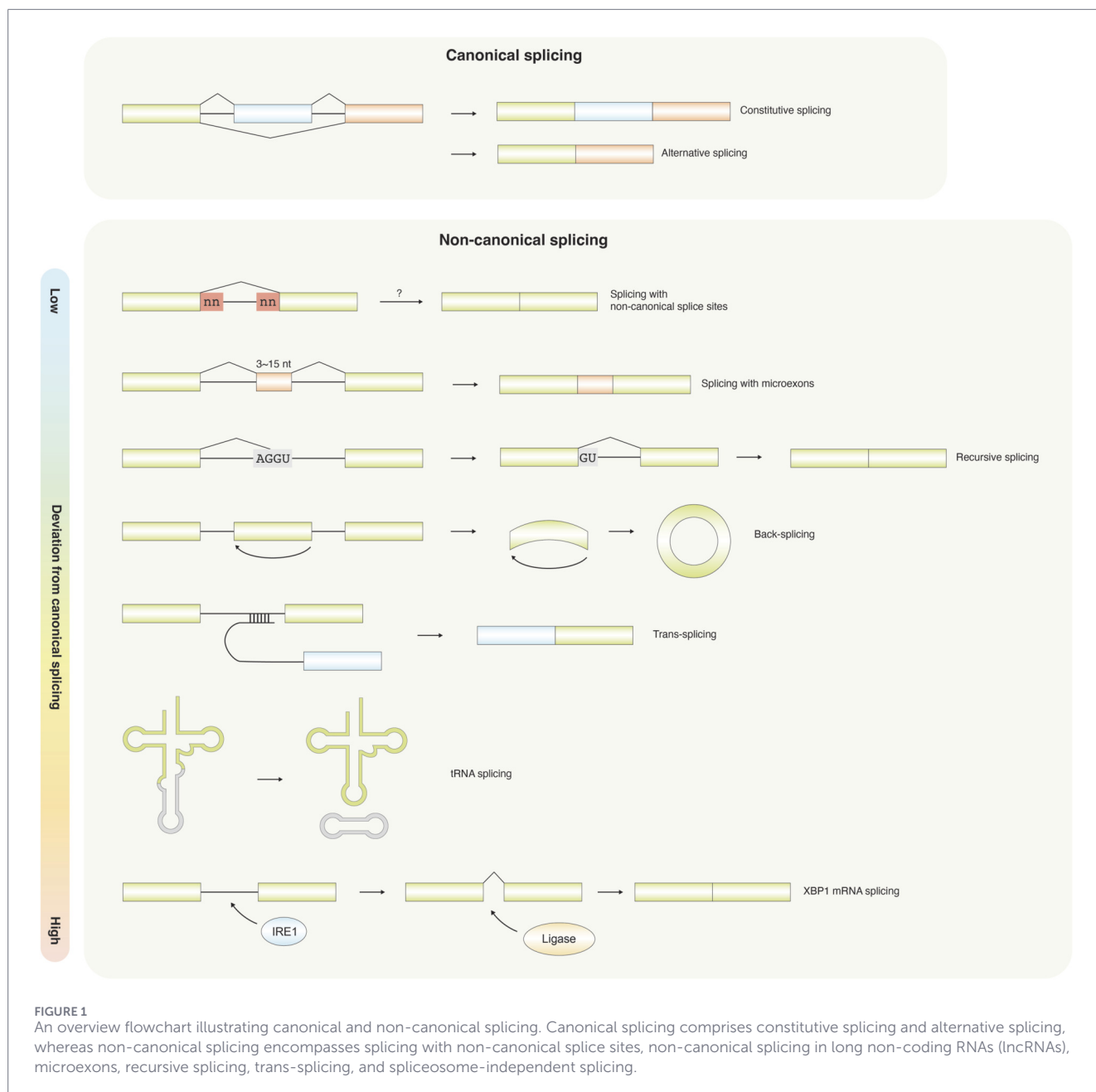
level of splice site sequences themselves and represent one of the most fundamental forms of non-canonical splicing. Accordingly, an examination of non-canonical splice sites provides a critical entry point for understanding how canonical splicing principles are relaxed, modified, or repurposed to achieve regulatory flexibility.

Splicing events involving non-canonical splice sites do not conform to the canonical GT-AG motif. Depending on the context, such events may involve a non-canonical donor site, a non-canonical acceptor site, or deviations at both ends (Figure 3). Although splice-site recognition by the spliceosome is dominated by canonical GT-AG signals, multiple lines of evidence indicate that spliceosomal processing is not strictly limited to GT-AG junctions. A subset of non-canonical splice sites can still be recognized and excised when their surrounding sequence context retains features compatible with U2- or U12-type splice-site recognition (Turunen et al., 2013; Parada et al., 2014; Akinyi and Frilander, 2021). In such cases, the terminal dinucleotides deviate from the canonical pattern, but the broader splice-site architecture remains sufficiently similar to known major or minor spliceosomal recognition signals. In addition, some non-canonical splice sites may become functionally more canonical-like through context-dependent mechanisms such as A-to-I RNA editing or may show evolutionary relationships to nearby or ancestral canonical splice sites (Parada et al., 2014). Non-canonical junctions that retain recognizable U2/U12-like consensus features are often described as U2/U12-like non-canonical junctions. By contrast, non-U2/U12-like non-canonical junctions are not readily explained by established major or minor spliceosomal recognition models and may represent a heterogeneous class that includes both mechanistically distinct RNA-processing events and lower-confidence events influenced by technical artifacts (Parada et al., 2014; Sibley et al., 2016).

Non-canonical splice sites are exceedingly rare among all splicing events. In a previous systematic baseline study, splice sites outside the canonical GT-AG class accounted for only about 1% of all events (Parada et al., 2014). Moreover, when the major U2/U12-like non-canonical classes, namely GC-AG and AT-AC, are excluded, the proportion of all remaining minor non-canonical splice-site combinations decreases further to only 0.08% (Parada et al., 2014). Similarly, another study focused on non-canonical 5' splice sites found that non-canonical 5' splice junctions represented only 0.4%–1% of all exon junctions (Erkelenz et al., 2018). Taken together, these observations indicate that splicing involving non-canonical splice sites constitutes a minority class rather than a mainstream mode of splicing.

For such rare splicing events, methodological differences can substantially affect prevalence estimates. For example, in the earlier systematic baseline study, after functional relevance filtering, the number of retained non-canonical splice sites was reduced to 28% of the broader set identified by library-guided detection (462 of 1,630), and after further classification into U2/U12-like sites, this number dropped even more sharply, to only 11% of the initially detected candidate set (184 of 1,630) (Parada et al., 2014). Likewise, in the study of non-canonical 5' splice sites, the raw RNA-seq estimate placed non-canonical 5' splice junctions at 2.2%, whereas removal of low-usage junctions likely to represent noise reduced the high-confidence proportion to only 0.4%–1% (Erkelenz et al., 2018). These findings show that differences in filtering stringency and functional criteria can materially alter the reported prevalence

Abbreviations: snRNPs, small nuclear ribonucleoproteins; PPT, polypyrimidine tract; PCGs, Protein Coding Genes; circRNA, Circular RNA; RBP, RNA-binding Protein; m6A, N6-methyladenosine; eIF3, Eukaryotic Translation Initiation Factor 3; PSI, Percent Spliced In; IDR, Intrinsically Disordered Region; ASD, Autism Spectrum Disorder; znRS, Zero-nucleotide Recursive Splice site; SL1 / SL2, Spliced Leader 1 / 2; cDNA, Complementary DNA; CPDase, Cyclic Phosphodiesterase; BHB, Bulge-Helix-Bulge; lncRNA, long non-coding RNA; SREs, splicing regulatory elements; EGFP, Enhanced Green Fluorescent Protein; AS, alternative splicing; IRES, internal ribosome entry site; SINEs, short interspersed nuclear elements; ciRNAs, circular intronic RNAs; IRE1, Inositol-requiring enzyme 1; XBP1, X-box binding protein 1; HAC1, Homologous to ATF/CREB 1; ER, Endoplasmic Reticulum; ERAD, Endoplasmic Reticulum-Associated Degradation; NMD, Nonsense-Mediated mRNA Decay; RNA-seq, RNA sequencing.



of non-canonical splice sites across studies; however, even when considered together, published estimates consistently place them within a very small fraction of total splicing events. Moreover, because splicing itself is inherently heterogeneous and dynamic, caution is warranted when applying functionality-based filters to non-canonical splice sites, as overly stringent or overly permissive thresholds may distort their apparent prevalence.

With respect to sequencing modality, the overall distribution of non-canonical splice sites inferred from long-read sequencing has not shown major divergence from estimates derived from short-read sequencing. Nevertheless, the large number of novel splice junctions uncovered by long-read approaches remains an important subject for further investigation. A genome-wide survey of non-canonical splice sites in plants reported that, under long-read sequencing,

all non-canonical splice sites together accounted for approximately 1.35%, whereas minor non-canonical splice sites represented only 0.09% (Pucker and Brockington, 2018). In addition, a long-read transcriptomic characterization of mouse neural tissue reported that only 0.1% of all known splice junctions were non-canonical (Tardaguila et al., 2018), and a study of rainbow trout genome annotation likewise reported approximately 0.1% non-canonical junctions (Ali et al., 2021). Collectively, these long-read-based studies are broadly consistent with the view that non-canonical splicing represents a very low-abundance class of splicing events.

At the same time, long-read sequencing has revealed a large number of previously undetected and uncharacterized splice sites, typically classified as novel splice junctions, which appear to have remained relatively underexplored. Many studies report these novel

TABLE 1 Comparison of the canonical and non-canonical splicing types discussed in this review, together with their principal characteristics.

Class	Characteristics
Canonical constitutive splicing	All exons within a single pre-mRNA are joined sequentially to produce a single transcript isoform
Canonical alternative splicing	Exons within a single pre-mRNA are joined in a selective manner, with one or more exons potentially skipped, thereby generating multiple transcript isoforms
Non-canonical splice sites	One or both nucleotides at the splice junction deviate from the canonical GT-AG consensus, resulting in splice sites that do not conform to the standard motif
Non-canonical lncRNA splicing	Commonly involves splicing at non-canonical splice sites, most notably GC-AG splice junctions
Microexon splicing	Splicing involving extremely short exons, typically only 3–15 nucleotides in length
Recursive splicing	Splicing of exceptionally long introns that contain multiple internal splice sites in addition to their terminal boundaries, allowing intron removal through a stepwise series of successive splicing reactions
Back splicing	A splicing event in which a downstream splice acceptor is joined to an upstream splice donor, often giving rise to circular RNAs (circRNAs)
Trans-splicing	A splicing process in which a transcript is generated by joining sequences derived from multiple pre-mRNAs of the same gene, different genes, or even different chromosomes
tRNA splicing	An essential step in tRNA maturation, mediated by a set of enzymes distinct from the spliceosome
IRE1-mediated splicing	A spliceosome-independent form of cytoplasmic splicing mediated by IRE1, occurring only in specific target transcripts such as XBP1 or HAC1

splice junctions as a separate category in long-read RNA-seq analyses; however, because the primary focus of those studies often lies elsewhere, these junctions are not always subjected to further filtering, remapping, or mechanistic analysis (Wright et al., 2022; Humphrey et al., 2025). Determining to what extent such novel junctions correspond to *bona fide* non-canonical splicing events therefore represents an important emerging challenge. In this regard, the earlier long-read transcriptomic study of mouse neural tissue is particularly informative, as it further analyzed novel splice junctions and found that 31% were classified as non-canonical

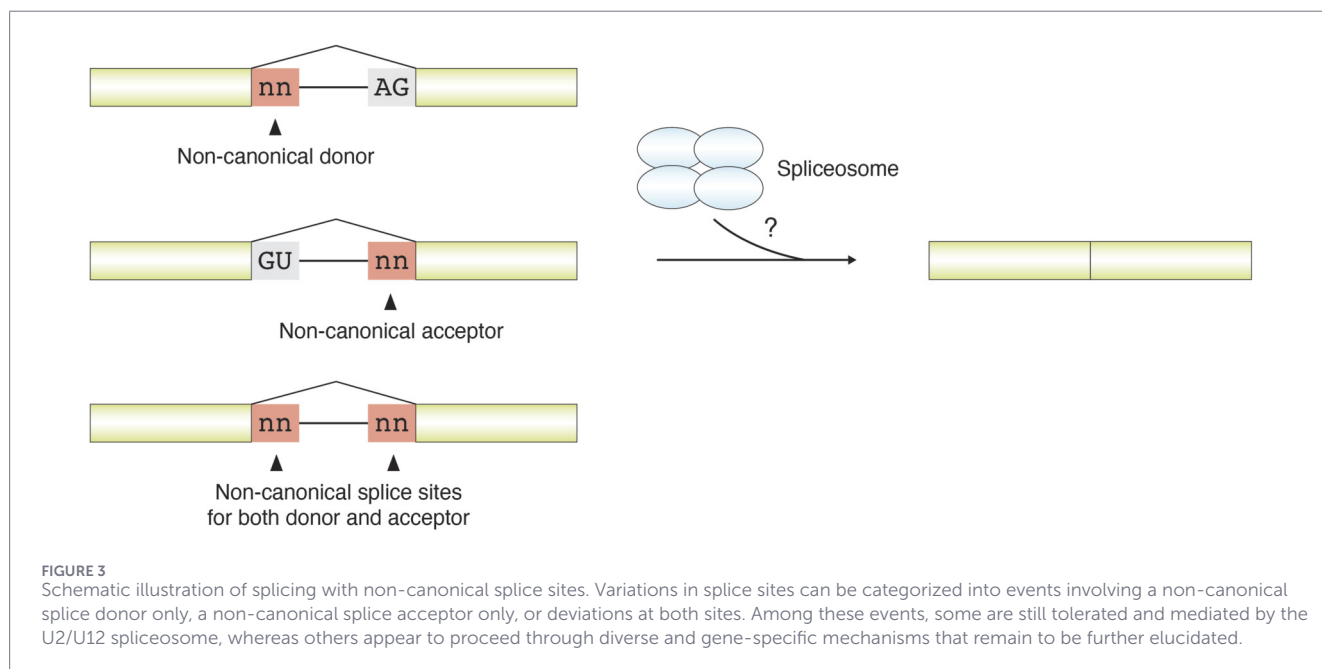
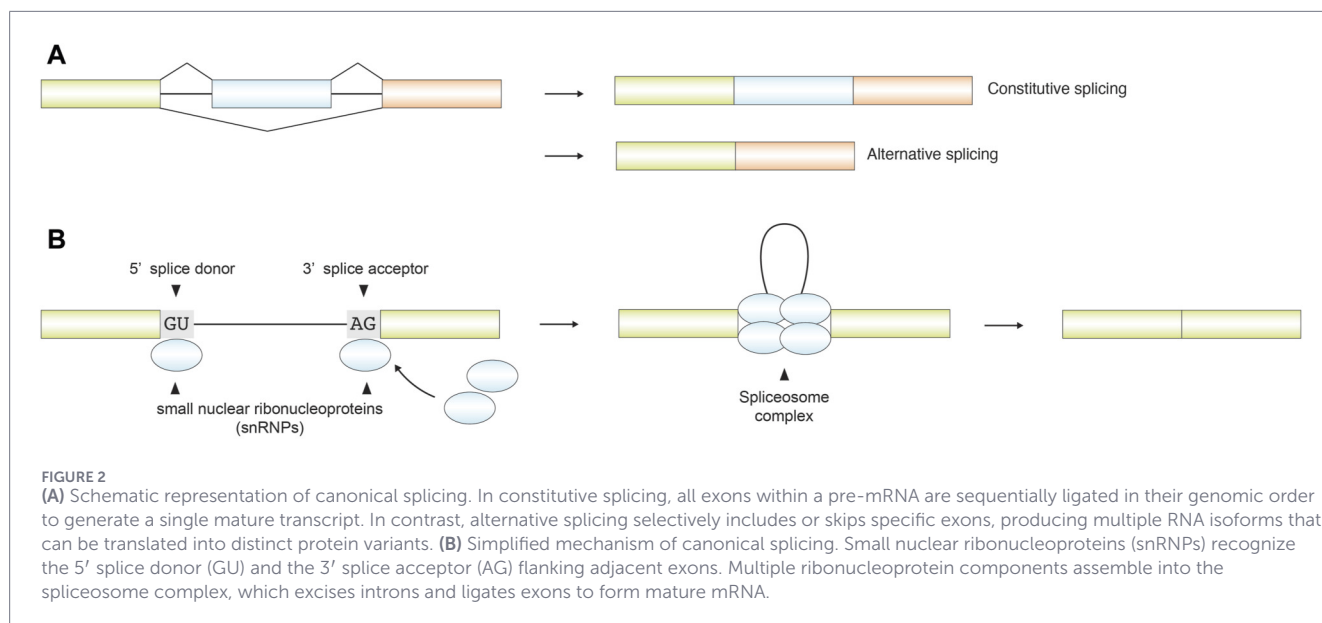
junctions (Tardaguila et al., 2018). This suggests that long-read sequencing can uncover a substantial number of novel isoforms and potentially informative splice features that may be missed, or at least insufficiently resolved, by short-read approaches. However, an important issue has also emerged: novel-calling in long-read data appears to be highly sensitive to pipeline design and artifact control (Tardaguila et al., 2018; Su et al., 2024), indicating that further improvements in quality-control procedures are still needed. Moreover, there is currently a lack of dedicated long-read-based baseline surveys of non-canonical splice sites, making it difficult to determine whether the large number of novel isoforms identified by long-read sequencing substantially alters the baseline frequencies previously inferred from short-read data.

Non-canonical splice sites can also display tissue-specific expression patterns. In terms of gene regulation, they are often surrounded by an increased density of splicing regulatory elements (SREs), indicating that more complex regulatory architectures may be required for their recognition (Parada et al., 2014; Vuong et al., 2016; Yang et al., 2023). Moreover, certain non-canonical splice sites themselves contribute to gene regulation; for example, some can trigger nonsense-mediated mRNA decay (NMD). In the mouse brain, specific 3′ splice sites within the *Syngap1* gene, which encodes the Ras GTPase-activating protein SYNGAP1, have been shown to induce NMD, leading to mRNA degradation and haploinsufficiency of SYNGAP1 protein and thereby affecting neuronal function and behavior (Yang et al., 2023). Coincidentally, additional non-canonical splice sites and truncated exon variants have also been reported in *Syngap1* (Kilinc et al., 2022). Building on these observations, we re-analyzed RNA-seq data from mouse neuronal cells described in the above studies (Kilinc et al., 2022; Yang et al., 2023) to further characterize the prevalence and sequence features of non-canonical splice sites.

2.1 Sequence enrichment of non-canonical splice sites

A sequence-enrichment analysis was performed on RNA-seq reads from EGFP+ cells sorted from E14.5 Tubb3-EGFP transgenic mice (Yang et al., 2023). Among all detected splice junctions, canonical GT-AG junctions were the most abundant, accounting for approximately 98.9% of the total (Table 2). These were followed by the major U2/U12 spliceosomal non-canonical preference sites, GC-AG and AT-AC, which accounted for approximately 0.83% and 0.12%, respectively, with the remaining fraction comprising other minor non-canonical splice sites (Table 2). The most prevalent splice-site classes and their relative proportions were highly similar to those reported in previous baseline studies (Parada et al., 2014).

Examination of the 5′- and 3′-terminal dinucleotides of all non-canonical introns revealed that AG remained by far the most frequently used non-canonical acceptor. Although GT was less frequent among non-canonical donor sites than GC and AT, which were observed mainly in the U2/U12-like GC-AG and AT-AC junctions, GT still occurred more frequently than most other minor donor classes (Figures 4A,B). In other words, GT donor sites and AG acceptor sites remained representative among non-canonical splicing events, which means a proportion of the non-canonical



junctions were only non-canonical at one end and retained a canonical dinucleotide at the opposite splice site. These findings further indicate that many mapped non-canonical splice sites still retain at least one canonical donor or acceptor dinucleotide, strongly suggesting that the spliceosome can accommodate a certain degree of non-canonical variation at one or both ends. Although the major U2/U12-like non-canonical classes, GC-AG and AT-AC, accounted for a large fraction of this category, it should also be recognized that this group still includes a subset of single-sided non-canonical splice sites (Table 2), whose underlying splicing mechanisms are also expected to be compatible with spliceosomal recognition. Among splice sites lacking the GT-AG motif, GC and AT emerged as two frequent alternative donor dinucleotides, whereas AC was the most common alternative acceptor (Figures 4C,D). This enrichment is

also likely attributable to the frequent occurrence of GC-AG and AT-AC junctions. In addition, TG and GG were also observed at relatively high frequencies as alternative acceptors (Figures 4C,D). Notably, the GG acceptor displayed potential functional links to the non-canonical splicing events reported in SYNGAP1 and to the activation of nonsense-mediated mRNA decay (NMD) (Kilinc et al., 2022; Yang et al., 2023), highlighting a promising direction for future investigation.

2.2 Difference in intron length between canonical and non-canonical introns

Within the same sample, introns excised at non-canonical splice sites were significantly shorter than canonical introns, with

TABLE 2 Distribution of splice-site dinucleotide combinations among all detected splice junctions. Each splice-site class is defined by the donor and acceptor terminal dinucleotides. The table reports the absolute count and relative frequency (%) of each splice-site combination across all detected junctions.

Splice site	Count	%	Splice site	Count	%	Splice site	Count	%
GTAG	149,786	98.9438	GCCC	2	0.0013	CCGT	1	0.0007
GCAG	1,253	0.8277	GCGG	2	0.0013	GCGA	1	0.0007
ATAC	185	0.1222	ACCG	1	0.0007	TGTT	1	0.0007
GTTG	33	0.0218	AGTG	1	0.0007	GGAA	1	0.0007
GTGG	21	0.0139	CTAG	1	0.0007	GGTC	1	0.0007
ATAG	11	0.0073	GTTC	1	0.0007	GGCC	1	0.0007
GAAG	9	0.0059	TGGA	1	0.0007	CTCC	1	0.0007
TTAG	8	0.0053	TCAG	1	0.0007	TGTC	1	0.0007
ATAA	8	0.0053	ACAC	1	0.0007	TTTT	1	0.0007
ATAT	6	0.0040	GCCT	1	0.0007	AGGG	1	0.0007
GGAG	6	0.0040	AGCC	1	0.0007	GGGT	1	0.0007
GTAA	4	0.0026	TTCT	1	0.0007	CTAT	1	0.0007
GGAC	3	0.0020	GGGC	1	0.0007	CTAA	1	0.0007
GTCA	3	0.0020	CCTT	1	0.0007	CAAC	1	0.0007
TGTG	2	0.0013	TCCC	1	0.0007	TCTT	1	0.0007
TATC	2	0.0013	CCCC	1	0.0007	AAGG	1	0.0007
GGGA	2	0.0013	CCGG	1	0.0007	TCGG	1	0.0007
AGAG	2	0.0013	CTGA	1	0.0007	GACA	1	0.0007
TAGA	2	0.0013	GATG	1	0.0007			

a median length difference of 415.5 bp (Figure 5; Mann–Whitney U test, $p = 3.657 \times 10^{-18}$). After further stratification of all non-canonical splice introns into U2/U12-like non-canonical and non-U2/U12-like non-canonical subsets, both groups likewise showed significantly shorter introns than canonical introns. Specifically, the median intron length of the U2/U12-like non-canonical subset was 343 bp shorter than that of canonical introns (Figure 5; $p = 2.904 \times 10^{-12}$), whereas the median intron length of the non-U2/U12-like non-canonical subset was 659.5 bp shorter (Figure 5; $p = 1.494 \times 10^{-9}$). These findings strongly suggest that, regardless of whether a non-canonical junction is still likely to be processed by the U2/U12 spliceosome, its associated intron tends to be shorter than those of canonical splice sites. Interestingly, introns in the non-U2/U12-like non-canonical subset were also significantly shorter than those in the U2/U12-like non-canonical subset (Figure 5; $p = 2.544 \times 10^{-3}$). These results indicate that non-U2/U12-like non-canonical introns exhibit a more pronounced shortening relative to canonical introns than do U2/U12-like non-canonical introns, which may reflect underlying differences in their splice-site recognition mechanisms, including possible differences in sequence conservation. Accordingly, a tendency toward shorter introns may serve as a useful indicator in future studies of non-canonical splicing for further identifying and prioritizing non-U2/U12-like splicing events.

2.3 Future perspectives of splicing with non-canonical splice sites

Our analysis indicates that a substantial proportion of non-canonical splicing events recognized in mouse neuronal cells involve splicing with only a single non-canonical splice site, remarkably for the acceptor. This suggests that many non-canonical splicing events remain mediated by the U2/U12 spliceosome. At the same time, beyond the major U2/U12-preferred non-canonical junction classes, such as GC–AG and AT–AC, some junctions with a splice-site change at only one end are also likely to be tolerated and successfully processed by the U2/U12 spliceosome. In addition, introns involved in splicing with non-canonical splice sites, especially non-U2/U12-like non-canonical splice sites, are significantly shorter than those associated with canonical splicing events. Given the persistent challenges in accurately identifying diverse forms of non-canonical splicing, a bias toward shorter intron length may serve as an auxiliary criterion for the detection and characterization of non-canonical splicing events. Moreover, as non-canonical splice sites in SYNGAP1 mRNA can potentially trigger NMD, leading to reduced dosage or altered forms of proteins essential for neurodevelopment, the role of aberrant splicing in the pathogenesis of various diseases warrants further investigation.

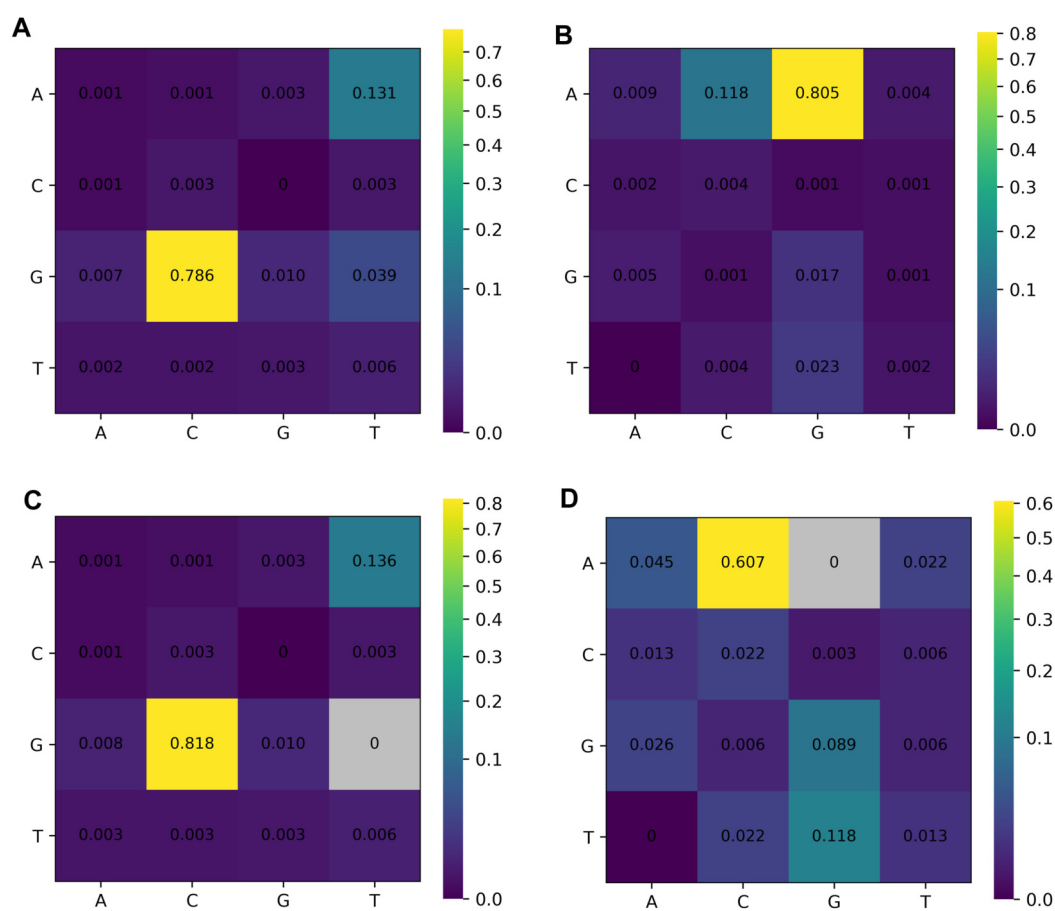


FIGURE 4

4 × 4 dinucleotide heat-map analysis of non-canonical splice sites: (A) donors, (B) acceptors, (C) donors with GT excluded, and (D) acceptors with AG excluded. All non-canonical splice sites were first counted and then converted to frequencies. Rows correspond to the first nucleotide of the splice site and columns to the second. Heat-map colors were scaled using power-law normalization (PowerNorm, gamma = 0.5) to enhance the visual separation of low-frequency minor non-canonical dinucleotide classes. This normalization was applied only to the color scale while the underlying counts/frequencies and the numeric labels remained untransformed.

The continued development of long-read sequencing has revealed many previously unclassified novel splicing isoforms, which are likely to contain substantial numbers of as-yet undiscovered, unclassified, and uncharacterized non-canonical splice sites. Accordingly, systematic investigation of non-canonical splice sites in the context of long-read sequencing, together with the establishment of comprehensive baseline surveys comparable to those developed in the short-read era, represents an important and timely direction for future research.

While the above features outline the general landscape of non-canonical splice sites across transcripts, their distribution and functional relevance are not uniform across RNA classes. Notably, long non-coding RNAs (lncRNAs) exhibit a distinct enrichment of non-canonical splice sites (Khan et al., 2023), providing a permissive context in which spliceosomal tolerance may be amplified and repurposed for regulatory functions.

3 Non-canonical splicing in long non-coding RNAs (lncRNAs)

Compared with protein-coding genes, lncRNAs display a markedly higher frequency of non-canonical splice sites, reflecting reduced coding constraints and increased splicing plasticity (Figure 6). lncRNAs are RNA transcripts longer than 200 nt but lack the potential to be translated into proteins (Mattick et al., 2023). lncRNAs also undergo splicing, and the vast majority are mediated by the spliceosome. However, compared to mRNA, lncRNA exhibits distinct differences in selective pressure, functional constraints, and splicing precision (Derrien et al., 2012; Chen and Kim, 2024). This renders lncRNA a domain where non-canonical splicing occurs more readily yet proves more challenging to define.

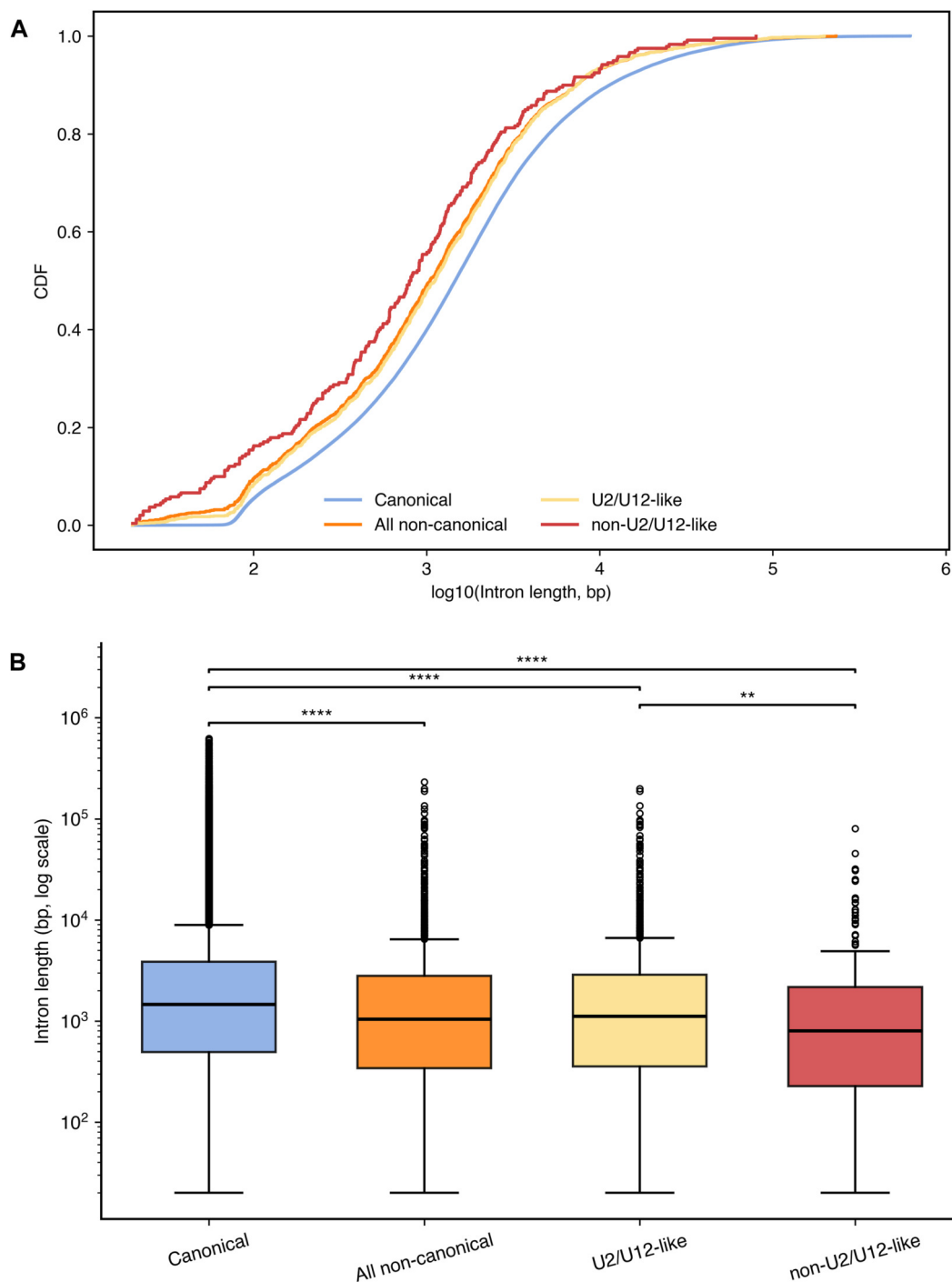


FIGURE 5

(A) Empirical cumulative distribution function (ECDF) plot and (B) box plot showing differences in intron lengths (log10-scaled) between canonical and non-canonical introns (Mann–Whitney U test, $p = 3.657 \times 10^{-18}$), canonical and non-canonical U2/U12-like introns ($p = 2.904 \times 10^{-12}$), canonical and non-canonical non-U2/U12-like introns ($p = 1.494 \times 10^{-9}$), and non-canonical U2/U12-like and non-U2/U12 like ($p = 2.544 \times 10^{-3}$). Individual points shown in the boxplots represent outlier observations only, defined as values lying beyond $1.5 \times$ the interquartile range from the box, rather than all underlying data points.

3.1 Instability of lncRNA splicing mechanisms

Compared with mRNA splicing, lncRNA splicing is generally characterized by weaker splice site strength, lower conservation

of exon-intron architecture, and a markedly higher prevalence of alternative splicing, with intron retention and exon skipping representing common features rather than exceptions in lncRNAs (Ouyang et al., 2022; Khan et al., 2023). Notably, alternative splicing of lncRNAs is particularly prevalent and can regulate



FIGURE 6

A representative and relatively conserved form of non-canonical splicing in lncRNAs. This scenario exemplifies splicing with a non-canonical donor only, in which a non-canonical dinucleotide (GC) replaces the canonical GU at the 5' splice donor of the first exon in an lncRNA transcript.

selective splicing through multiple mechanisms, including interactions with splicing factors, base pairing with pre-mRNAs, and modulation of chromatin remodeling (Pisignano and Ladomery, 2021; Ouyang et al., 2022). Such variability and instability in splicing provide a broad landscape for the emergence and diversification of non-canonical splicing events.

3.2 Non-canonical splice sites in lncRNAs

The presence of weak splice sites and the high splicing plasticity characteristic of lncRNAs provide a structural basis for non-canonical splicing, particularly for the emergence and utilization of non-canonical splice sites. The occurrence of non-canonical splice sites is markedly higher in lncRNAs than in protein-coding genes (PCGs) (Khan et al., 2023). The GC-AG pattern represents a non-canonical splicing mode that is markedly enriched in lncRNAs. In both human and mouse lncRNAs, GC-AG motifs occur at frequencies approximately two-to four-fold higher than those observed in protein-coding genes and are preferentially enriched within the first intron (Abou Alezz et al., 2020). In splicing events involving non-canonical sites, both donor and acceptor splice sites are significantly weaker, and the associated polypyrimidine tracts are shorter (Abou Alezz et al., 2020; Statello et al., 2021). Such structural features render these sites more susceptible to further alternative splicing and show a strong association with alternative polyadenylation, collectively indicating that non-canonical splice sites play a regulatory role in lncRNA splicing (Abou Alezz et al., 2020).

Non-canonical splicing in lncRNAs remains a relatively infrequent class of RNA-splicing events overall, yet its reported prevalence is consistently higher than that observed in protein-coding genes (PCGs), with no major disagreement across published studies. A previous review summarized that atypical splice sites account for approximately ~3% of splice sites in lncRNAs, compared with ~0.8% in PCGs (Khan et al., 2023). Another study similarly reported that non-GT-AG introns comprise approximately 3.4% of lncRNA splice junctions in humans, compared with 1.4% in PCGs; in mice, the corresponding values were 3.4% and 1.2%, respectively (Abou Alezz et al., 2020). Together, these studies support a consistent overall conclusion: although non-canonical splicing in lncRNAs remains a minority phenomenon, its usage is higher than in protein-coding genes.

3.3 Future perspectives of non-canonical splicing in lncRNAs

lncRNAs undergo pervasive alternative splicing, frequently generating aberrant isoforms implicated in oncogenesis (Ouyang et al., 2022; Wang et al., 2022). The extensive splicing

variability of lncRNAs, together with their documented roles in tumorigenesis, demonstrates that lncRNAs are far from mere transcriptional noise but instead represent a long-overlooked class of regulatory molecules with substantial functional significance in gene regulation. Beyond the already highly diverse landscape of canonical alternative splicing, lncRNAs that undergo non-canonical splicing also warrant close attention, as they may uncover conserved splicing mechanisms or serve as indicators of aberrant splicing events that contribute to disease pathogenesis.

4 Splicing with microexons

While non-canonical splice sites challenge the sequence constraints of spliceosome recognition, other non-canonical splicing events arise even when canonical splice signals are preserved. Microexons exemplify such cases, in which extreme exon length rather than splice site identity imposes unique regulatory and kinetic constraints on splicing.

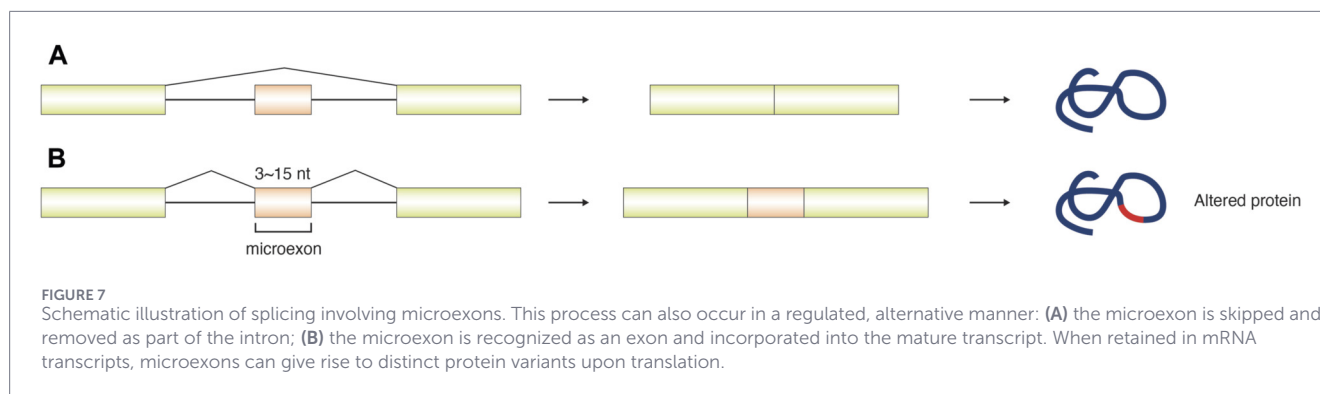
Microexons are very short exonic sequences, typically only 3–15 nucleotides in length, with tightly regulated splicing and high evolutionary conservation (Li et al., 2015; Mackensen and Irimia, 2025) (Figure 7). Once largely overlooked, they are now recognized as modulators of diverse protein functions and interactions, particularly in the central nervous system (Irimia et al., 2014; Li et al., 2015; Mackensen and Irimia, 2025). Microexons show a marked tendency for in-frame insertion, which can fine-tune protein properties (Irimia et al., 2014; Li et al., 2015). Mechanistically, microexons are similar to canonical exons, but their unusually short length has long impeded their reliable detection by standard RNA-seq alignment and transcriptome annotation methods.

4.1 Microexons are often tightly regulated

The inclusion of microexons in mature mRNA is stringently regulated across tissues and developmental stages. RNA-seq analyses have shown that, in both humans and mice, the vast majority of microexons are detected in transcripts associated with the central nervous system, whereas they are seldom found in non-neural tissues (Irimia et al., 2014; Li et al., 2015). These observations indicate that microexon splicing is highly tissue-specific and closely linked to neural development and function.

4.2 Many microexons exhibit strong evolutionary conservation

Microexons display highly similar sequences, splicing structures, and regulatory patterns across species. Cross-species analyses have revealed nearly identical exon sequences and flanking



splice sites in numerous neural microexons, and the patterns of percent spliced in (PSI) likewise show similar inclusion trends between mammals (Irimia et al., 2014; Ustianenko et al., 2017). Most microexons are integer multiples of three nucleotides—typically 3 to 15 bases, and occasionally up to 27 or even 51 bases—allowing their inclusion without disrupting the coding frame (Li et al., 2015). Remarkably, even the shortest microexons, containing only 3–6 nucleotides, display highly conserved patterns that are rarely observed in typical exons (Li et al., 2015).

4.3 Alterations and functions of microexons on mRNAs and proteins

Through in-frame splicing, microexons can modify the amino acid sequence without compromising the overall structural integrity of the encoded protein. The microexon-encoded residues are typically inserted into intrinsically disordered regions (IDRs)—flexible linkers between structured domains—or into loops on domain surfaces, thereby fine-tuning protein interactions and dynamics. These additional amino acids are frequently charged residues, such as glutamate (E), aspartate (D), lysine (K), or arginine (R) (Dergai et al., 2010; Irimia et al., 2014; Li et al., 2015).

4.4 Relevance of microexons to neural tissues and neurological disorders

Microexons—short exons of typically 3–15 nucleotides—exhibit a strong neural tissue-specific expression pattern and participate in diverse neuronal processes, including neuronal maturation, synaptic plasticity, cytoskeletal remodeling, and signal integration (Irimia et al., 2014; Quesnel-Vallières et al., 2016). Dysregulation of neuronal microexon splicing has been reported in the brains of individuals with autism spectrum disorder (ASD) (Irimia et al., 2014; Quesnel-Vallières et al., 2016), suggesting that modulating microexon inclusion could provide novel avenues for the treatment of neurodevelopmental or neurodegenerative disorders.

4.5 Discussion about prevalence of microexon splicing

Microexon splicing has likewise been widely reported as a relatively rare class of splicing event; however, despite substantial

differences in study design and statistical definitions, microexons are also consistently described as having a high degree of involvement in neural-regulated alternative splicing (AS). Microexon-type AS events account for only about 1% of all AS events (Irimia et al., 2014; Li et al., 2015). In a study of autistic brains, 60.7% of alternative microexons showed increased neural PSI, whereas only 9.5% of longer alternative exons did so, indicating a marked enrichment of microexons in neural function-related splicing programs (Irimia et al., 2014). Interestingly, in another study of micro-exons in the human brain that did not explicitly report a comparable neural-PSI-based probability, the proportion of brain-expressed microexons among all detected microexons was also 60.7% (7,949 of 13,095) (Li et al., 2015). However, this numerical agreement is likely coincidental rather than biologically equivalent, because the two studies addressed different questions: the former quantified neural inclusion at the level of alternative splicing, whereas the latter measured only expression in brain-associated transcripts. This distinction implies that the actual proportion of microexons showing preferential inclusion in the brain in the latter study was likely lower than 60.7%.

Such differences may reflect variation in tissue scope and dataset scale (mouse and human vs. human only), RNA-seq processing strategies (a multi-module exon–exon junction library approach vs. Augmented Transcriptome Mapping), and microexon filtering criteria (3–15 nt vs. ≤51 nt, with candidates shorter than 6 nt discarded). Taken together, these observations suggest that microexons have not yet been studied within a unified baseline framework; instead, they have often been filtered and analyzed according to disease-specific or regulatory priorities. Combined with the inherent technical difficulty of detecting microexons, this has left the field without a broadly accepted and widely used standard analytical framework, and it therefore remains comparatively underexplored. Given that the continued development of long-read sequencing is enabling the detection of increasing numbers of microexons (Rahimi et al., 2021a; 2021b), long-read-based studies of microexons, particularly in the context of neural regulation, represent a promising direction for future research.

4.6 Future perspectives of microexon splicing

Due to their extremely short length and consequent difficulty in detection, microexons are systematically underestimated by

current short-read RNA-seq approaches, and their annotation remains broadly incomplete. Furthermore, because many studies of RNA splicing remove very short introns and exons during filtering by treating them as noise, many potential microexons may be discarded at an early stage and consequently never enter downstream analyses. Advances in long-read sequencing, specialized alignment algorithms, and single-cell full-length RNA-seq are expected to substantially enhance the identification and characterization of microexons (Fuentes-Beals et al., 2023; Liu Z. et al., 2023; Fu et al., 2025). Meanwhile, the regulatory mechanisms governing microexon splicing remain incompletely understood. For example, how the spliceosome reliably recognizes highly conserved microexons, and why microexons are selectively included under specific cellular contexts, are still open questions. Given that microexon inclusion can generate distinct protein variants and that microexons are evolutionarily conserved in the nervous system, elucidating the mechanisms underlying microexon splicing may provide novel insights into diverse neuronal processes as well as neurodegenerative diseases.

5 Recursive splicing

Beyond exon definition, the spliceosome also faces challenges imposed by intron length. In genes containing exceptionally long introns, recursive splicing provides a stepwise strategy to maintain splicing fidelity and efficiency. Recursive splicing (RS) is a stepwise RNA splicing mechanism that segments the removal of very long introns—typically tens to hundreds of kilobases—into multiple successive steps (Figure 8). This strategy enhances splicing efficiency and reduces the risk of mis-splicing or the formation of aberrant isoforms. It proceeds through embedded recursive splice sites (RS sites) within the intron, which act as internal splice junctions and allow the intron to be excised fragment by fragment.

5.1 Structure and function of recursive splice sites

Recursive splicing typically relies on multiple AG|GU motifs embedded within long introns, which serve as relay points during the stepwise removal of a single long intron. These consecutive splice signals, in which a 3' AG acceptor is immediately followed by a 5' GU donor, are known as zero-nucleotide recursive splice sites (znRS sites), reflecting the absence of intervening nucleotides. In the first splicing step, a znRS site functions as a relay 3' acceptor, pairing with the original 5' donor to excise the initial intronic segment. The same znRS element then acts as a 5' donor in the next step, joining with the downstream AG acceptor, and this cycle can repeat when multiple znRS sites are present. Ultimately, znRS elements are either removed together with the intron (RS-excluded) or retained as optional exonic segments (RS-included), potentially generating alternative protein isoforms (Burnette et al., 2005; Duff et al., 2015; Sibley et al., 2015; Georgomanolis et al., 2016; Hayashi et al., 2018). The stepwise reuse of recursive splice sites inherently links recursive splicing to splicing dynamics and kinetics, as each splicing event creates a transient intermediate that temporally defines subsequent splice site recognition (Liu M. et al., 2023; Merens et al., 2024).

5.2 The distribution of recursive splicing

Recursive splicing was first identified in specific *Drosophila* genes and corresponding mRNA isoforms (Hatton et al., 1998), and has since been documented in mice and humans, with more than 100 RS sites reported (Duff et al., 2015; Sibley et al., 2015). These sites are typically located in the middle of long introns exceeding 10 kb in length. Moreover, recursive splicing occurs frequently in genes involved in nervous-system function, developmental regulation, and signal transduction (Duff et al., 2015; Sibley et al., 2015). Its high prevalence in the nervous system likely reflects the greater intron length and the need for complex splicing programs that generate multiple isoforms. This stepwise splicing mechanism is thought to enhance overall splicing efficiency, reduce mis-splicing events and transcriptional backlogs, and coordinate with co-transcriptional processing in these gene contexts.

Because recursive splicing (RS) remains a comparatively underexplored field, there is currently no baseline-level estimate of its overall prevalence. Conservatively, RS appears to be extremely rare based on the evidence currently available, and its prevalence may plausibly be even lower than that of non-canonical splice sites or microexons (i.e., below 1%), although the possibility of substantial underestimation cannot be excluded. In one study of RS in long vertebrate genes, approximately 1.04% of a broad candidate set of unique, unannotated splice sites (419 of 40,163) matched the RS-site motif, yet only 9 high-confidence RS-sites remained after further filtering, suggesting that RS may indeed be highly uncommon (Sibley et al., 2015). At the same time, contemporaneous and subsequent studies have still not been able to establish a definitive estimate of the overall prevalence of RS (Duff et al., 2015; Moon and Zhao, 2022; Hoppe et al., 2023).

What can be stated more confidently at present is that RS is strongly enriched in long introns. A study using *Drosophila* as a genetic model reported that approximately 70% of long *Drosophila* introns (>40 kb) contain at least one RS-site (Kumari et al., 2022). Another genome-wide study in *Drosophila* further showed that 97% of all introns are smaller than the smallest recursive intron, providing additional indirect support for a close association between RS and very long introns (Duff et al., 2015). However, one review of RS has pointed out that previous searches for RS relied too heavily on very long introns as a defining criterion, potentially restricting the detectable scope of RS events (Hoppe et al., 2023). Given the uncertainty surrounding its overall prevalence, together with possible biases in search scope, RS remains a field with substantial room for further investigation.

5.3 Future perspectives of recursive splicing

Although recursive splicing has gained increasing recognition as an efficient strategy for processing long introns, its overall prevalence and regulatory mechanisms remain incompletely understood. Given that recursive splicing is a dynamic, multi-step process, many events are likely to evade detection by conventional short-read sequencing approaches. Long-read and nascent RNA sequencing may therefore reveal a substantially larger and more dynamic landscape of recursive splicing (Wan et al., 2021; Liu M. et al., 2023). From an evolutionary perspective, the reasons why long introns

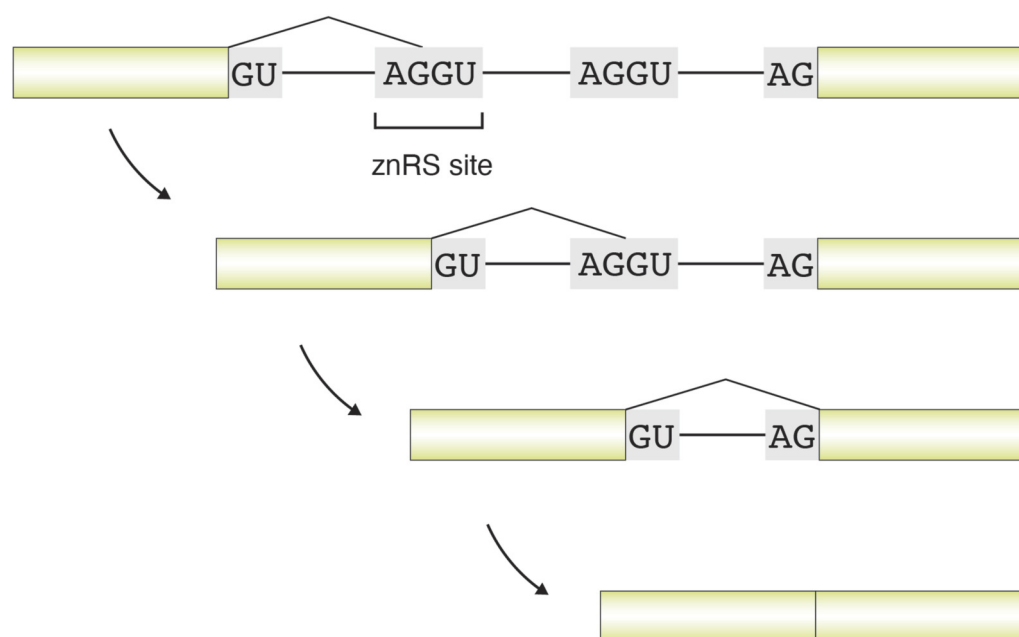


FIGURE 8

Schematic illustration of recursive splicing involving three consecutive splicing steps. In this example, a long intron contains two zero-nucleotide recursive splice (znRS) sites. In the first splicing step, the 5' donor contributed by the upstream exon is joined to the AG acceptor of the first znRS site. In the second step, the remaining GU of the first znRS site functions as a donor and is spliced to the AG acceptor of the second znRS site. In the third step, the remaining GU of the second znRS site serves as the donor and is joined to the 3' acceptor contributed by the downstream exon. This schematic depicts recursive splicing as a temporally ordered, kinetic splicing process.

and recursive splicing are conserved in nervous-system-associated genes warrant further investigation, and such insights may provide new entry points for the study of neurodegenerative diseases. Mechanistically, the multi-step and dynamic nature of recursive splicing suggest the presence of temporal and context-dependent regulation. How the spliceosome selectively recognizes functional recursive splice sites within long introns, and how this selection is coordinated with transcriptional dynamics, remain open questions.

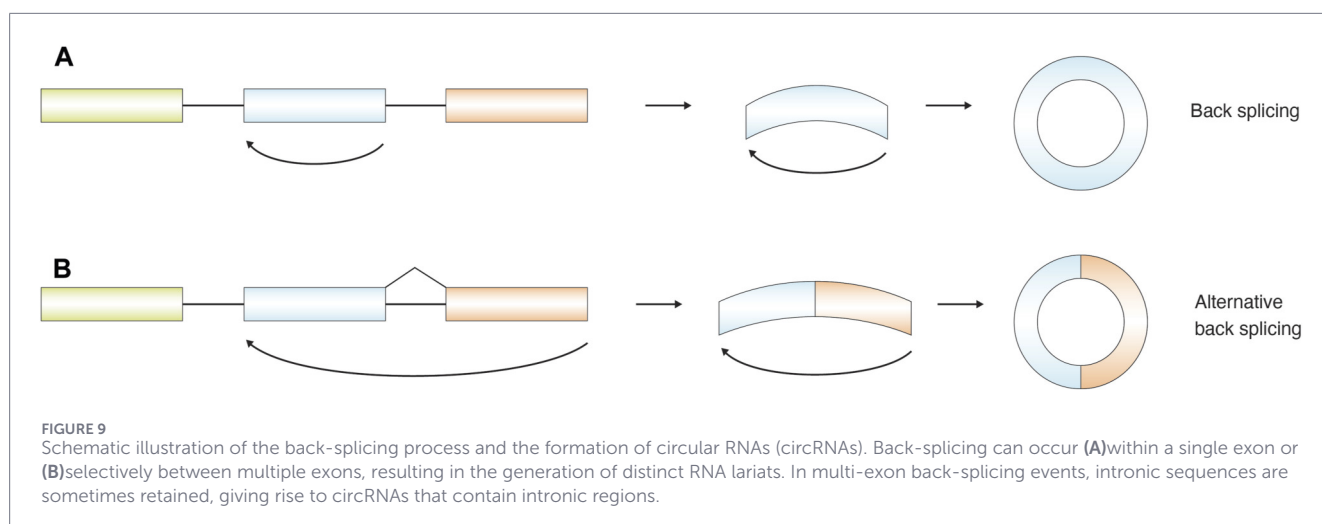
6 Back splicing and circular RNAs (circRNAs)

While recursive splicing extends splicing across time, other non-canonical events fundamentally alter the topology of exon ligation. Back-splicing represents such a topological departure, in which splice sites are joined in a reverse orientation to generate circular RNAs.

Back splicing is a special form of RNA splicing in which a downstream 5'splice donor site of an exon is joined to an upstream 3'splice acceptor, rather than to the 3'acceptor of a downstream exon as in canonical splicing. This atypical ligation generates a covalently closed circular RNA (circRNA) (Figure 9A). In some cases, back-splicing can occur across multiple exons. Canonical splicing events still take place internally, while the most upstream splice donor is joined to the most downstream splice acceptor, resulting in alternative back-splicing and the formation of circular RNAs containing multiple exons (Zhang et al., 2016; Chen, 2020) (Figure 9B).

6.1 Mechanism of back splicing and circRNA biogenesis

A covalently closed circular RNA (circRNA) is generated when a downstream 5'splice donor joins an upstream 3'splice acceptor—opposite to the direction of canonical (linear) splicing. Back-splicing can be mechanistically categorized into three principal models: intron-pairing-driven circularization, RNA-binding-protein (RBP)-mediated circularization, and lariat-driven circularization (Kristensen et al., 2019). In the intron-pairing-driven model, complementary Alu elements—short interspersed nuclear elements (SINEs) within flanking introns—form base-paired secondary structures that bring the 5'and 3'splice sites into proximity, thereby promoting exon back-splicing (Jeck et al., 2013). In the RBP-mediated model, specific RBPs recognize cis-acting motifs in upstream and downstream introns and bridge them to form a loop, facilitating circRNA formation (Ashwal-Fluss et al., 2014; Conn et al., 2015). The lariat-driven model involves the coexistence of exon skipping and back splicing: a spliced lariat containing skipped exons undergoes internal splicing to yield a circRNA (Zhang et al., 2013; 2016; Kelly et al., 2015). Notably, circRNA formation itself often undergoes alternative splicing, producing multiple circRNAs from a single gene through the use of alternative 5'or 3'back-splice sites, a process largely governed by competition among complementary intronic sequences (Zhang et al., 2016). Moreover, intronic regions can also circularize, generating circular intronic RNAs (ciRNAs) that lack exonic sequences (Zhang et al., 2013).



6.2 Molecular structures and translational potential of circRNAs

Owing to their covalently closed, head-to-tail circular structure, circRNAs lack free 5' and 3' ends and canonical splice junctions, making them largely resistant to exonucleolytic degradation and thereby substantially more stable than their linear RNA counterparts (Jeck et al., 2013; Enuke et al., 2016; Kristensen et al., 2019). Moreover, because circRNAs lack both a 5' cap and a 3' poly(A) tail, they are unable to initiate cap-dependent translation (Jeck et al., 2013; Enuke et al., 2016; Kristensen et al., 2019). Nevertheless, certain circRNAs have been shown to undergo translation via alternative mechanisms. For example, some harbor internal ribosome entry sites (IRES) that directly recruit ribosomal subunits in a cap-independent manner (Legnini et al., 2017; Yang et al., 2017). Others rely on an epitranscriptomic trigger in which N6-methyladenosine (m6A) modification recruits eIF3 and ribosomes to initiate translation independently of the 5' cap (Pamudurti et al., 2017; Yang et al., 2017; Zhou et al., 2019).

6.3 circRNAs as microRNA regulators

For a long time, circRNAs were primarily viewed as miRNA decoys or sponges that modulate miRNA-mediated regulation (Hansen et al., 2013). In parallel, numerous studies have attempted to engineer synthetic circRNAs targeting specific miRNAs, with the aim of exploiting their regulatory potential for the treatment of miRNA-associated diseases, including hepatitis C, gastric cancer, and type 2 diabetes (Rossbach, 2019; Avendaño-Portugal et al., 2025). However, it has gradually become apparent that circRNA-miRNA interactions are not limited to a simple sponge model. For example, CDR1as (CDR1 antisense) circRNA was initially proposed to function as a miR-7 sponge (Hansen et al., 2013; Memczak et al., 2013; Chen, 2020), but was later found to protect miR-7 from degradation mediated by the long noncoding RNA Cyrano (Kleaveland et al., 2018). This interplay forms a highly conserved regulatory network in the mammalian brain, where CDR1as modulates miR-7 availability and downstream target repression (Kleaveland et al., 2018). Loss of CDR1as leads to

downregulation of miR-7 targets and to behavioral changes in mice, highlighting the physiological relevance of circRNA-miRNA interactions. In addition, emerging evidence suggests that the relationship between neuronal circRNAs and miRNAs may be better characterized by coordinated localization and local regulatory interactions, rather than by straightforward sequestration (Woods and Van Vactor, 2021; Silenzi et al., 2024).

Taken together, these findings suggest that circRNAs may influence miRNA regulation through a broader range of mechanisms beyond sponge activity. At the same time, a purely sponge-based regulation depends on several stringent parameters, including expression level, number of binding sites, binding affinity, subcellular localization, and stoichiometric relationship (Nielsen et al., 2022). Because such conditions are unlikely to be broadly met under physiological circumstances (Thomson and Dinger, 2016; Nielsen et al., 2022), this further supports the view that many circRNA regulators may affect miRNA function through mechanisms other than simple sponging.

6.4 Discussion about back splicing and circRNAs

Back-splicing appears to be more prevalent than several of the other non-canonical splicing events discussed above, and this field is correspondingly more mature. Nevertheless, as research progresses and sequencing technologies continue to improve, additional conserved back-splicing events and circRNAs are still being identified, and their tissue-specific expression patterns warrant further investigation. Long-read sequencing analyses have indicated that 4.44% of reads in the human brain and 6.23% of reads in the mouse brain contain circRNA back-splice junctions (Rahimi et al., 2021b). Other studies have identified more than 880,000 back-splicing events in humans, potentially giving rise to more than 1.8 million human circRNA isoforms (Margvelani et al., 2025). Together, these observations indicate that back-splicing and circRNAs are widespread in mammals. Additional studies have further shown that circRNAs are present at broadly comparable orders of magnitude in the brain, pituitary, ovary, and liver, suggesting that circRNA expression is pervasive across

diverse tissues (Li et al., 2021); however, their tissue-specific and stage-specific expression patterns still require close attention.

Another important issue concerns the ongoing debate over the biological significance of back-splicing and circRNAs. CircRNAs have been widely reported to possess diverse biological functions, including the translational potential and microRNA-regulatory roles discussed above (Legnini et al., 2017; Pamudurti et al., 2017; Kleaveland et al., 2018), and their functions in regulation and coding have been broadly validated (Shi et al., 2023; Nemeth et al., 2024). At the same time, other studies have argued that mammalian circRNAs are produced primarily as by-products of splicing errors, and that as many as 97% of such events are deleterious (Xu and Zhang, 2021). One possible interpretation is that these two views are not necessarily contradictory. Rather, circRNAs with specific biological functions or conserved regulatory roles may represent only a minority, whereas the majority of circRNAs produced *in vivo* may indeed arise from erroneous back-splicing events. Accordingly, determining whether a circRNA is generated incidentally or is recurrently expressed within a given biological system may become one of the key criteria for distinguishing functional circRNAs from aberrant splicing products.

6.5 Future perspectives of back splicing and circRNAs

Benefiting from their exceptional resistance to exonucleolytic degradation, circRNAs hold promise as therapeutic agents, vaccine platforms, and gene-delivery vectors (Wesselhoeft et al., 2018; O'Leary et al., 2025). Engineered circRNAs exhibit prolonged half-lives and can support robust, sustained protein expression (Wesselhoeft et al., 2018; Chen, 2020). They have also been explored as vaccine platforms, including for SARS-CoV-2, where circRNA vaccines show improved stability compared to conventional mRNA vaccines (Qu et al., 2022; Niu et al., 2023). In cancer therapy, circRNAs can function as gene-therapy vectors, persisting and acting as gene medicines (Wesselhoeft et al., 2018). Moreover, shared molecular features of circRNAs in tumor cells provide a rationale for their use as cancer biomarkers, further broadening their diagnostic potential (Vo et al., 2019; Wang et al., 2021; Zhang et al., 2021). In addition, the regulatory effects of circRNAs on miRNAs, particularly those involving regulatory networks beyond simple sponging, remain an area with considerable potential for further investigation (Kleaveland et al., 2018; Woods and Van Vactor, 2021; Silenzi et al., 2024). Long-read sequencing has also provided critical evidence for the detection of circRNAs and the elucidation of back-splicing mechanisms, thereby compensating for the limitations of second-generation sequencing in identifying these non-canonical splicing events (Rahimi et al., 2021a). Back-splicing and circRNAs exhibit substantial potential in RNA regulation, such as through microRNA sponge activity, as well as in disease prevention and treatment, including applications as circRNA-based vaccines and cancer biomarkers.

7 Trans-splicing

Whereas back-splicing reshapes exon connectivity within a single transcript, trans-splicing further expands splicing plasticity

by linking exons derived from distinct pre-mRNA molecules, sometimes even across genes or chromosomes.

Trans-splicing is a form of non-canonical splicing in which exons derived from separate pre-mRNA precursors are joined to generate a single mature mRNA (Figure 10). Unlike canonical cis-like splicing, which ligates exons within a single transcript, trans-splicing can link exons originating from different pre-mRNAs of the same gene, from distinct genes, or even from different chromosomes, thereby expanding the diversity of mRNA transcripts and protein isoforms.

7.1 Structure and pattern of trans-splice sites

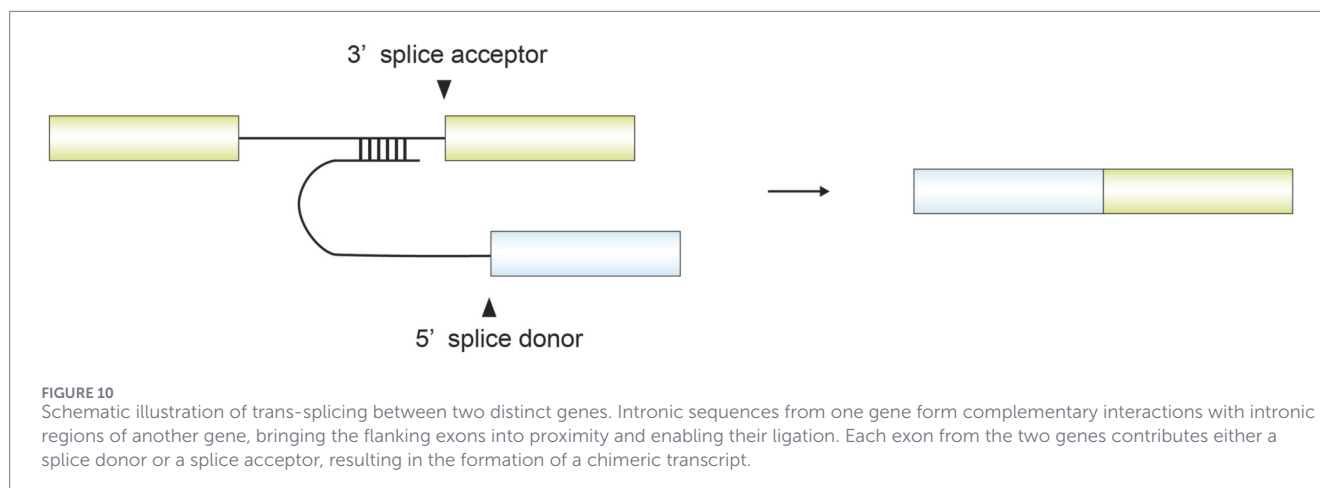
Trans-splicing follows the canonical GT-AG splice motif even though it joins exons derived from two different pre-mRNA molecules (Horiuchi and Aigaki, 2006). These splice junctions can be intragenic, arising from different promoters within the same gene, or intergenic, connecting exons from different genes or even different chromosomes (Horiuchi and Aigaki, 2006). The *mod* (*mdg4*) gene in *Drosophila* exemplifies an intragenic trans-splicing event: exon 5 is transcribed in the opposite direction to exons 1-4 and is joined to them through trans-splicing (Büchner et al., 2000; Dorn et al., 2001; Dorn and Krauss, 2003; McManus et al., 2010). At the other extreme, the human *ACAT1* gene provides an example of interchromosomal trans-splicing, in which a functional chimeric mRNA is formed from exons located on chromosome 7 and chromosome 1 (Li et al., 1999). This interchromosomal splicing has been shown to generate proteins with defined biological functions (Chen et al., 2008). Although all of these events share the canonical GT-AG splice motif, their underlying mechanisms vary, with some cases relying on RNA-RNA base pairing—such as complementary arm sequences—to bring distant transcripts into physical proximity (Dorn et al., 2001; Dorn and Krauss, 2003).

7.2 Functions of trans-splicing

Trans-splicing events between distinct pre-mRNAs directly promote the production of additional mRNA transcripts and protein isoforms, thereby diversifying gene expression and regulatory networks (Gingeras, 2009). In *Drosophila*, trans-splicing of the *mod* (*mdg4*) gene contributes more than 30 protein isoforms that influence chromatin organization and regulation (Büchner et al., 2000; Dorn and Krauss, 2003). In certain contexts, trans-splicing can also trigger interallelic complementation, in which homologous recombination between two alleles regenerates a functional mRNA transcript. This process can mediate gene repair or functional restoration (Horiuchi et al., 2003).

7.3 The prevalence of trans-splicing in *Caenorhabditis elegans* (*C. elegans*) is remarkably high

Early transcriptome-wide surveys estimated that roughly 70%–85% of *C. elegans* genes undergo trans-splicing, predominantly mediated by SL1 or SL2 leader sequences (Allen et al., 2011; Ramani et al., 2011). More recent long-read RNA-seq analyses have confirmed and extended these findings, showing that about



87.4% of genes are trans-spliced (Bernard et al., 2023), indicating that transcript processing in *C. elegans* follows a mechanism distinct from that of other eukaryotes.

7.4 Discussion about trans-splicing

The prevalence of trans-splicing is strongly dependent on both species and tissue context. For example, in the mitochondrial genomes of Pinaceae and Cupressaceae plants, where exons are dispersed, approximately 50%–70% of mitochondrial introns require trans-splicing (Guo et al., 2020). In *C. elegans*, 87.4% of genes undergo trans-splicing (Bernard et al., 2023). By contrast, in humans, trans-splicing is generally considered a rare event, and a substantial fraction of human trans-spliced chimeric RNAs appears to arise from chromosomal rearrangements in cancer cells (Yokomori et al., 2024). For instance, in one analytical set compiled from previously published human trans-spliced genes, only 52 genes were classified as trans-spliced, compared with 24,808 non-trans-spliced genes (Yokomori et al., 2024). In addition, one integrative transcriptome study reported that, among 24,498 detected non-co-linear splicing events, approximately 20%–40% represented mixtures of trans-spliced RNAs and circRNA isoforms (Chuang et al., 2018), further suggesting a potential connection between trans-splicing and circRNA biogenesis.

Although a subset of experimentally observed trans-splicing events has been confirmed to be functional, many are likely to represent biological noise or technical byproducts. Studies have shown that numerous reported cases may arise as artifacts during cDNA preparation or sequencing, in which reverse transcriptase and related enzymes can induce template switching when encountering direct repeat sequences, thereby generating chimeric reads that mimic trans-splicing events (Cocquet et al., 2006; Houseley and Tollervey, 2010; Kong et al., 2015; Balázs et al., 2019). Consequently, careful experimental validation and stringent bioinformatic filtering are essential to distinguish *bona fide* trans-splicing from artifacts and background noise.

Taken together, trans-splicing can exhibit strikingly different patterns of enrichment across species, tissues, and even individual genes. This indicates that current research has already achieved a

certain breadth in characterizing trans-splicing, but its relationships with chromosomal rearrangements, circRNAs, and disease warrant much deeper investigation. In addition, background noise and false-positive signals remain important concerns that must be carefully addressed in studies of trans-splicing.

7.5 Future perspectives of trans-splicing

Although false positives and background noise remain major challenges in the study of trans-splicing, emerging technologies and bioinformatic tools have begun to dispel some of the uncertainty surrounding this field (Liu et al., 2020). Meanwhile, attempts to integrate trans-splicing with CRISPR–Cas platforms to develop novel RNA-editing tools have also been reported (Fiflis et al., 2024; Zhu et al., 2024; Chandrasekaran et al., 2026). Continued development and application of these approaches will facilitate a more comprehensive understanding of trans-splicing. At the functional level, trans-splicing directly expands transcript and protein diversity and holds research potential in chromatin regulation, transcript complementation, and gene repair. *C. elegans*, in which trans-splicing operates as a predominant mechanism of transcript processing, offers a valuable evolutionary framework for exploring the origins and significance of trans-splicing. Moreover, as a shared question across many spliceosome-mediated non-canonical splicing events, the mechanisms by which the spliceosome tolerates trans-splicing between different genes or even across chromosomes remain to be elucidated.

8 Spliceosome-independent splicing

Collectively, the above non-canonical events illustrate the remarkable tolerance and adaptability of the spliceosome. However, not all RNA splicing reactions rely on spliceosomal machinery. In certain biological contexts, splicing is executed through entirely distinct enzymatic pathways.

Spliceosome-independent RNA splicing represents a mechanistically distinct pathway from canonical, spliceosome-mediated mRNA splicing. Two well-characterized examples include the endonuclease-mediated excision of introns from

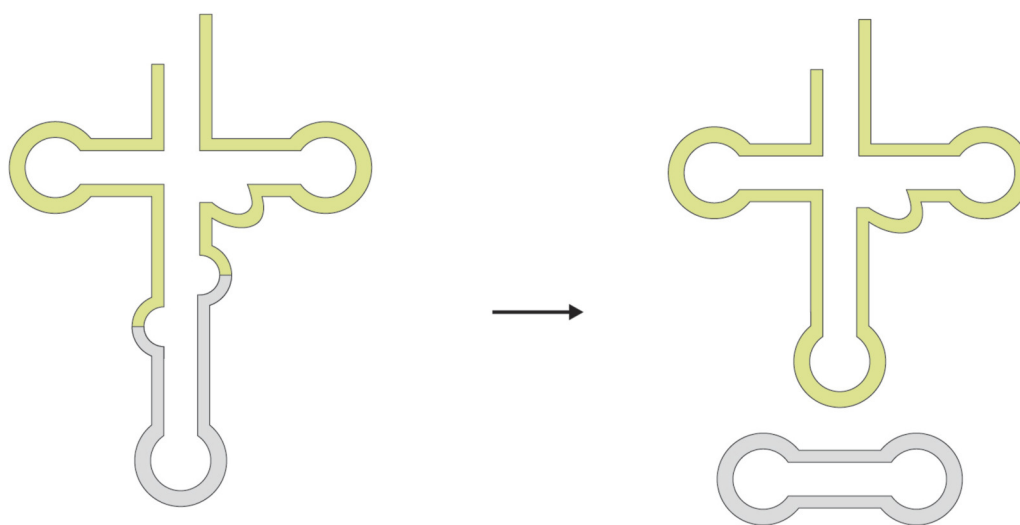


FIGURE 11

Simplified schematic of tRNA splicing. The intron within a precursor tRNA is excised by a site-specific endonuclease, generating an RNA lariat intermediate, after which the exons are ligated by a tRNA ligase. This process is carried out independently of the spliceosome.

precursor tRNAs in eukaryotes and archaea, and the non-canonical cytoplasmic splicing mediated by the ER stress sensor IRE1 during unfolded protein response.

8.1 tRNA splicing

tRNA splicing does not rely on the spliceosome, the ribonucleoprotein machinery used for canonical mRNA splicing but instead depends directly on specialized protein enzymes (Figure 11). Consequently, its reaction pathway differs substantially from canonical splicing. In eukaryotes, each pre-tRNA is first cleaved by a site-specific endonuclease, yielding two tRNA exon halves and a linear intron (Abelson et al., 1998). The exon halves are then ligated by tRNA ligase, producing a full-length tRNA bearing a 2'-phosphate at the splice junction, and this residual phosphate is subsequently removed by cyclic phosphodiesterase (CPDase) to yield a functional tRNA (Abelson et al., 1998). In archaea, pre-tRNAs contain a bulge-helix-bulge (BHB) structural motif that is precisely recognized and cleaved by the endonuclease (Belfort and Weiner, 1997; Trotta et al., 1997). Both eukaryotic and archaeal systems thus rely on spatial constraints and RNA secondary structure—rather than the GT-AG sequence motifs of canonical splicing—for accurate intron removal (Reinhold-Hurek and Shub, 1992; Abelson et al., 1998; Fujishima and Kanai, 2014). By contrast, enzyme-dependent tRNA splicing is rare in bacteria, where most introns are excised through self-splicing mediated by the tRNA's own structure (Belfort and Weiner, 1997; Trotta et al., 1997; Abelson et al., 1998).

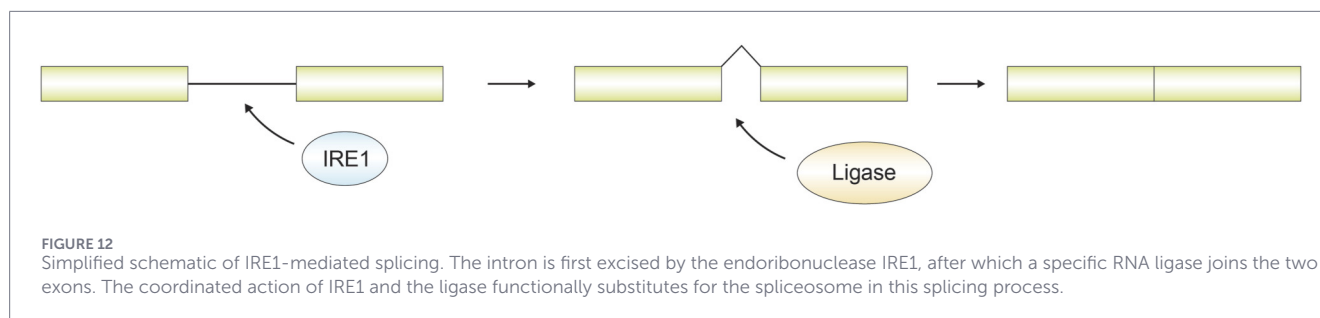
tRNA splicing is an obligate maturation step during pre-translation preparations, and is therefore widely observed in eukaryotes and archaea, although its prevalence varies across lineages and its underlying mechanisms differ substantially. In contrast, it is not a general feature of bacterial tRNA maturation. Although tRNA splicing has been widely studied, ongoing discoveries at the mechanistic and structural levels continue to

expand our understanding of this process. Certain enzymes have been identified that do not participate directly in tRNA splicing itself but instead modify tRNAs during post-splicing maturation (Hayne et al., 2022). During tRNA splicing, a circular RNA can also be generated, and the potential regulatory roles of this non-coding lariat in RNA regulation warrant further investigation. Moreover, several enzymes involved in tRNA splicing have been implicated in integrated stress responses and certain neurological disorders (Hurtig and van Hoof, 2023; Phizicky and Hopper, 2023), suggesting that further studies may provide important insights into the treatment of these diseases.

8.2 IRE1-mediated cytoplasmic splicing

Inositol-requiring enzyme 1 (IRE1) is an endoplasmic reticulum transmembrane sensor that is responsible for maintaining endoplasmic reticulum homeostasis and overall cellular function (Siwecka et al., 2021; Le Goupil et al., 2024). In certain contexts, IRE1 can function analogously to the spliceosome by mediating the splicing of specific mRNAs in the cytoplasm (Figure 12). This process is spliceosome-independent and, unlike canonical splicing, does not occur in the nucleus.

A representative example is the cytoplasmic splicing of X-box binding protein 1 (XBP1) mRNA in metazoans. The splicing of XBP1, a core transcription factor in the mammalian endoplasmic reticulum (ER) stress response, occurs in the cytoplasm and is mediated by the endoribonuclease IRE1 (Yoshida et al., 2001; Lee et al., 2002; Back et al., 2006; Uemura et al., 2009; Yanagitani et al., 2011). Under ER stress, IRE1 cleaves XBP1 pre-mRNA at two conserved RNA stem-loop structures that serve as splice sites, and an RNA ligase then joins the exons, allowing XBP1 to become functional and drive the transcription of molecular chaperones and ER-associated degradation (ERAD) components (Yoshida et al., 2001; Lee et al., 2002; Back et al., 2006; Uemura et al., 2009; Yanagitani et al., 2011). This cytoplasmic



cleavage enables a rapid restoration of ER homeostasis. Notably, XBP1 splicing bypasses the canonical GU–AG motif and does not generate a lariat intermediate; instead, the intron is directly removed by IRE1 (Yoshida et al., 2001; Lee et al., 2002; Back et al., 2006; Uemura et al., 2009; Yanagitani et al., 2011).

Another example of IRE1-mediated cytoplasmic splicing is the splicing of Homologous to ATF/CREB 1 (HAC1) in *Saccharomyces cerevisiae*. HAC1 is the functional homolog of XBP1 in yeast, and its splicing likewise occurs in the cytoplasm, with the intron being directly cleaved by IRE1 and the exons subsequently ligated by an RNA ligase (Xia, 2019). In higher plants, the bZIP60 mRNA likewise undergoes IRE1-mediated cytoplasmic splicing in response to ER stress (Kaur and Kaitheri Kandoth, 2021; Li et al., 2025).

Compared with the non-canonical splicing events discussed above, in which the spliceosome remains involved, IRE1-mediated splicing is target-restricted rather than globally prevalent and has been reported only in a very small number of distinct regulatory systems. IRE1 can cleave the transcripts described above to maintain endoplasmic reticulum homeostasis and support cell survival. Conversely, IRE1 can also target ER-associated RNAs for cleavage and subsequent degradation, a process that may, in some contexts, lead to cell death (Tam et al., 2014; Bartoszewska et al., 2023). How IRE1 discriminates between transcripts destined for productive splicing and those targeted for degradation remains an important open question. This dual role of IRE1 as a “double-edged sword” in ER signaling underscores its central regulatory importance and highlights its potential as a therapeutic target for drug development.

9 Conclusions and future directions

In this review, we have summarized and discussed multiple forms of non-canonical alternative splicing, outlining their principal structures, current landscapes, challenges, and prospective directions for future research (Table 3).

Non-canonical splicing comprises a broad and mechanistically diverse set of RNA-processing events. Accordingly, not all forms of non-canonical splicing should be discussed at the same analytical level. A useful organizing principle is to classify these events according to their degree of deviation from spliceosome-dependent splicing. For example, many cases of splicing with non-canonical splice sites, particularly the U2/U12-like subset, can still be tolerated by the spliceosome and are best regarded as spliceosome-mediated. Events that depart from canonical splicing in certain mechanistic aspects but still remain partially compatible with spliceosomal recognition may be more appropriately grouped as

spliceosome-compatible, including microexons, recursive splicing, and trans-splicing. By contrast, forms of non-canonical splicing that operate through fundamentally distinct mechanisms, such as tRNA splicing or the splicing of XBP1 and HAC1, are more appropriately classified as spliceosome-independent. These three categories differ substantially in mechanism and should therefore be analyzed separately. At the same time, this framework is not absolute. Some reported subclasses of non-canonical splicing may span conceptual boundaries; for instance, certain minor non-canonical splice junctions that do not conform to current U2/U12-like models may nonetheless rely on as-yet-uncharacterized mechanisms, potentially placing them closer to the spliceosome-compatible or even spliceosome-independent end of the spectrum. In such cases, individual events may require case-by-case discussion, as is already common for XBP1 and HAC1.

A second general conclusion is that reported prevalence is highly method dependent. Because published studies differ substantially in scope, period, and analytical design, the resulting prevalence estimates and reporting frameworks are often difficult to compare directly. Two issues are particularly important. First, the increasing use of long-read sequencing has not fundamentally overturned the overall view of non-canonical splicing, but it has revealed a large number of junctions that were not detectable in the short-read era, thereby exposing a substantial and still underexplored landscape. These newly detected long-read junctions represent an important resource for future investigation. Second, studies often differ markedly in pipeline stringency and filtering strategy, because different analytical aims naturally lead to different data-processing workflows and parameter choices. As a result, cross-study comparisons of non-canonical splicing require careful attention to methodology. When necessary, reanalysis of public raw datasets under a consistent filtering framework may provide a more reliable basis for direct comparison.

Species specificity and tissue specificity also emerge as important dimensions of non-canonical splicing. For example, trans-splicing is almost a dominant mode of transcript processing in *C. elegans*, whereas microexons, recursive splicing, and non-canonical splice sites have all been repeatedly linked to the nervous system in multiple studies. More broadly, these forms of potentially neural-associated non-canonical splicing may contribute to a deeper understanding of neuronal RNA processing and of neurological disease.

Finally, mechanistic stratification is likely to change interpretation. In the earlier RNA-seq reanalysis presented in this review, non-canonical introns were compared with canonical introns only as a single combined category. However, after

TABLE 3 Comparative summary of the major forms of non-canonical splicing discussed in this review, highlighting their current research landscape, key challenges, and future directions.

Class	Current landscape	Challenges	Possible future directions
Non-canonical splice sites	Long-read sequencing has revealed large numbers of novel, uncharacterized junctions Reported to be associated with NMD and neurological disease	Prevalence estimates are highly sensitive to analytical stringency Long-read novel-calling is pipeline- and artifact-sensitive Non-U2/U12-like junctions remain mechanistically heterogeneous and poorly resolved	Build long-read baseline surveys comparable to short-read-era studies Test disease relevance and tissue specificity, especially in neural and NMD-linked contexts Further investigate the different mechanisms of non-U2/U12-like non-canonical splicing
Non-canonical lncRNA splicing	Non-canonical splice site usage is more prevalent than in PCGs GC-AG introns are enriched, especially in first introns Weak splice signals may underlie high splicing plasticity	Mechanistic basis of lncRNA-specific non-canonical splicing remains unclear Current evidence remains limited compared with PCGs	Define the regulatory roles of non-canonical splicing in lncRNAs Clarify disease relevance, especially in cancer-associated lncRNA isoforms
Microexon splicing	Long-read sequencing is revealing additional microexons Specialized pipelines are improving microexon detection Single-cell full-length RNA-seq is enabling context-specific analysis	Reported prevalence is difficult to compare across studies because the underlying definitions differ Short-read RNA-seq and standard filtering likely underestimate microexons	Develop long-read-based and specialized pipelines for microexon detection Clarify how the spliceosome recognizes extremely short microexons Define the roles of microexons in neural regulation and neurological disease
Recursive splicing	Strongly enriched in very long introns Reported to be enriched in neuronal cells Recent studies suggest that earlier searches may have underestimated its scope	Overall prevalence remains poorly defined Detection is limited by short-read sequencing and search-scope bias Mechanistic regulation and site selection remain incompletely understood	Use long-read and nascent RNA sequencing to detect dynamic RS events Clarify how functional RS-sites are recognized within long introns Determine how RS is coordinated with transcriptional dynamics and neural gene regulation
Back splicing and circRNAs	Long-read sequencing is expanding the detectable landscape Recognized as widespread across multiple mammalian tissues; circRNAs are increasingly explored for RNA vaccines and biomarker applications	Distinction between functional circRNAs and aberrant by-products remains unresolved Most circRNAs may not be functionally equivalent despite widespread detection Tissue-specific and stage-specific expression patterns remain incompletely characterized	Use long-read sequencing to improve detection and analysis Develop criteria to distinguish recurrent functional circRNAs from incidental products Expand therapeutic and biomarker studies of circRNAs in disease contexts Explore broader miRNA-regulatory roles of circRNAs beyond conventional sponge activity
Trans-splicing	Emerging sequencing and bioinformatic tools are improving detection Beginning to be explored in CRISPR-based RNA-editing platforms	False positives and background noise remain major concerns Mechanistic links to chromosomal rearrangements, circRNAs, and disease remain incompletely resolved	Develop more stringent validation frameworks for <i>bona fide</i> trans-splicing Expand RNA-editing and gene-repair applications based on trans-splicing
tRNA splicing	Widely conserved in eukaryotes and archaea Reported to be implicated in integrated stress responses and certain neurological disorders	Functions of certain involving enzymes remain unclear Connection to specific stress responses and disorders remains unclear	Define how tRNA-splicing factors intersect with stress-response pathways Investigate the mechanistic links between tRNA splicing and neurological disease

(Continued)

TABLE 3 Continued

Class	Current landscape	Challenges	Possible future directions
IRE1-mediated splicing	A representative example of spliceosome-independent splicing Restricted to a small number of key ER stress-responsive transcripts Conserved across distinct systems	Additional forms of spliceosome-independent splicing are likely to remain unidentified How IRE1 distinguishes productive splicing from RNA decay remains unclear	Explore broader landscape of similar spliceosome-independent cases Clarify how IRE1 selects transcripts for splicing versus degradation Develop therapeutic strategies targeting IRE1-dependent stress signaling

further classification based on U2/U12-like scoring, a previously unanticipated conclusion emerged: non-canonical non-U2/U12-like introns also tended to be shorter than non-canonical U2/U12-like introns. Deeper subdivision and mechanistic categorization are essential for the study of non-canonical splicing. With the support of long-read sequencing, the field has already achieved considerable breadth; however, more refined mechanistic dissection at greater depth is likely to be especially valuable in the future.

10 Supplementary methods

10.1 RNA-seq mapping

Paired-end RNA-seq reads from EGFP+ cells sorted from E14.5 Tubb3-EGFP transgenic mice (NCBI Accession: SRR23308049) (Yang et al., 2023) were aligned to the mouse GRCm39 reference genome using STAR (v2.7.10b) (Dobin et al., 2013) with a GRCm39 genome index from NCBI RefSeq using --sjdbOverhang 59. Alignment was performed in single-pass mode with default splice-junction output filtering, and splice junctions were extracted from SJ.out.tab.

10.2 Data categorization and filtering

Using Python (v3.13.2), splice junctions in SJ.out.tab were first separated into canonical and non-canonical junctions and written to two corresponding output files. Only junctions supported by more than one uniquely mapped read (`unique_reads > 1`) were retained. Relevant features were preserved for downstream analysis, including chromosome, genomic start and end coordinates, strand, junction motif, gene name, and intron length.

Among all splice junctions identified by STAR, all junctions other than those with the GT-AG pattern (`motif_code ≠ 1`) were initially treated as non-canonical junctions and were subsequently subjected to further classification in downstream analyses based on their surrounding sequence features. This strategy was adopted because, in addition to GT-AG junctions, STAR also assigns several major classes of non-canonical junctions to predefined motif categories, for example `motif_code = 2` for CT-AC and `motif_code = 3` for GC-AG. Because the downstream analysis involved a more detailed PWM-based scoring framework for identifying U2/U12-like junctions, STAR's built-in classification of these major non-canonical classes was not used for the final classification step.

In addition, because STAR does not directly assign gene names to non-canonical junctions, non-canonical splice sites were further

remapped using the GRCm39 GTF annotation index. Putative gene assignment was inferred by determining whether each junction fell within the genomic span of an annotated gene.

All source code has been uploaded to GitHub (<https://github.com/SpringShann/Non-canonical-RNA-splicing-tools-for-STAR>), where the attached example inputs and outputs are taken from the original data of this study.

10.3 U2/U12-like categorization for non-canonical splice sites

Using canonical splice sites identified from the same sample, position weight matrices (PWMs) were constructed to assess whether non-canonical splice sites were consistent with U2/U12-like splicing. For canonical donor sites, a window comprising 3 nt upstream + the 2-nt donor site + 4 nt downstream (total length, 9 nt) was extracted. For canonical acceptor sites, a window comprising 18 nt upstream + the 2-nt acceptor site + 3 nt downstream (total length, 23 nt) was extracted. A longer window was used for acceptor sites because the natural sequence intervals typically associated with canonical donor and acceptor sites span approximately 10 bp and 28 bp, respectively (Caminsky et al., 2014). The trained PWMs were then used to score all non-canonical junctions, yielding a donor score, an acceptor score, and a total score for each junction. Scoring thresholds were determined using scrambled controls, in which randomized sequence windows were scored to generate a null score distribution. The threshold derived from this null distribution was then applied to the total score: junctions with total scores greater than the threshold were classified as U2/U12-like, whereas those with total scores less than or equal to the threshold were classified as non-U2/U12-like.

Non-canonical junctions were classified as not classified if no valid PWM score could be obtained, including cases in which strand orientation could not be resolved, sequence-window extraction failed, or the extracted window contained bases other than A, C, G, or T. These junctions were excluded from downstream statistical analyses.

The source code for this analysis is also available at GitHub (same depository).

10.4 4 × 4 dinucleotide heatmap analysis

Using Python (v3.13.2), the dinucleotide composition (i.e., motif patterns) of the non-canonical splice sites extracted in Section 10.2, Data categorization and filtering, was quantified. The results were visualized as separate 4 × 4 heatmaps for donor and acceptor

sites. For each site class, dinucleotides were first counted as raw occurrences and then normalized to relative frequencies by dividing by the total number of valid sites. For better visualization of the minor non-canonical dinucleotides, color intensities were mapped using a power-law normalization (PowerNorm, $\gamma = 0.5$) rather than a strictly linear scale. This transformation was applied only to the color mapping to improve the visual separation of low-frequency minor non-canonical dinucleotide classes when one or a few high-frequency classes dominated the distribution. The underlying count and frequency values, as well as the numeric annotations shown in each cell, were not transformed.

Because many non-canonical splicing events are non-canonical on only one side, with the opposite side still carrying a canonical dinucleotide, additional filtering was applied to remove GT from donor sites and AG from acceptor sites. Additional 4×4 heatmaps were then generated to visualize dinucleotide distributions after exclusion of these canonical site signals. In all heatmaps, rows correspond to the first nucleotide of the splice site and columns correspond to the second nucleotide.

The source code for this analysis and plotting is also available at GitHub (same depository).

10.5 Intron length comparison

Using Python (v3.13.2), intron lengths were extracted separately for canonical, non-canonical, non-canonical U2/U12-like, non-canonical non-U2/U12-like splice sites, respectively. The resulting distributions were visualized as (1) an empirical cumulative distribution function (ECDF) of intron lengths after $\log_{10}$ transformation, and (2) a boxplot of the raw intron lengths displayed on a log-scaled y-axis. In addition, a two-sided Mann–Whitney U test (Wilcoxon rank-sum test) was used to assess whether the distributions differed significantly. This nonparametric, rank-based test was selected because intron lengths are typically highly skewed and may span several orders of magnitude, making it more robust than parametric tests that assume approximate normality.

For the boxplot, outliers were defined according to the standard $1.5 \times$ interquartile range rule and were displayed as individual points; all other observations were represented by the box, median line, and whiskers rather than plotted individually.

The source code for this analysis and plotting is also available at GitHub (same depository).

Author contributions

JG: Formal Analysis, Methodology, Visualization, Writing – original draft, Writing – review and editing.

References

Abelson, J., Trotta, C. R., and Li, H. (1998). tRNA splicing*. *J. Biol. Chem.* 273, 12685–12688. doi:10.1074/jbc.273.21.12685

Abou Alezz, M., Celli, L., Belotti, G., Lisa, A., and Bione, S. (2020). GC-AG introns features in long non-coding and protein-coding genes suggest their role in gene expression regulation. *Front. Genet.* 11, 488. doi:10.3389/fgene.2020.00488

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Acknowledgements

The author would like to thank Jesse M. Gardner at Advanced Biotechnology Core (ABCORE) Facilities, Department of Biology, Georgia State University for providing servers for RNA-seq mappings and analysis. The author would also like to thank Junjie Guo at School of Medicine, Yale University for providing strong and constructive suggestions for this review.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. After the finishing of the original manuscript, the author used ChatGPT (OpenAI, version 4o and 5.4) for the purposes of language refinement and clarity improvements. All scientific content was written and reviewed by the author. The author has reviewed and edited the output and takes full responsibility for the content of this publication.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Akinyi, M. V., and Frilander, M. J. (2021). At the intersection of major and minor spliceosomes: crosstalk mechanisms and their impact on gene expression. *Front. Genet.* 12, 700744. doi:10.3389/fgene.2021.700744

Ali, A., Thorgaard, G. H., and Salem, M. (2021). PacBio Iso-Seq improves the rainbow trout genome annotation and identifies alternative splicing associated with economically important phenotypes. *Front. Genet.* 12, 683408. doi:10.3389/fgene.2021.683408

- Allen, M. A., Hillier, L. W., Waterston, R. H., and Blumenthal, T. (2011). A global analysis of *C. elegans* trans-splicing. *Genome Res.* 21, 255–264. doi:10.1101/gr.113811.110
- Ashwal-Fluss, R., Meyer, M., Pamudurti, N. R., Ivanov, A., Bartok, O., Hanan, M., et al. (2014). circRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell* 56, 55–66. doi:10.1016/j.molcel.2014.08.019
- Avendaño-Portugal, C., Montañó-Samaniego, M., Guttman-Bazbaz, R., Bravo-Estupiñan, D. M., and Ibáñez-Hernández, M. (2025). Therapeutic applications of Poly-miRNAs and miRNA sponges. *Int. J. Mol. Sci.* 26, 4535. doi:10.3390/ijms26104535
- Back, S. H., Lee, K., Vink, E., and Kaufman, R. J. (2006). Cytoplasmic IRE1 α -mediated XBP1 mRNA splicing in the absence of nuclear processing and endoplasmic reticulum stress. *J. Biol. Chem.* 281, 18691–18706. doi:10.1074/jbc.M602030200
- Balázs, Z., Tombácz, D., Csabai, Z., Moldován, N., Snyder, M., and Boldogkői, Z. (2019). Template-switching artifacts resemble alternative polyadenylation. *BMC Genomics* 20, 824. doi:10.1186/s12864-019-6199-7
- Bartoszewski, S., Sławski, J., Collawn, J. F., and Bartoszewski, R. (2023). Dual RNase activity of IRE1 as a target for anticancer therapies. *J. Cell Commun. Signal.* 17, 1145–1161. doi:10.1007/s12079-023-00784-5
- Belfort, M., and Weiner, A. (1997). Another bridge between kingdoms: tRNA splicing in archaea and eukaryotes. *Cell* 89, 1003–1006. doi:10.1016/S0092-8674(00)80287-1
- Bernard, F., Dargère, D., Rechavi, O., and Dupuy, D. (2023). Quantitative analysis of *C. elegans* transcripts by nanopore direct-cDNA sequencing reveals terminal hairpins in non trans-spliced mRNAs. *Nat. Commun.* 14, 1229. doi:10.1038/s41467-023-36915-0
- Bradley, R. K., and Anczukow, O. (2023). RNA splicing dysregulation and the hallmarks of cancer. *Nat. Rev. Cancer* 23, 135–155. doi:10.1038/s41568-022-00541-7
- Büchner, K., Roth, P., Schotta, G., Krauss, V., Saumweber, H., Reuter, G., et al. (2000). Genetic and molecular complexity of the position effect variegation modifier mod(mdg4) in drosophila. *Genetics* 155, 141–157. doi:10.1093/genetics/155.1.141
- Burnette, J. M., Miyamoto-Sato, E., Schaub, M. A., Conklin, J., and Lopez, A. J. (2005). Subdivision of large introns in drosophila by recursive splicing at nonexonic elements. *Genetics* 170, 661–674. doi:10.1534/genetics.104.039701
- Caminsky, N., Mucaki, E. J., and Rogan, P. K. (2014). Interpretation of mRNA splicing mutations in genetic disease: review of the literature and guidelines for information-theoretical analysis. *PLoS ONE* 9, 282. doi:10.1371/journal.pone.0092843
- Chandrasekaran, S. S., Tau, C., Fu, B. X. H., Nemeth, M., Bartie, L., Pawluk, A., et al. (2026). Rewriting endogenous human transcripts with dual CRISPR-guided 3' trans-splicing. *Cell Syst.* 17, 101487. doi:10.1016/j.cels.2025.101487
- Chen, L.-L. (2020). The expanding regulatory mechanisms and cellular functions of circular RNAs. *Nat. Rev. Mol. Cell Biol.* 21, 475–490. doi:10.1038/s41580-020-0243-y
- Chen, L.-L., and Kim, V. N. (2024). Small and long non-coding RNAs: past, present, and future. *Cell* 187, 6451–6485. doi:10.1016/j.cell.2024.10.024
- Chen, J., Zhao, X.-N., Yang, L., Hu, G.-J., Lu, M., Xiong, Y., et al. (2008). RNA secondary structures located in the interchromosomal region of human ACAT1 chimeric mRNA are required to produce the 56-kDa isoform. *Cell Res.* 18, 921–936. doi:10.1038/cr.2008.66
- Chuang, T.-J., Chen, Y.-J., Chen, C.-Y., Mai, T.-L., Wang, Y.-D., Yeh, C.-S., et al. (2018). Integrative transcriptome sequencing reveals extensive alternative trans-splicing and cis-backsplicing in human cells. *Nucleic Acids Res.* 46, 3671–3691. doi:10.1093/nar/gky032
- Cocquet, J., Chong, A., Zhang, G., and Veitia, R. A. (2006). Reverse transcriptase template switching and false alternative transcripts. *Genomics* 88, 127–131. doi:10.1016/j.ygeno.2005.12.013
- Conn, S. J., Pillman, K. A., Toubia, J., Conn, V. M., Salamanidis, M., Phillips, C. A., et al. (2015). The RNA binding protein quaking regulates formation of circRNAs. *Cell* 160, 1125–1134. doi:10.1016/j.cell.2015.02.014
- Dergai, M., Tsyba, L., Dergai, O., Zlatskii, I., Skrypkina, I., Kovalenko, V., et al. (2010). Microexon-based regulation of ITS1 and Src SH3 domains specificity relies on introduction of charged amino acids into the interaction interface. *Biochem. Biophysical Res. Commun.* 399, 307–312. doi:10.1016/j.bbrc.2010.07.080
- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–1789. doi:10.1101/gr.132159.111
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi:10.1093/bioinformatics/bts635
- Dorn, R., and Krauss, V. (2003). The modifier of mdg4 locus in drosophila: functional complexity is resolved by trans splicing. *Genetica* 117, 165–177. doi:10.1023/A:1022983810016
- Dorn, R., Reuter, G., and Loewendorf, A. (2001). Transgene analysis proves mRNA trans-splicing at the complex mod(mdg4) locus in drosophila. *Proc. Natl. Acad. Sci.* 98, 9724–9729. doi:10.1073/pnas.151268698
- Duff, M. O., Olson, S., Wei, X., Garrett, S. C., Osman, A., Bolisetty, M., et al. (2015). Genome-wide identification of zero nucleotide recursive splicing in drosophila. *Nature* 521, 376–379. doi:10.1038/nature14475
- El Marabti, E., Malek, J., and Younis, I. (2021). Minor intron splicing from basic science to disease. *Int. J. Mol. Sci.* 22, 6062. doi:10.3390/ijms22116062
- Enuka, Y., Lauriola, M., Feldman, M. E., Sas-Chen, A., Ulitsky, I., and Yarden, Y. (2016). Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor. *Nucleic Acids Res.* 44, 1370–1383. doi:10.1093/nar/gkv1367
- Erkelenz, S., Theiss, S., Kaisers, W., Ptok, J., Walotka, L., Müller, L., et al. (2018). Ranking noncanonical 5' splice site usage by genome-wide RNA-seq analysis and splicing reporter assays. *Genome Res.* 28, 1826–1840. doi:10.1101/gr.235861.118
- Fiflis, D. N., Rey, N. A., Venugopal-Lavanya, H., Sewell, B., Mitchell-Dick, A., Clements, K. N., et al. (2024). Repurposing CRISPR-Cas13 systems for robust mRNA trans-splicing. *Nat. Commun.* 15, 2325. doi:10.1038/s41467-024-46172-4
- Fu, Y., Kim, H., Roy, S., Huang, S., Adams, J. I., Grimes, S. M., et al. (2025). Single cell and spatial alternative splicing analysis with nanopore long read sequencing. *Nat. Commun.* 16, 6654. doi:10.1038/s41467-025-60902-2
- Fuentes-Beals, C., Olivares-Costa, M., Andrés, M. E., Haeger, P. A., Riadi, G., Oliva, C., et al. (2023). Bioinformatic analysis predicts that ethanol exposure during early development causes alternative splicing alterations of genes involved in RNA post-transcriptional regulation. *PLOS ONE* 18, e0284357. doi:10.1371/journal.pone.0284357
- Fujishima, K., and Kanai, A. (2014). tRNA gene diversity in the three domains of life. *Front. Genet.* 5, 142. doi:10.3389/fgene.2014.00142
- Georgomanolis, T., Sofiadis, K., and Papantonis, A. (2016). Cutting a long intron short: recursive splicing and its implications. *Front. Physiol.* 7, 598. doi:10.3389/fphys.2016.00598
- Gingras, T. R. (2009). Implications of chimaeric non-co-linear transcripts. *Nature* 461, 206–211. doi:10.1038/nature08452
- Guo, W., Zhu, A., Fan, W., Adams, R. P., and Mower, J. P. (2020). Extensive shifts from cis-to trans-splicing of gymnosperm mitochondrial introns. *Mol. Biol. Evol.* 37, 1615–1620. doi:10.1093/molbev/msaa029
- Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388. doi:10.1038/nature11993
- Hatton, A. R., Subramaniam, V., and Lopez, A. J. (1998). Generation of alternative ultrathorax isoforms and stepwise removal of a large intron by resplicing at exon–exon junctions. *Mol. Cell* 2, 787–796. doi:10.1016/S1097-2765(00)80293-2
- Hayashi, T., Ozaki, H., Sasagawa, Y., Umeda, M., Danno, H., and Nikaido, I. (2018). Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. *Nat. Commun.* 9, 619. doi:10.1038/s41467-018-02866-0
- Hayne, C. K., Lewis, T. A., and Stanley, R. E. (2022). Recent insights into the structure, function, and regulation of the eukaryotic transfer RNA splicing endonuclease complex. *WIREs RNA* 13, e1717. doi:10.1002/wrna.1717
- Hoppe, E. R., Udy, D. B., and Bradley, R. K. (2023). Recursive splicing discovery using lariats in total RNA sequencing. *Life Sci. Alliance* 6, e202201889. doi:10.26508/lsa.202201889
- Horiuchi, T., and Aigaki, T. (2006). Alternative trans-splicing: a novel mode of pre-mRNA processing. *Biol. Cell* 98, 135–140. doi:10.1042/BC20050002
- Horiuchi, T., Giniger, E., and Aigaki, T. (2003). Alternative trans-splicing of constant and variable exons of a drosophila axon guidance gene, lola. *Genes Dev.* 17, 2496–2501. doi:10.1101/gad.1137303
- Houseley, J., and Tollervey, D. (2010). Apparent non-canonical trans-splicing is generated by reverse transcriptase *in vitro*. *PLOS ONE* 5, e12271. doi:10.1371/journal.pone.0012271
- Humphrey, J., Brophy, E., Kosoy, R., Zeng, B., Coccia, E., Mattei, D., et al. (2025). Long-read RNA sequencing atlas of human microglia isoforms elucidates disease-associated genetic regulation of splicing. *Nat. Genet.* 57, 604–615. doi:10.1038/s41588-025-02099-0
- Hurtig, J. E., and van Hoof, A. (2023). An unknown essential function of tRNA splicing endonuclease is linked to the integrated stress response and intron debranching. *Genetics* 224, iyad044. doi:10.1093/genetics/iyad044
- Irimia, M., Weatheritt, R. J., Ellis, J. D., Parikhshak, N. N., Gonatopoulos-Pournatzis, T., Babor, M., et al. (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* 159, 1511–1523. doi:10.1016/j.cell.2014.11.035
- Jeck, W. R., Sorrentino, J. A., Wang, K., Slevin, M. K., Burd, C. E., Liu, J., et al. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 19, 141–157. doi:10.1261/rna.035667.112
- Kaur, N., and Kaitheri Kandoth, P. (2021). Tomato bZIP60 mRNA undergoes splicing in endoplasmic reticulum stress and in response to environmental stresses. *Plant Physiol. Biochem.* 160, 397–403. doi:10.1016/j.plaphy.2021.01.033
- Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., et al. (2013). Function of alternative splicing. *Gene* 514, 1–30. doi:10.1016/j.gene.2012.07.083
- Kelly, S., Greenman, C., Cook, P. R., and Papantonis, A. (2015). Exon skipping is correlated with exon circularization. *J. Mol. Biol.* 427, 2414–2417. doi:10.1016/j.jmb.2015.02.018

- Khan, M. R., Avino, M., Wellinger, R. J., and Laurent, B. (2023). Distinct regulatory functions and biological roles of lncRNA splice variants. *Mol. Ther. Nucleic Acids* 32, 127–143. doi:10.1016/j.omtn.2023.03.004
- Kilinc, M., Arora, V., Creson, T. K., Rojas, C., Le, A. A., Lauterborn, J., et al. (2022). Endogenous Syngap1 alpha splice forms promote cognitive function and seizure protection. *Elife* 11, e75707. doi:10.7554/eLife.75707
- Kleaveland, B., Shi, C. Y., Stefano, J., and Bartel, D. P. (2018). A network of noncoding regulatory RNAs acts in the mammalian brain. *Cell* 174, 350–362.e17. doi:10.1016/j.cell.2018.05.022
- Kong, Y., Zhou, H., Yu, Y., Chen, L., Hao, P., and Li, X. (2015). The evolutionary landscape of intergenic trans-splicing events in insects. *Nat. Commun.* 6, 8734. doi:10.1038/ncomms9734
- Kretova, M., Selicky, T., Cipakova, I., and Cipak, L. (2023). Regulation of Pre-mRNA splicing: indispensable role of post-translational modifications of splicing factors. *Life* 13, 604. doi:10.3390/life13030604
- Kristensen, L. S., Andersen, M. S., Stagsted, L. V. W., Ebbesen, K. K., Hansen, T. B., and Kjems, J. (2019). The biogenesis, biology and characterization of circular RNAs. *Nat. Rev. Genet.* 20, 675–691. doi:10.1038/s41576-019-0158-7
- Kumari, A., Sedehizadeh, S., Brook, J. D., Kozlowski, P., and Wojciechowska, M. (2022). Differential fates of introns in gene expression due to global alternative splicing. *Hum. Genet.* 141, 31–47. doi:10.1007/s00439-021-02409-6
- Le Goupil, S., Laprade, H., Aubry, M., and Chevet, E. (2024). Exploring the IRE1 interactome: from canonical signaling functions to unexpected roles. *J. Biol. Chem.* 300, 107169. doi:10.1016/j.jbc.2024.107169
- Lee, K., Tirasophon, W., Shen, X., Michalak, M., Prywes, R., Okada, T., et al. (2002). IRE1-mediated unconventional mRNA splicing and S2P-mediated ATF6 cleavage merge to regulate XBP1 in signaling the unfolded protein response. *Genes Dev.* 16, 452–466. doi:10.1101/gad.964702
- Legnini, I., Timoteo, G. D., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., et al. (2017). Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Mol. Cell* 66, 22–37.e9. doi:10.1016/j.molcel.2017.02.017
- Li, B. L., Li, X. L., Duan, Z. J., Lee, O., Lin, S., Ma, Z. M., et al. (1999). Human acyl-CoA:cholesterol acyltransferase-1 (ACAT-1) gene organization and evidence that the 4.3-kilobase ACAT-1 mRNA is produced from two different chromosomes. *J. Biol. Chem.* 274, 11060–11071. doi:10.1074/jbc.274.16.11060
- Li, Y. I., Sanchez-Pulido, L., Haerty, W., and Ponting, C. P. (2015). RBFOX and PTBP1 proteins regulate the alternative splicing of micro-exons in human brain transcripts. *Genome Res.* 25, 1–13. doi:10.1101/gr.181990.114
- Li, J., Shi, B., Wang, C., Shao, C., Liu, X., and Zhang, D. (2021). Comprehensive CircRNA profiling and selection of key CircRNAs reveal the potential regulatory roles of CircRNAs throughout ovarian development and maturation in *Cynoglossus semilaevis*. *Biology* 10, 830. doi:10.3390/biology10090830
- Li, H., Qin, Z., Geng, X., Cao, J., Yuan, X., Peng, H., et al. (2025). TaIRE1-mediated unconventional splicing of the TabZIP60 mRNA and the miR172 precursor regulates heat stress tolerance in wheat. *J. Integr. Plant Biol.* 67, 2388–2400. doi:10.1111/jipb.13963
- Lin, C.-F., Mount, S. M., Jarmolowski, A., and Makiolowski, W. (2010). Evolutionary dynamics of U12-type spliceosomal introns. *BMC Evol. Biol.* 10, 47. doi:10.1186/1471-2148-10-47
- Liu, C., Zhang, Y., Li, X., Jia, Y., Li, F., Li, J., et al. (2020). Evidence of constraint in the 3D genome for trans-splicing in human cells. *Sci. China Life Sci.* 63, 1380–1393. doi:10.1007/s11427-019-1609-6
- Liu, M., Zhu, J., Huang, H., Chen, Y., and Dong, Z. (2023a). Comparative analysis of nascent RNA sequencing methods and their applications in studies of cotranscriptional splicing dynamics. *Plant Cell* 35, 4304–4324. doi:10.1093/plcell/koad237
- Liu, Z., Zhu, C., Steinmetz, L. M., and Wei, W. (2023b). Identification and quantification of small exon-containing isoforms in long-read RNA sequencing data. *Nucleic Acids Res.* 51, e104. doi:10.1093/nar/gkad810
- Mackensen, T., and Irimia, M. (2025). From tiny exons to big insights: the expanding field of microexons. *Annu. Rev. Genomics Hum. Genet.* 26, 77–102. doi:10.1146/annurev-genom-121323-103648
- Marasco, L. E., and Kornblihtt, A. R. (2023). The physiology of alternative splicing. *Nat. Rev. Mol. Cell Biol.* 24, 242–254. doi:10.1038/s41580-022-00545-z
- Margvelani, G., Maquera, K. A. A., Welden, J. R., Rodgers, D. W., and Stamm, S. (2025). Translation of circular RNAs. *Nucleic Acids Res.* 53, gkae1167. doi:10.1093/nar/gkae1167
- Mattick, J. S., Amaral, P. P., Carninci, P., Carpenter, S., Chang, H. Y., Chen, L.-L., et al. (2023). Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat. Rev. Mol. Cell Biol.* 24, 430–447. doi:10.1038/s41580-022-00566-8
- McManus, C. J., and Graveley, B. R. (2011). RNA structure and the mechanisms of alternative splicing. *Curr. Opin. Genet. Dev.* 21, 373–379. doi:10.1016/j.gde.2011.04.001
- McManus, C. J., Duff, M. O., Eipper-Mains, J., and Graveley, B. R. (2010). Global analysis of trans-splicing in drosophila. *Proc. Natl. Acad. Sci.* 107, 12975–12979. doi:10.1073/pnas.1007586107
- Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333–338. doi:10.1038/nature11928
- Merens, H. E., Choquet, K., Baxter-Koenigs, A. R., and Churchman, L. S. (2024). Timing is everything: advances in quantifying splicing kinetics. *Trends Cell Biol.* 34, 968–981. doi:10.1016/j.tcb.2024.03.007
- Modrek, B., and Lee, C. (2002). A genomic view of alternative splicing. *Nat. Genet.* 30, 13–19. doi:10.1038/ng0102-13
- Moon, S., and Zhao, Y.-T. (2022). Recursive splicing is a rare event in the mouse brain. *PLOS ONE* 17, e0263082. doi:10.1371/journal.pone.0263082
- Nemeth, K., Bayraktar, R., Ferracin, M., and Calin, G. A. (2024). Non-coding RNAs in disease: from mechanisms to therapeutics. *Nat. Rev. Genet.* 25, 211–232. doi:10.1038/s41576-023-00662-1
- Nielsen, A. F., Bindereif, A., Bozzoni, I., Hanan, M., Hansen, T. B., Irimia, M., et al. (2022). Best practice standards for circRNA research. *Nat. Methods* 19, 1208–1220. doi:10.1038/s41592-022-01487-2
- Niu, D., Wu, Y., and Lian, J. (2023). Circular RNA vaccine in disease prevention and treatment. *Sig Transduct. Target Ther.* 8, 341. doi:10.1038/s41392-023-01561-x
- Ouyang, J., Zhong, Y., Zhang, Y., Yang, L., Wu, P., Hou, X., et al. (2022). Long non-coding RNAs are involved in alternative splicing and promote cancer progression. *Br. J. Cancer* 126, 1113–1124. doi:10.1038/s41416-021-01600-w
- O'Leary, E., Jiang, Y., Kristensen, L. S., Hansen, T. B., and Kjems, J. (2025). The therapeutic potential of circular RNAs. *Nat. Rev. Genet.* 26, 230–244. doi:10.1038/s41576-024-00806-x
- Pamudurti, N. R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., et al. (2017). Translation of CircRNAs. *Mol. Cell* 66, 9–21.e7. doi:10.1016/j.molcel.2017.02.021
- Parada, G. E., Munita, R., Cerda, C. A., and Gysling, K. (2014). A comprehensive survey of non-canonical splice sites in the human transcriptome. *Nucleic Acids Res.* 42, 10564–10578. doi:10.1093/nar/gku744
- Phizicky, E. M., and Hopper, A. K. (2023). The life and times of a tRNA. *RNA* 29, 898–957. doi:10.1261/rna.079620.123
- Pisignano, G., and Ladomery, M. (2021). Epigenetic regulation of alternative splicing: how lncRNAs tailor the message. *Non-Coding RNA* 7, 21. doi:10.3390/ncrna7010021
- Pucker, B., and Brockington, S. F. (2018). Genome-wide analyses supported by RNA-seq reveal non-canonical splice sites in plant genomes. *BMC Genomics* 19, 980. doi:10.1186/s12864-018-5360-z
- Qu, L., Yi, Z., Shen, Y., Lin, L., Chen, F., Xu, Y., et al. (2022). Circular RNA vaccines against SARS-CoV-2 and emerging variants. *Cell* 185, 1728–1744.e16. doi:10.1016/j.cell.2022.03.044
- Quesnel-Vallières, M., Dargaei, Z., Irimia, M., Gonatopoulos-Pournatzis, T., Ip, J. Y., Wu, M., et al. (2016). Misregulation of an activity-dependent splicing network as a common mechanism underlying autism spectrum disorders. *Mol. Cell* 64, 1023–1034. doi:10.1016/j.molcel.2016.11.033
- Rahimi, K., Færch Nielsen, A., Venø, M. T., and Kjems, J. (2021a). Nanopore long-read sequencing of circRNAs. *Methods* 196, 23–29. doi:10.1016/j.ymeth.2021.09.010
- Rahimi, K., Venø, M. T., Dupont, D. M., and Kjems, J. (2021b). Nanopore sequencing of brain-derived full-length circRNAs reveals circRNA-specific exon usage, intron retention and microexons. *Nat. Commun.* 12, 4825. doi:10.1038/s41467-021-24975-z
- Ramani, A. K., Calarco, J. A., Pan, Q., Mavandadi, S., Wang, Y., Nelson, A. C., et al. (2011). Genome-wide analysis of alternative splicing in *Caenorhabditis elegans*. *Genome Res.* 21, 342–348. doi:10.1101/gr.114645.110
- Reinhold-Hurek, B., and Shub, D. A. (1992). Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature* 357, 173–176. doi:10.1038/357173a0
- Rosbach, O. (2019). Artificial circular RNA sponges targeting MicroRNAs as a novel tool in molecular biology. *Mol. Ther. Nucleic Acids* 17, 452–454. doi:10.1016/j.omtn.2019.06.021
- Shi, X., Liao, S., Bi, Z., Liu, J., Li, H., and Feng, C. (2023). Newly discovered circRNAs encoding proteins: recent progress. *Front. Genet.* 14, 1264606. doi:10.3389/fgene.2023.1264606
- Sibley, C. R., Emmett, W., Blazquez, L., Faro, A., Haberman, N., Briese, M., et al. (2015). Recursive splicing in long vertebrate genes. *Nature* 521, 371–375. doi:10.1038/nature14466
- Sibley, C. R., Blazquez, L., and Ule, J. (2016). Lessons from non-canonical splicing. *Nat. Rev. Genet.* 17, 407–421. doi:10.1038/nrg.2016.46
- Silenzi, V., D'Ambr, E., Santini, T., D'Uva, S., Setti, A., Salvi, N., et al. (2024). A tripartite circRNA/mRNA/miRNA interaction regulates glutamatergic signaling in the mouse brain. *Cell Rep.* 43, 114766. doi:10.1016/j.celrep.2024.114766
- Siwecka, N., Rozpedek-Kamińska, W., Wawrzynkiewicz, A., Pytel, D., Diehl, J. A., and Majsterek, I. (2021). The structure, activation and signaling of IRE1 and its role in determining cell fate. *Biomedicines* 9, 156. doi:10.3390/biomedicines9020156

- Statello, L., Guo, C.-J., Chen, L.-L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* 22, 96–118. doi:10.1038/s41580-020-00315-9
- Su, Y., Yu, Z., Jin, S., Ai, Z., Yuan, R., Chen, X., et al. (2024). Comprehensive assessment of mRNA isoform detection methods for long-read sequencing data. *Nat. Commun.* 15, 3972. doi:10.1038/s41467-024-48117-3
- Tam, A. B., Koong, A. C., and Niwa, M. (2014). Ire1 has distinct catalytic mechanisms for XBP1/HAC1 splicing and RIDD. *Cell Rep.* 9, 850–858. doi:10.1016/j.celrep.2014.09.016
- Tardaguila, M., de la Fuente, L., Marti, C., Pereira, C., Pardo-Palacios, F. J., Del Risco, H., et al. (2018). SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome Res.* 28, 396–411. doi:10.1101/gr.222976.117
- Thomson, D. W., and Dinger, M. E. (2016). Endogenous microRNA sponges: evidence and controversy. *Nat. Rev. Genet.* 17, 272–283. doi:10.1038/nrg.2016.20
- Trotta, C. R., Miao, F., Arn, E. A., Stevens, S. W., Ho, C. K., Rauhut, R., et al. (1997). The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell* 89, 849–858. doi:10.1016/S0092-8674(00)80270-6
- Turunen, J. J., Niemelä, E. H., Verma, B., and Frilander, M. J. (2013). The significant other: splicing by the minor spliceosome. *Wiley Interdiscip. Rev. RNA* 4, 61–76. doi:10.1002/wrna.1141
- Uemura, A., Oku, M., Mori, K., and Yoshida, H. (2009). Unconventional splicing of XBP1 mRNA occurs in the cytoplasm during the mammalian unfolded protein response. *J. Cell Sci.* 122, 2877–2886. doi:10.1242/jcs.040584
- Ustianenko, D., Weyn-Vanhentenryck, S. M., and Zhang, C. (2017). Microexons: discovery, regulation, and function. *WIREs RNA* 8, e1418. doi:10.1002/wrna.1418
- Vo, J. N., Cieslik, M., Zhang, Y., Shukla, S., Xiao, L., Zhang, Y., et al. (2019). The landscape of circular RNA in cancer. *Cell* 176, 869–881.e13. doi:10.1016/j.cell.2018.12.021
- Vuong, C. K., Black, D. L., and Zheng, S. (2016). The neurogenetics of alternative splicing. *Nat. Rev. Neurosci.* 17, 265–281. doi:10.1038/nrn.2016.27
- Wan, Y., Anastasakis, D. G., Rodriguez, J., Palangat, M., Gudla, P., Zaki, G., et al. (2021). Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. *Cell* 184, 2878–2895.e20. doi:10.1016/j.cell.2021.04.012
- Wang, S., Zhang, K., Tan, S., Xin, J., Yuan, Q., Xu, H., et al. (2021). Circular RNAs in body fluids as cancer biomarkers: the new frontier of liquid biopsies. *Mol. Cancer* 20, 13. doi:10.1186/s12943-020-01298-z
- Wang, X., Hua, J., Li, J., Zhang, J., Dzakah, E. E., Cao, G., et al. (2022). Mechanisms of non-coding RNA-modulated alternative splicing in cancer. *RNA Biol.* 19, 541–547. doi:10.1080/15476286.2022.2062846
- Wesselhoeft, R. A., Kowalski, P. S., and Anderson, D. G. (2018). Engineering circular RNA for potent and stable translation in eukaryotic cells. *Nat. Commun.* 9, 2629. doi:10.1038/s41467-018-05096-6
- Woods, B. J., and Van Vactor, D. (2021). miRNA: local guardians of presynaptic function in plasticity and disease. *RNA Biol.* 18, 1014–1024. doi:10.1080/15476286.2020.1871214
- Wright, D. J., Hall, N. A. L., Irish, N., Man, A. L., Glynn, W., Mould, A., et al. (2022). Long read sequencing reveals novel isoforms and insights into splicing regulation during cell state changes. *BMC Genomics* 23, 42. doi:10.1186/s12864-021-08261-2
- Wu, T., and Fu, X.-D. (2015). Genomic functions of U2AF in constitutive and regulated splicing. *RNA Biol.* 12, 479–485. doi:10.1080/15476286.2015.1020272
- Xia, X. (2019). Translation control of HAC1 by regulation of splicing in *Saccharomyces cerevisiae*. *Int. J. Mol. Sci.* 20, 2860. doi:10.3390/ijms20122860
- Xu, C., and Zhang, J. (2021). Mammalian circular RNAs result largely from splicing errors. *Cell Rep.* 36, 109439. doi:10.1016/j.celrep.2021.109439
- Yanagitani, K., Kimata, Y., Kadokura, H., and Kohno, K. (2011). Translational pausing ensures membrane targeting and cytoplasmic splicing of XBP1u mRNA. *Science* 331, 586–589. doi:10.1126/science.1197142
- Yang, Q., Du, W. W., Wu, N., Yang, W., Awan, F. M., Fang, L., et al. (2017). A circular RNA promotes tumorigenesis by inducing c-myc nuclear translocation. *Cell Death Differ.* 24, 1609–1620. doi:10.1038/cdd.2017.86
- Yang, R., Feng, X., Arias-Cavieres, A., Mitchell, R. M., Polo, A., Hu, K., et al. (2023). Upregulation of SYNGAP1 expression in mice and human neurons by redirecting alternative splicing. *Neuron* 111, 1637–1650.e5. doi:10.1016/j.neuron.2023.02.021
- Yokomori, R., Kusakabe, T. G., and Nakai, K. (2024). Characterization of trans-spliced chimeric RNAs: insights into the mechanism of trans-splicing. *Nar. Genom. Bioinform.* 6, lqae067. doi:10.1093/nargab/lqae067
- Yoshida, H., Matsui, T., Yamamoto, A., Okada, T., and Mori, K. (2001). XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell* 107, 881–891. doi:10.1016/S0092-8674(01)00611-0
- Zhang, Y., Zhang, X.-O., Chen, T., Xiang, J.-F., Yin, Q.-F., Xing, Y.-H., et al. (2013). Circular intronic long noncoding RNAs. *Mol. Cell* 51, 792–806. doi:10.1016/j.molcel.2013.08.017
- Zhang, X.-O., Dong, R., Zhang, Y., Zhang, J.-L., Luo, Z., Zhang, J., et al. (2016). Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* 26, 1277–1287. doi:10.1101/gr.202895.115
- Zhang, Y., Wang, Y., Su, X., Wang, P., and Lin, W. (2021). The value of circulating circular RNA in cancer diagnosis, monitoring, prognosis, and guiding treatment. *Front. Oncol.* 11, 736546. doi:10.3389/fonc.2021.736546
- Zhou, K. I., Shi, H., Lyu, R., Wylder, A. C., Matuszek, Z., Pan, J. N., et al. (2019). Regulation of Co-transcriptional Pre-mRNA splicing by m6A through the low-complexity protein hnRNP G. *Mol. Cell* 76, 70–81.e9. doi:10.1016/j.molcel.2019.07.005
- Zhu, G., Zhou, X., Wen, M., Qiao, J., Li, G., and Yao, Y. (2024). CRISPR-Cas13: pioneering RNA editing for nucleic acid therapeutics. *BioDesign Res.* 6, 0041. doi:10.34133/bdr.0041