Cutting a Long Intron Short: Recursive Splicing and Its Implications

Theodore Georgomanolis, Konstantinos Sofiadis and Argyris Papantonis *

Chromatin Systems Biology Laboratory, Center for Molecular Medicine, University of Cologne, Cologne, Germany

Over time eukaryotic genomes have evolved to host genes carrying multiple exons separated by increasingly larger intronic, mostly non-protein-coding, sequences. Initially, little attention was paid to these intronic sequences, as they were considered not to contain regulatory information. However, advances in molecular biology, sequencing, and computational tools uncovered that numerous segments within these genomic elements do contribute to the regulation of gene expression. Introns are differentially removed in a cell type-specific manner to produce a range of alternatively-spliced transcripts, and many span tens to hundreds of kilobases. Recent work in human and fruitfly tissues revealed that long introns are extensively processed cotranscriptionally and in a stepwise manner, before their two flanking exons are spliced together. This process, called "recursive splicing," often involves non-canonical splicing elements positioned deep within introns, and different mechanisms for its deployment have been proposed. Still, the very existence and widespread nature of recursive splicing offers a new regulatory layer in the transcript maturation pathway, which may also have implications in human disease.

Keywords: recursive splicing, variant U1 RNAs, processing, exon definition, RNA polymerase, co-transcriptional

INTRODUCTION

The interruption of a gene's open reading frame by a non-protein-coding sequence, i.e., by an intron, is an exclusive feature of eukaryotes. It is now thought that the course of evolution has brought about such an exon-intron gene structure concomitantly with the emergence and diversification of multicellular eukaryotes (Rogozin et al., 2012) and the need for complex gene regulation (Jeffares et al., 2008). However, introns are not "genomic junk"; they have been shown to confer important regulatory capacity, they typically carry *cis*-regulatory elements important for both transcription and splicing (Wang and Burge, 2008; Levine, 2010), and have even been found to be partially or fully coding (Marquez et al., 2015).

An average mammalian gene will contain 8–9 introns; >3000 human introns are longer than 50 kbp, and >1200 longer than 100 kbp (Bradnam and Korf, 2008; Shepard et al., 2009). This poses the following problem. In long introns the three sites reactive in a splicing reaction (i.e., the 5′ splicing site, the branch-point, and the 3′ splice site; Hollander et al., 2016) will be separated by large stretches of RNA sequence. Thus, it becomes difficult to explain how the sites required for splicing can find one another in three-dimensional space, or how a primary transcript spanning tens to hundreds of kbp can be protected from unspecific hydrolytic cleavage in the time it takes an RNA polymerase to copy it as one continuous RNA (e.g., at an average speed of 3 kbp/min, >30 min are required to fully transcribe a 100 kbp-long intron; Wada et al., 2009).

An elegant solution to this problem was proposed for Drosophila long introns—recursive splicing (RS). According to this, long introns are removed in a stepwise manner by splicing at intronic sites that carry the expected acceptor and donor splice sequences in the three gene examples studied (consensus sequence: $5'$-$(Y)_n$NCAG|GTAAGT-$3'$; the vertical line represents the splicing junction; Burnette et al., 2005). Similarly, a "zero-length" exon was identified between the 2nd and 3rd exon of the rat α-tropomyosin gene (Grellscheid and Smith, 2006), as well as "dual specificity" splicing sites in human pre-mRNAs (Zhang et al., 2007). Still, despite computational efforts (Shepard et al., 2009), the RS concept was not verified in humans until 2015. A study in human primary endothelial cells (Kelly et al., 2015), followed by two back-to-back studies across Drosophila tissues (Duff et al., 2015) and in human brain (Sibley et al., 2015), revealed that RS is a conserved and widespread splicing mechanism. Nonetheless, the fruitfly and human RS-sites differ in composition, and their molecular recognition and processing remains unknown. Here, we discuss different scenarios by which recursive splicing might manifest, as well as its potential implications in gene expression regulation and deregulation.

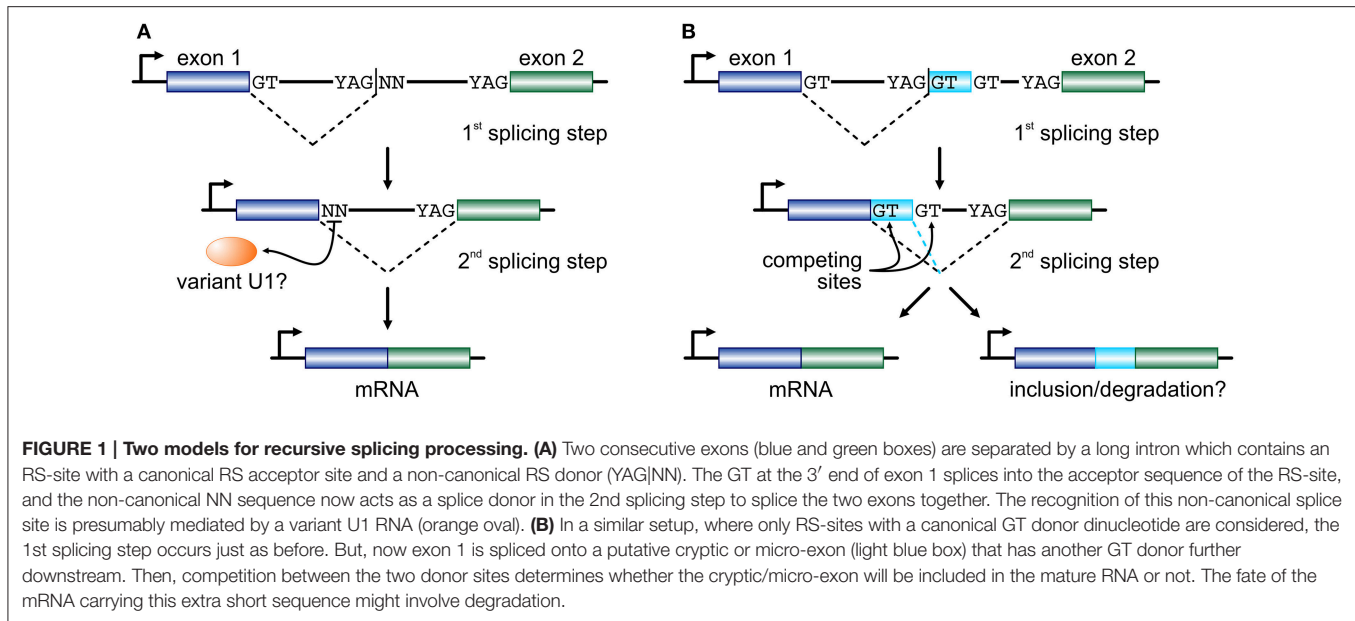## MODELS FOR THE PROCESSING OF RECURSIVE SPLICING INTERMEDIATES

The idea that intronic sequences are not evolutionarily constrained, because they do not code for proteins, pervades our thinking; however, the conservation of parts of these non-coding sequences between three diverse mammalian genomes (human, whale, and seal) amounts to almost 50% in pairwise comparisons, and to 28% amongst the three taxa (Hare and Palumbi, 2003). This hints to the existence of underappreciated classes of intronic regulatory elements. Recent work on recursive splicing in human cells (Kelly et al., 2015; Sibley et al., 2015) in part confirms this by using deep RNA sequencing and data analysis to find potential "ratchet" RS points. A large number of RS-sites was discovered (albeit different in the two studies, due to the different approaches and cutoffs used), the conservation of which was higher than that of similar, adjacent, intronic regions. These do not carry the consensus sequence identified in Drosophila, but rather one that contains a typical acceptor site followed by a donor sequence that is not the expected GT/GC/GA in >60% of cases (Kelly et al., 2015). This, of course, raises the question of how these non-canonical sites are recognized by the splicing machinery and processed accurately to produce a mature messenger RNA (although RNase R-resistant lariats as a result of recursive splicing were detected; Duff et al., 2015; Kelly et al., 2015).

One scenario could be that the vast majority of RS events detected, especially those with non-GT sequences at donor sites, represent "dead-end" products targeted for degradation. But, in human primary endothelial cells, a number of evidence does not concur with this scenario. First, the ~2400 RS high-confidence events recorded occur at ~15% the level of primary transcription; second, targeted genome editing of three different RS-sites in the 134 kbp-long intron of the *SAMD4A* gene showed that they

are necessary for efficient mRNA production; third, knocking-down exosome components did not affect the levels of RS intermediates, either GT- or non-GT-containing (Kelly et al., 2015). Thus, splicing at RS-sites occurs at significant levels, is widespread, and does not appear linked to exosomal degradation, but rather to RNA maturation.

If RS intermediates lie on the productive pathway of mRNAs, the dinucleotide immediately downstream of an RS-junction will subsequently need to act as an efficient splicing donor. In endothelial cells, ~45% of RS-sites encode a GN dinucleotide and it has been shown that they can efficiently function as donors provided strong acceptor and "splicing enhancer" sequences also partake in that reaction (Twigg et al., 1998; Thanaraj and Clark, 2001; Dewey et al., 2006). For the remaining 55% of RS-sites, a combination of mechanisms might come into play. We now know that the U1-containing snRNPs, designed to identify the GT donor dinucleotide, are able to expand their base-pairing repertoire via mispairing (Roca et al., 2012; Tan et al., 2016). We have also come to find out that the human genome encodes a large number of "variant" U1 snRNAs (Kyriakopoulou et al., 2006; O'Reilly et al., 2013). Their expression is markedly higher in primary, embryonic, and pluripotent cells (O'Reilly et al., 2013; Kelly et al., 2015; Vazquez-Arango et al., 2016) and they are able to form proper RNPs *in vitro* (Somarelli et al., 2014). In endothelial cells, the repertoire of expressed variant U1, together with the minor spliceosome (Turunen et al., 2013), would suffice for the recognition of the vast majority of all non-canonical RS donor dinucleotides recorded (Kelly et al., 2015). In addition, efficient splicing has been shown to occur independently of U1-mediated recognition (Raponi and Baralle, 2008) or of the physical continuity of the nascent transcript (via "exon tethering"; Dye et al., 2006). With the aforementioned into account, we propose that long human introns are cotranscriptionally removed by splicing at RS-sites that may equally carry a canonical or a non-canonical donor dinucleotide, before the two flanking exons are joined together (**Figure 1A**).

Another model, proposed on the basis of data from human brain, sees RS-sites as a means for establishing a "binary splicing switch" (Sibley et al., 2015). However, it is worth noting here that this study focuses specifically on RS-sites that conform to the YAG|GT consensus, and thus investigated ~400 such junctions. According to this model, each RS-site may also act as an RS-exon whereby the GT dinucleotide immediately downstream of the splice site will compete with an alternative GT further downstream for splicing into the canonical acceptor site at the $3'$ end of the long intron. This inter-site competition determines whether the very short RS-exon sequence will be retained as part of the final spliced transcript or not (**Figure 1B**; a mechanism similar to "intrasplicing"; Parra et al., 2008). It is suggested that inclusion of such RS-exons will target the mature transcript for degradation, as they encode premature termination codons (Sibley et al., 2015). However, their inclusion (if in-frame) will act on top of alternative splicing, and brain tissue was shown to be uniquely prone to the inclusion of microexons into mature mRNAs (Scheckel and Darnell, 2015), and this may not be perfectly reconciled with this RS model. Still, despite their

FIGURE 1 | Two models for recursive splicing processing. (A) Two consecutive exons (blue and green boxes) are separated by a long intron which contains an RS-site with a canonical RS acceptor site and a non-canonical RS donor (YAG|NN). The GT at the 3′ end of exon 1 splices into the acceptor sequence of the RS-site, and the non-canonical NN sequence now acts as a splice donor in the 2nd splicing step to splice the two exons together. The recognition of this non-canonical splice site is presumably mediated by a variant U1 RNA (orange oval). **(B)** In a similar setup, where only RS-sites with a canonical GT donor dinucleotide are considered, the 1st splicing step occurs just as before. But, now exon 1 is spliced onto a putative cryptic or micro-exon (light blue box) that has another GT donor further downstream. Then, competition between the two donor sites determines whether the cryptic/micro-exon will be included in the mature RNA or not. The fate of the mRNA carrying this extra short sequence might involve degradation.

differences, both models favor "noisy splicing," which is thought to drive mRNA isoform diversity in human cells (Pickrell et al., 2010).
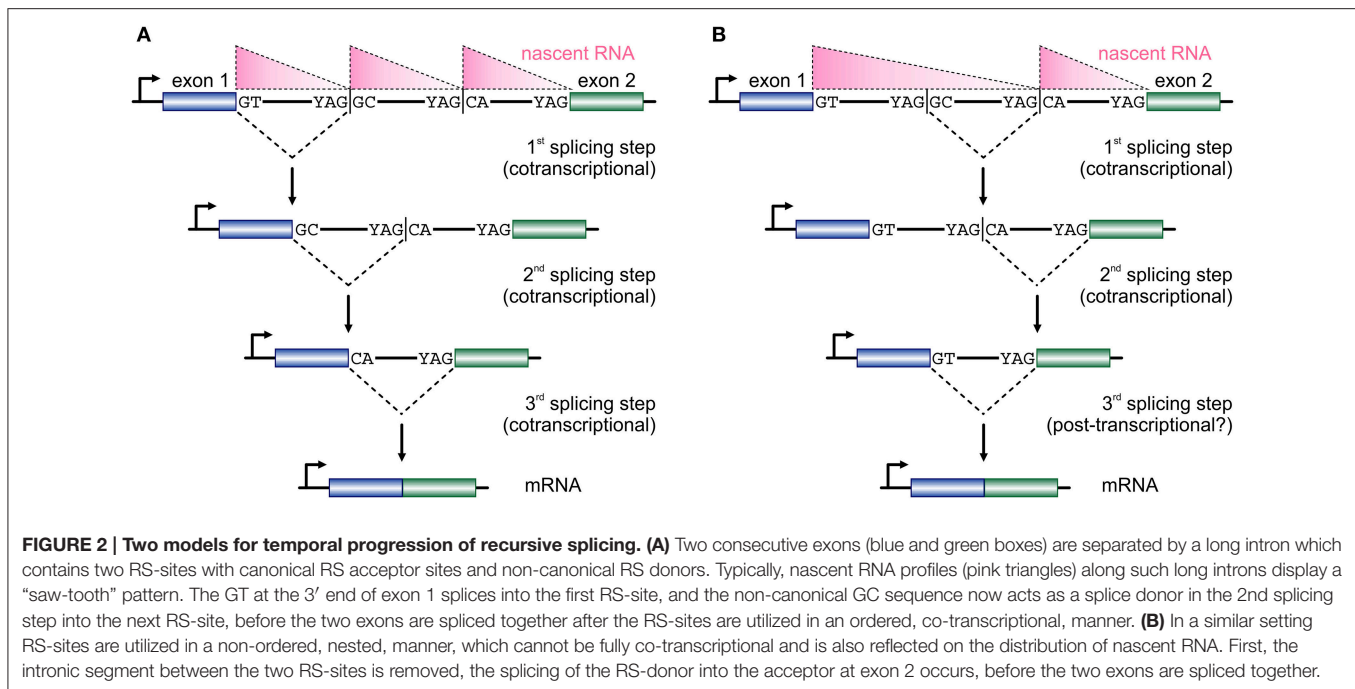
REGULATORY AND DISEASE IMPLICATIONS OF RECURSIVE SPLICING

The size of first introns in higher eukaryotes is such that, on average, exceeds all other downstream introns in length (Bradnam and Korf, 2008). This structural property of eukaryotic genomes has been linked with programmed delays in gene transcription cycles (Swinburne and Silver, 2008). As a result, the preferential positioning of RS-sites in such long introns (Kelly et al., 2015; Sibley et al., 2015) creates a novel regulatory layer for the processing of the nascent transcripts copied from these loci. Given that the majority of splicing in human cells occurs cotranscriptionally (Aitken et al., 2011; Tilgner et al., 2012), it would be reasonable to assume that the RS-junctions in one long intron are used successively at more or less the moment they are produced by the RNA polymerase (**Figure 2A**). This is supported by the study of TNF-inducible *SAMD4A*; upon induction, nascent RNA production progresses synchronously along its first intron and intronic RNA FISH fails to return evidence in favor of a single, long, transcript from this intron (Wada et al., 2009; Kelly et al., 2015). Intermediate splicing products at the 8 RS-sites in this 134-kbp intron appear and disappear in sync with the production of nascent RNA, and the half-life of each such RS-intermediate is ∼1/15 the time it takes the RNA polymerase to fully transcribe this intron (Kelly et al., 2015). This evidence, plus the "saw-tooth" patterns observed in brain RNA-seq data (Sibley et al., 2015; see **Figure 2**), are in support of the successive use of RS-sites. Nonetheless, there have been reports of non-ordered ("nested") use of such sites (Suzuki et al., 2013; Gazzoli et al., 2016), whereby the RS-sites can engage

in splicing reactions decoupled from cotranscriptionality and in which long primary transcripts survive degradation (**Figure 2B**). In fact, such decoupling of RS has been proposed for yeast splicing (Lopez and Séraphin, 2000).

Another question that arises is: Are the RS-sites in a given long intron all used in every transcription cycle or is their usage more stochastic? Again, studies from the *SAMD4A* locus using CRISPR-Cas9 technology (Ran et al., 2013) to specifically mutate 3 RS-sites, showed that abolishing any one RS-site results in a 35–50% reduction in mRNA levels (Kelly et al., 2015). Similarly, reducing RS-site usage by antisense oligonucleotides in the zebrafish *cadm2a* gene led to a ∼2-fold reduction in its mRNA levels *in vivo* (Sibley et al., 2015). These results (albeit based a limited number of example loci) point to a stochastic usage of multiple RS-sites along one intron and/or to compensatory mechanisms that prevent a complete loss of mRNA output. Additionally, it is necessary to investigate the connection between RS, exon skipping, and the formation of circular RNAs from a given gene locus, as they could all be functionally linked (Kelly et al., 2014).

RS-sites were found to be more conserved than equivalent intronic regions of similar composition in humans (Kelly et al., 2015; Sibley et al., 2015), and this hinted in favor of their functional role. As more than 90% of human genetic variation maps outside protein-coding regions, at inter- or intragenic sequences, and >40% maps within introns (Maurano et al., 2012), it is attractive to hypothesize that mutations at RS-sites may contribute to disease manifestation. Splicing defects are now well-established contributors in various diseases (Chabot and Shkreta, 2016), and RS, yet another layer of splicing regulation, remains unexplored. In fact, when we intersected a list of high-confidence RS-sites from human brain (Sibley et al., 2015) or endothelial cells (Kelly et al., 2015) to an ensemble of all putatively disease-causative human SNPs, they overlapped (within the 40 preceding the RS-junction) those associated with

FIGURE 2 | Two models for temporal progression of recursive splicing. (A) Two consecutive exons (blue and green boxes) are separated by a long intron which contains two RS-sites with canonical RS acceptor sites and non-canonical RS donors. Typically, nascent RNA profiles (pink triangles) along such long introns display a "saw-tooth" pattern. The GT at the 3′ end of exon 1 splices into the first RS-site, and the non-canonical GC sequence now acts as a splice donor in the 2nd splicing step into the next RS-site, before the two exons are spliced together after the RS-sites are utilized in an ordered, co-transcriptional, manner. **(B)** In a similar setting RS-sites are utilized in a non-ordered, nested, manner, which cannot be fully co-transcriptional and is also reflected on the distribution of nascent RNA. First, the intronic segment between the two RS-sites is removed, the splicing of the RS-donor into the acceptor at exon 2 occurs, before the two exons are spliced together.

neurological (e.g., Parkinson's disease, cognitive performance) or circulatory disorders/traits (e.g., retinal vascular caliper, blood pressure), respectively, more than what was expected by chance (A. Papantonis; unpublished data). Such a potential role of RS should be further investigated in both disease models and in GWAS datasets, as it can—in conjunction with alternative splicing—impact heavily on the mRNA isoform that a given cell generates.

CONCLUSIONS AND OUTLOOK

We think that there is still much to be discovered about the molecular basis and the regulatory implications of recursive splicing. The presence of non-canonical splicing sequences at RS-sites, the possibility of splice-site competition, the proposed involvement of U1 variants, even the cotranscriptional and/or non-sequential processing of long introns all need to be systematically dissected. To cite just a few pertinent questions: How widespread is recursive splicing across mammalian tissues and developmental stages? Is it affected once cell homeostasis is challenged, and how does this affect transcript maturation? How

are RS-sites defined, recognized, and marked epigenetically? Are they being utilized in a stochastic or a deterministic temporal order? Addressing these questions, amongst others, will be important for understanding this unforeseen regulatory layer of transcript processing in higher eukaryotes.

AUTHOR CONTRIBUTIONS

TG, KS, and AP reviewed the bibliography and wrote the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Aitken, S., Alexander, R. D., and Beggs, J. D. (2011). Modelling reveals kinetic advantages of co-transcriptional splicing. *PLoS Comput. Biol.* 7:e1002215. doi: 10.1371/journal.pcbi.1002215

Bradnam, K. R., and Korf, I. (2008). Longer first introns are a general property of eukaryotic gene structure. *PLoS ONE* 3:e3093. doi: 10.1371/journal.pone.0003093

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

Chabot, B., and Shkreta, L. (2016). Defective control of pre-messenger RNA splicing in human disease. *J. Cell Biol.* 212, 13–27. doi: 10.1083/jcb.201510032

Dewey, C. N., Rogozin, I. B., and Koonin, E. V. (2006). Compensatory relationship between splice sites and exonic splicing signals depending on the length of vertebrate introns. *BMC Genomics* 7:311. doi: 10.1186/1471-2164-7-311

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

Dye, M. J., Gromak, N., and Proudfoot, N. J. (2006). Exon tethering in transcription by RNA polymerase II. *Mol. Cell* 21, 849–859. doi: 10.1016/j.molcel.2006.01.032

Gazzoli, I., Pulyakhina, I., Verwey, N. E., Ariyurek, Y., Laros, J. F., 't Hoen, P. A., et al. (2016). Non-sequential and multi-step splicing of the dystrophin transcript. *RNA Biol.* 13, 290–305. doi: 10.1080/15476286.2015.1125074

Grellscheid, S. N., and Smith, C. W. (2006). An apparent pseudo-exon acts both as an alternative exon that leads to nonsense-mediated decay and as a zero-length exon. *Mol. Cell. Biol.* 26, 2237–2246. doi: 10.1128/MCB.26.6.2237-2246.2006

Hare, M. P., and Palumbi, S. R. (2003). High intron sequence conservation across three mammalian orders suggests functional constraints. *Mol. Biol. Evol.* 20, 969–978. doi: 10.1093/molbev/msg111

Hollander, D., Naftelberg, S., Lev-Maor, G., Kornblihtt, A. R., and Ast, G. (2016). How are short exons flanked by long introns defined and committed to splicing? *Trends Genet.* 32, 596–606. doi: 10.1016/j.tig.2016.07.003

Jeffares, D. C., Penkett, C. J., and Bähler, J. (2008). Rapidly regulated genes are intron poor. *Trends Genet.* 24, 375–378. doi: 10.1016/j.tig.2008.05.006

Kelly, S., Georgomanolis, T., Zirkel, A., Diermeier, S., O'Reilly, D., Murphy, S., et al. (2015). Splicing of many human genes involves sites embedded within introns. *Nucleic Acids Res.* 43, 4721–4732. doi: 10.1093/nar/gkv386

Kelly, S., Greenman, C., Cook, P. R., and Papantonis, A. (2014). Exon skipping is correlated with exon circularization. *J. Mol. Biol.* 427, 2414–2417. doi: 10.1016/j.jmb.2015.02.018

Kyriakopoulou, C., Larsson, P., Liu, L., Schuster, J., Söderbom, F., Kirsebom, L. A., et al. (2006). U1-like snRNAs lacking complementarity to canonical 5′ splice sites. *RNA* 12, 1603–1611. doi: 10.1261/rna.26506

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. *Curr. Biol.* 20, R754–R763. doi: 10.1016/j.cub.2010.06.070

Lopez, P. J., and Séraphin, B. (2000). Uncoupling yeast intron recognition from transcription with recursive splicing. *EMBO Rep.* 1, 334–339. doi: 10.1093/embo-reports/kvd065

Marquez, Y., Höpfler, M., Ayatollahi, Z., Barta, A., and Kalyna, M. (2015). Unmasking alternative splicing inside protein-coding exons defines exitrons and their role in proteome plasticity. *Genome Res.* 25, 995–1007. doi: 10.1101/gr.186585.114

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190–1195. doi: 10.1126/science.1222794

O'Reilly, D., Dienstbier, M., Cowley, S. A., Vazquez, P., Drozdz, M., Taylor, S., et al. (2013). Differentially expressed, variant U1 snRNAs regulate gene expression in human cells. *Genome Res.* 23, 281–291. doi: 10.1101/gr.142968.112

Parra, M. K., Tan, J. S., Mohandas, N., and Conboy, J. G. (2008). Intrasplicing coordinates alternative first exons with alternative splicing in the protein 4.1R gene. *EMBO J.* 27, 122–131. doi: 10.1038/sj.emboj.7601957

Pickrell, J. K., Pai, A. A., Gilad, Y., and Pritchard, J. K. (2010). Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet.* 6:e1001236. doi: 10.1371/journal.pgen.1001236

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308. doi: 10.1007/978-1-4939-1862-1_10

Raponi, M., and Baralle, D. (2008). Can donor splice site recognition occur without the involvement of U1 snRNP? *Biochem. Soc. Trans.* 36, 548–550. doi: 10.1042/BST0360548

Roca, X., Akerman, M., Gaus, H., Berdeja, A., Bennett, C. F., and Krainer, A. R. (2012). Widespread recognition of 5′ splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes Dev.* 26, 1098–1109. doi: 10.1101/gad.190173.112

Rogozin, I. B., Carmel, L., Csuros, M., and Koonin, E. V. (2012). Origin and evolution of spliceosomal introns. *Biol. Direct.* 7:11. doi: 10.1186/1745-6150-7-11

Scheckel, C., and Darnell, R. B. (2015). Microexons—tiny but mighty. *EMBO J.* 34, 273–274. doi: 10.15252/embj.201490651

Shepard, S., McCreary, M., and Fedorov, A. (2009). The peculiarities of large intron splicing in animals. *PLoS ONE* 4:e7853. doi: 10.1371/journal.pone.0007853

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

Somarelli, J. A., Mesa, A., Rodriguez, C. E., Sharma, S., and Herrera, R. J. (2014). U1 small nuclear RNA variants differentially form ribonucleoprotein particles *in vitro*. *Gene* 540, 11–15. doi: 10.1016/j.gene.2014.02.054

Suzuki, H., Kameyama, T., Ohe, K., Tsukahara, T., and Mayeda, A. (2013). Nested introns in an intron: evidence of multi-step splicing in a large intron of the human dystrophin pre-mRNA. *FEBS Lett.* 587, 555–561. doi: 10.1016/j.febslet.2013.01.057

Swinburne, I. A., and Silver, P. A. (2008). Intron delays and transcriptional timing during development. *Dev. Cell* 14, 324–330. doi: 10.1016/j.devcel.2008.02.002

Tan, J., Ho, J. X., Zhong, Z., Luo, S., Chen, G., and Roca, X. (2016). Noncanonical registers and base pairs in human 5′ splice-site selection. *Nucleic Acids Res.* 44, 3908–3921. doi: 10.1093/nar/gkw163

Thanaraj, T. A., and Clark, F. (2001). Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. *Nucleic Acids Res.* 29, 2581–2593. doi: 10.1093/nar/29.12.2581

Tilgner, H., Knowles, D. G., Johnson, R., Davis, C. A., Chakrabortty, S., Djebali, S., et al. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* 22, 1616–1625. doi: 10.1101/gr.134445.111

Turunen, J. J., Niemelä, E. H., Verma, B., and Frilander, M. J. (2013). The significant other: splicing by the minor spliceosome. *Wiley Interdiscipl. Rev. RNA* 4, 61–76. doi: 10.1002/wrna.1141

Twigg, S. R., Burns, H. D., Oldridge, M., Heath, J. K., and Wilkie, A. O. (1998). Conserved use of a non-canonical 5′ splice site (/GA) in alternative splicing by fibroblast growth factor receptors 1, 2 and 3. *Hum. Mol. Genet.* 7, 685–691.

Vazquez-Arango, P., Vowles, J., Browne, C., Hartfield, E., Fernandes, H. J., Mandefro, B., et al. (2016). Variant U1 snRNAs are implicated in human pluripotent stem cell maintenance and neuromuscular disease. *Nucleic Acids Res.* doi: 10.1093/nar/gkw711. [Epub ahead of print].

Wada, Y., Ohta, Y., Xu, M., Tsutsumi, S., Minami, T., Inoue, K., et al. (2009). A wave of nascent transcription on activated human genes. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18357–18361. doi: 10.1073/pnas.0902573106

Wang, Z., and Burge, C. B. (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* 14, 802–813. doi: 10.1261/rna.876308

Zhang, C., Hastings, M. L., Krainer, A. R., and Zhang, M. Q. (2007). Dual-specificity splice sites function alternatively as 5′ and 3′ splice sites. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15028–15033. doi: 10.1073/pnas.0703773104