Recent advances of studies on alternative intron retention

Lei-Lei Zhan

Institute of Biochemistry, College of Life Sciences, Zhejiang University (Zijingang Campus), Hangzhou, Zhejiang, China

Abstract

Alternative splicing, allowing multiple mRNAs to be generated from a single gene, is a major source of proteome diversity in eukaryotes. Intron retention (IR), one type of alternative splicing, is the complete retention of an intron in a mature transcript. IR is believed to be associated with failure of the recognition of weak splice sites flanking introns. Mutations in DNA sequences, including point mutations and sequence insertions or deletions, can be at the origin and evolution of IR. The strength of weaker splice sites is the main cause of IR, but some *cis*-regulatory elements and *trans*-regulatory factors can also play crucial roles in regulating IR. IR can result in the regulation of gene expression and may contribute to increase protein diversity. IR has been shown to occur in a variety of diseases, and it frequently leads to aberrant splicing.

Introduction

Alternative splicing (AS), allowing multiple mRNAs to be generated from a single gene, is considered to be one of the main mechanisms for regulating gene expression and augmenting protein diversity.[1] For example, recent studies indicated that more than 90% of human genes are alternatively spliced.[2,3] Moreover, while most genes are thought to create only two alternatively spliced mRNA isoforms, many genes have the capacity to encode a much greater repertoire of mRNA variants.[1,4] The most dramatic example is the *Dscam* gene of *Drosophila melanogaster*, which can potentially generate a large number of mRNA isoforms via AS.[5]

There are several different types of AS events (Figure 1).[4,6] Intron retention (IR), one of the least common types of AS, is the complete retention of an intron in a mature transcript. Generally, IR is considered to be the result of intron, rather than exon, definition associated with failure of the recognition of weak splice sites flanking short introns. IR is the least characterized event of all AS types, mainly because of the exclusion of this phenomenon in many studies, due to the difficulties to differentiate it from incompletely processed transcripts.[7] However, many discoveries have been made about IR in recent years, including studies focusing human diseases. Here, I review the most recent studies about IR, and my aim is to provide some useful information for future research in this field.

Models for mechanisms of intron retention

Splicing is ubiquitous in eukaryotes, which removes introns from pre-mRNA and joins exons together to form a mature mRNA. Exon and intron recognition is achieved by two mechanisms: exon definition (ED) and intron definition (ID). ED defines pairs of splice sites that flank the same exons, and ID defines pairs of splice sites located on both ends of the same introns (Figure 2).[6,8] ED can identify relatively short exon sequences (~400 nucleotides) located within large intronic sequences, which is the case for most of the exons in higher eukaryotic cells.[8,9] ID seems to be the ancient mechanism that allows the recognition of introns embedded in large exonic sequences, which is the case for most of the introns in lower eukaryotic cells.[9]

Intron retention (IR) is thought to be associated with failure of ED and ID. There are two proposed models.[10,11] In the first model (Model A) (Figure 3A), it suggests that both ID and ED should be associated with IR. For the relatively short introns in *exon* + *intron* + *exon* units, Mode1 A is plausible, and supported by some evidence.[10,12] Model B (Figure 3B), suggesting only the ED should be associated with IR and failure in E to A or A to B complex transition would lead to IR during the splicing process.[11] For the relatively long retained introns, Model B is more reasonable. In fact, the two models do not contradict each other.

Prevalence of intron retention

Although there are introns in the genomes of most eukaryotes, AS is only prevalent in multicellular organisms, and most of the studies were focused on higher multicellular organisms. In fungi,[13-17] as well as in the protozoa and in lower metazoans,[18-20] the studies indicated that AS are extremely rare. However, in these organisms, the most prevalent type of AS was found to be IR.[20,21]

Among higher multicellular organisms, plants show relatively low levels of alternatively spliced genes but high levels of gene families,[20-22] and this phenomenon could be explained by the negative correlation between AS and gene duplication.[23] However, the plants show high levels of IR events in AS. For example, the studies revealed that IR is a major phenomenon in AS in *Arabidopsis thaliana*, and responsible for about 30% of AS events.[24-26] In the higher metazoans, IR is only a minor form of AS. It has been shown that IR is the rarest forms of AS events in seven analyzed species,[26] including nematode (*Caenorhabditis elegans*), fly (*D. melanogaster*), sea squirt (*Ciona intestinalis*), chicken (*Gallus gallus*), mouse (*Mus musculus*), rat (*Rattus norvegicus*) and human (*Homo sapiens*).

Origin and evolution of intron retention

There are at least two possible origin and evolution models of AS. The first model emphasizes the natural mutations of DNA sequences, whereas the second emphasizes the evolution of splicing regulatory factors. It is of great importance to realize that the two models do not necessarily contradict each other, however the splicing regulatory factor model has not received much experimental attention and

Correspondence: Lei-Lei Zhan, Institute of Biochemistry, College of Life Sciences, Zhejiang University (Zijingang Campus), Hangzhou, Zhejiang, ZJ310058, China.
Tel. +86.571.88206479 - Fax: +86.571.88206478
E-mail: zhanleiray@sina.com

Key words: alternative splicing, intron retention, mechanism, origin and evolution.

Acknowledgements: I would thank Dr Can Liu from Beihang University for numerous insightful discussions. I am grateful to Dr Yi-fang Zhang from East China Normal University for critical reading of the manuscript.

Conflict of interests: the author declares no potential conflict of interests.

Funding: LLZ received funding from a General Financial Grant, number 2012M511357 from the China Postdoctoral Science Foundation.

Received for publication: 14 November 2012.
Revision received: 30 December 2012.
Accepted for publication: 23 January 2013

Trends in Evolutionary Biology 2013; 5:e1
doi:10.4081/eb.2013.e1

remains a possibility only.[6]

Some evidence supporting the first model suggest that there should be at least three possible ways leading to the origin and evolution of IR. Firstly, mutations in DNA sequences, including point mutations and sequence deletions may enable IR (Figure 4A). Previous studies showed that experimental mutations of optimal splice sites lead to IR.[9,10,12,27] In *D. melanogaster*, an intronic partial deletion also induced alternative IR at the *Rieske Iron Sulphur Protein* (*RFeSP*) *locus*.[28] Secondly, natural point mutations could create weak splice sites on the exonic sequences, thereby promoting IR (Figure 4B). For example, in *C. elegans*, there are 16 cases of retained introns, which were created by *intronization* of exonic sequences.[29] Some studies revealed that some mobile elements of other DNA fragments can be integrated into a proto-splice site location of thus creating introns,[30-32] and it is possible that some of these new introns may be retained (Figure 4C).

Regulations on intron retention

Strength of splice site

Generally, IR is associated with weaker splice sites, and the strength of splice sites often plays very important roles in regulation of IR. Firstly, previous bioinformatic analyses indicated that IR correlate with weaker splice sites.[33,34] In human, it was found that there is a negative correlation between the strength of splice sites and the frequency of IR.[11] Secondly, some experimental observations demonstrated that changes of splice site strength have great effect on IR. Studies with human genes have shown that strengthening weak splice sites flanking retained introns could cause an increase in their removal levels,[35] or completely abolish retention.[36] Other studies also revealed that weakening the strength of optimal splice sites may lead to IR or cause an increase in their retention levels.[9,10,12,27]

Cis-regulator and trans-factors

Cis-regulatory elements, including exonic splicing enhancer (ESE), exonic splicing silencer (ESS), intronic splicing enhancer (ISE) and intronic splicing silencer (ISS), and *trans*-factors also play important roles in regulating IR. Bioinformatic analyses, have shown that sets of *cis*-regulatory elements are associated with regulation of IR of many genes by acting as binding sites for *trans*-factors which promote (ESEs and ISEs) or inhibit (ESSs and ISSs) splicing, and the distribution and density of these *cis*-regulators plays an important role in regulation of IR[11]. In bovine growth hor-

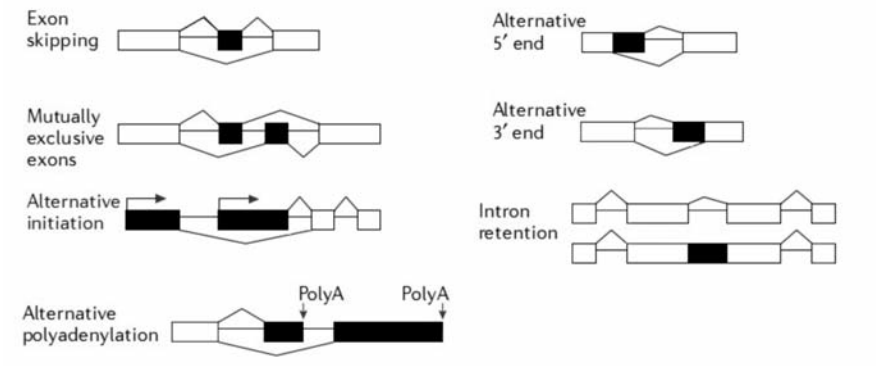

Figure 1. Types of alternative splicing. Black boxes represent alternatively spliced exons, and white boxes represent constitutive exons. The lines between the boxes represent introns.

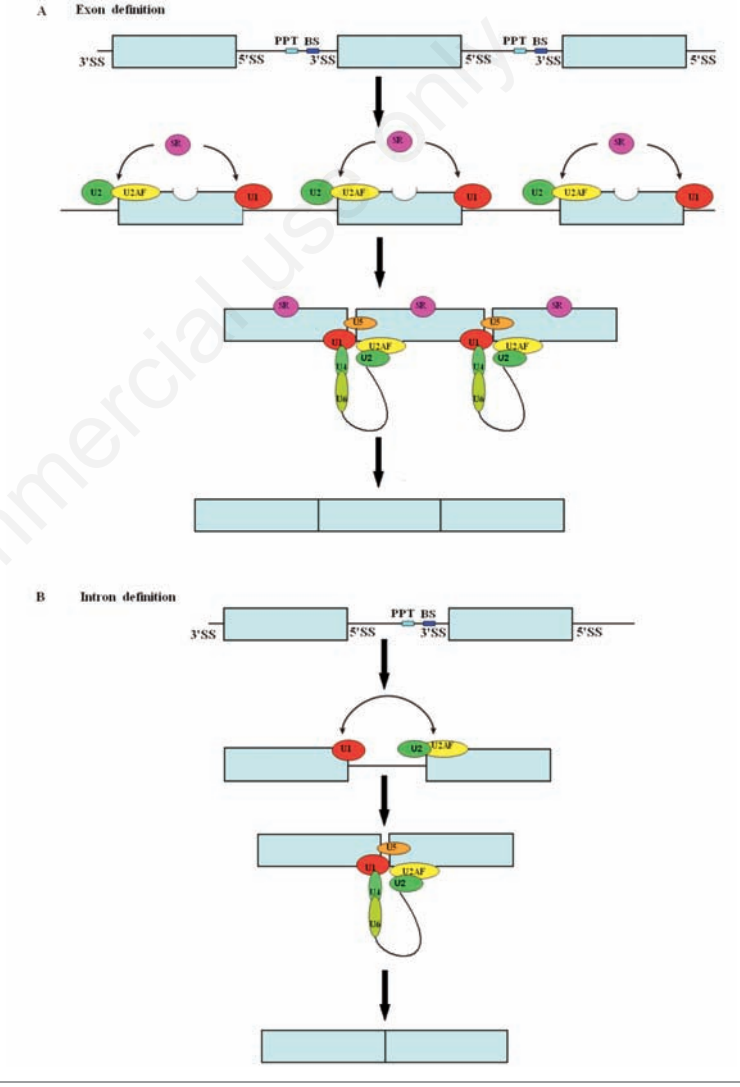

Figure 2. Mechanisms of exon and intron definition in the splicing process of mRNAs.[6,8] (A) In exon definition (ED), SR proteins bind to exonic splicing enhancers (ESE), recruiting U1 to the downstream 5' splice site (SS) and the splicing factor U2AF to the upstream polypyrimidine tract (PPT) and the 3'SS. U2AF then recruits U2 to the branch site (BS) which is a splicing signal located upstream of 3' end of the intron. So, the ED can form a *cross-exon* recognition complex by placing the basal splicing machinery in the splice sites that flank the same exon. (B) In intron definition (ID), U1 binds to the upstream 5'SS and U2AF and U2 bind to the downstream PPT and BS of the same intron, respectively. Therefore, ID selects pairs of splice sites located on both ends of the same intron.

mone (bGH) pre-mRNAs, mutagenesis experiments suggested that a purine-rich ESE is crucial to counterbalance retention of intron D.[36] In the *FosB* human gene, the polypyrimidine tract binding protein (PTB) can regulate the retention of intron 4.[37] It was also shown that the *trans*-factor Ptbp1 play a crucial role in regulating the retention of 3'-terminal introns of many genes encoding important proteins which are tightly regulated during neuronal differentiation.[38] In addition, a recent study indicated that methylation and transcriptional efficiency have influence on AS patterns,[39] and it is possible to find that methylation and transcriptional efficiency is associated with alternative IR.

Consequences of intron retention

Intron retention regulates gene expression

During the expression of genes, the phenomenon of IR is very common. A number of studies indicated that up to 15% of human genes present at least one IR events, and that at least 22% of all informative IR events are also present in the mouse transcriptome.[40] IR can insert premature stop codons (PTC) in some mature transcripts that then be degraded by non-sense mediated decay (NMD). NMD is one of several RNA surveillance pathways that ensure the fidelity of gene expression. So, decreasing the level of mRNAs of given genes is a function of IR. This phenomenon should not be seen as a cellular waste because unusable transcripts are produced, but rather as a means for the cell to regulate gene expression at the post-transcriptional level in a precise spatiotemporal manner.

Interestingly, IR could also regulate the expression of genes by pathways other than NMD. For example, it was revealed that intron-retained transcripts of *LY6G5B* and *LY6G6D* are not subjected to NMD, but form chimeric transcripts with adjacent genes' transcripts respectively, and regulate the expression of two genes.[7]

Intron retention and protein diversity

Often, IR affects mRNA transport to the cytoplasm and can insert PTCs in the mature transcripts. However, the NMD is not the fate of all mRNAs with retained introns. Some isoforms of genes with the retained introns could be transported into cytoplasm because of some special elements. For example, the sequence from *Tap* intron 10 could function as a constitutive transport element (CTE) to allow export and expression of a mRNA with a retained intron.[41] Some isoforms with retained introns do not cause open-reading-frame (ORF) shifts, so they could be translated into functional proteins with biological activities. IR events of the *H3R* (histamine H receptor) gene do not create PTCs, but make several kinds of functional proteins in mice and rat.[42] In addition, the isoforms with retained introns that lead to PTCs may also be translated into functional proteins. In *KCMNA1* gene of human, the retention of intron 16 or intron 17a leads to PTCs in some mRNAs, but these mRNAs would generate significantly truncated BKCa channel proteins.[43] In *D. melanogaster*, alternative IR at *RFeSP locus* created a PTC, but the isoforms with the PTCs could be translated into functional proteins.[28]

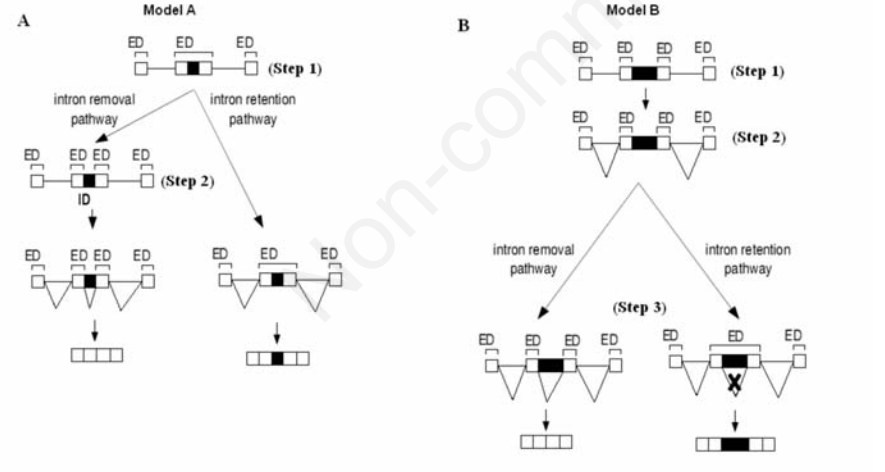

Figure 3. Models for mechanisms of intron retention (IR). (A) In Model A,[10,11] both intron definition (ID) and exon definition (ED) are associated with IR. In this mode, there are two steps in splice site recognition for short introns in short *exon* + *intron* + *exon* units (~400 bp, close to the size limit for exon recognition). In step 1, all the exons are defined, including the exon + intron + exon unit. Step 2 is alternative, if ID occurs in the *exon* + *intron* + *exon* unit, the intron will be removed. Abrogation of this step will lead to IR. (B) In Model B,[11] only ED is involved with IR. In this mode, three steps are associated with splice site recognition for long introns in *exon* + *intron* + *exon* units. In step 1, all exons are defined, including those flanking the intron to be retained. In step 2, flanking introns are spliced. Step 3 is alternative, if the exons flanking the intron to be retained are joined, the intron is removed; otherwise, skipping of this step by hypothetical failure in E to A or A to B complex transition would lead to IR. Black boxes represent alternatively spliced exons, and white boxes represent constitutive exons. The lines between the boxes represent introns.

Intron retention and disease

AS is a major source of proteome diversity in humans and thus is highly relevant to diseases.[44] IR has been shown to occur in a variety of diseases, and it frequently leads to erroneous splicing. Some studies have illustrated how IR event can play a relevant role in several diseases. Aberrant IR was shown to cause the dysfunction of *ATRX*, leading to acquired alpha thalassemia.[45] A new mutation of *XPC* gene causing the abnormal IR of intron 12, has led to the highest worldwide prevalence of xeroderma pigmentosum in black Mahori patients.[46] A point mutation of *POMGnT1* gene leading to aberrant IR of intron 21, is associated with muscle-eye-brain (MEB) disease.[47] Interestingly, it was found that normal IR can inhibit the occurrence of illness. For example, it was revealed that retention of intron 1 of *Id3* gene would generate a novel Id3 isoform following vascular injury and inhibit lesion formation.[48]

Conclusions

IR is more common in lower metazoans, fungi, protozoas and plants, but was found to be the rarest AS event in vertebrates and invertebrates. Generally, mutations in DNA sequences can give rise to origin of IR. The strength of splice sites, *cis*-regulatory elements and *trans*-regulatory factors play crucial roles in regulating IR. IR has the potential to regulate the expression of genes and increase protein diversity. IR has been shown to play a role in various diseases, and often lead to aber-

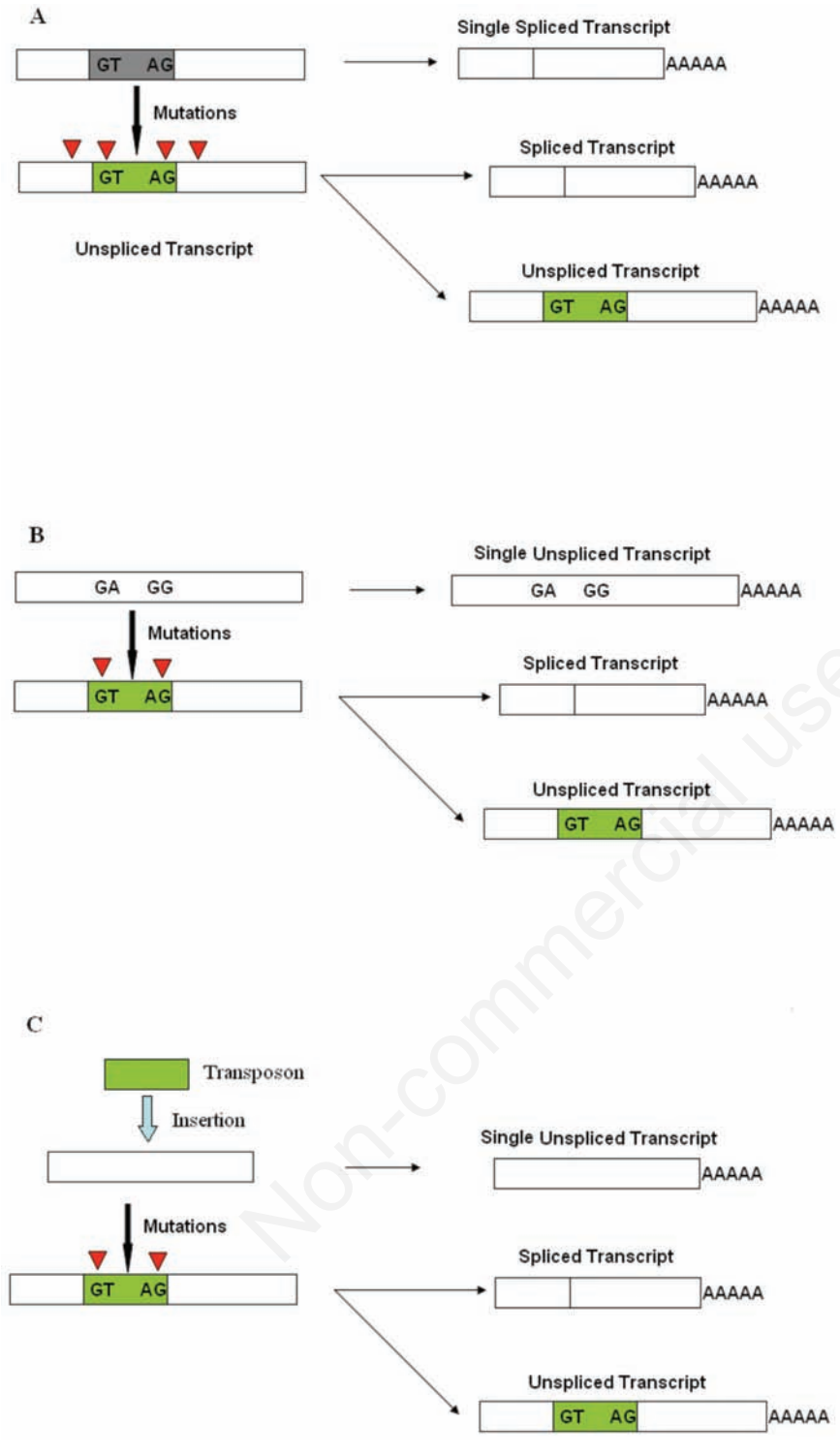

Figure 4. The origin and evolution of intron retention. (A) Mutations of DNA sequence, including point mutations and sequence deletions or insertions, have the possibilities to weaken the strength of the splice sites and change the conditions of splicing cis-regulators (ESE, ISE, ESS and ISS), and would make the normal introns becoming introns to be retained. (B) The retained introns might originate from exonic sequences. Point mutations on the exonic sequences have the possibilities to create weaker splice sites and introns to be retained. (C) The mobile elements of other DNA fragments can be integrated into a proto-splice site location of thus creating introns, and it is possible that some of these new introns may be retained. The inverted red triangles represent the mutations of DNA sequence. Grey box represents the normal intron, while green boxes represent the retained introns. White boxes represent constitutive exons.

rant splicing.

References

1. Black DL. Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem 2003;72:291-336.
2. Wang ET, Sandberg R, Luo S, et al. Alternative isoform regulation in human tissue transcriptomes. Nature 2008;456:470-6.
3. Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. Cell 2009;136:701-18.
4. Matlin AJ, Clark F, Smith CW. Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol 2005;6:386-98.
5. Schmucker D, Clemens JC, Shu H, et al. Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell 2000;101:671-84.
6. Ast G. How did alternative splicing evolve? Nat Rev Genet 2004;5:773-82.
7. Calvanese V, Mallya M, Campbell RD, Aguado B. Regulation of expression of two LY-6 family genes by intron retention and transcription induced chimerism. BMC Mol Biol 2008;9:81.
8. Berget SM. Exon recognition in vertebrate splicing. J Biol Chem 1995;270:2411-4.
9. Romfo CM, Alvarez CJ, van Heeckeren WJ, et al. Evidence for splice site pairing via intron definition in Schizosaccharomyces pombe. Mol Cell Biol 2000;20:7955-70.
10. Sterner DA, Berget SM. In vivo recognition of a vertebrate mini-exon as an exon-intron-exon unit. Mol Cell Biol 1993;13:2677-87.
11. Sakabe NJ, de Souza SJ. Sequence features responsible for intron retention in human. BMC Genomics 2007;8:59.
12. Talerico M, Berget SM. Intron definition in splicing of small Drosophila introns. Mol Cell Biol 1994;14:3434-45.
13. Yatzkan E, Yarden O. The B regulatory subunit of protein phosphatase 2A is required for completion of macroconidiation and other developmental processes in Neurospora crassa. Mol Microbiol 1999;31:197-209.
14. Davis CA, Grate L, Spingola M, Ares M Jr. Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. Nucleic Acids Res 2000;28:1700-6.
15. Okazaki K, Niwa O. mRNAs encoding zinc finger protein isoforms are expressed by alternative splicing of an in-frame intron in fission yeast. DNA Res 2000;7:27-30.
16. Vilardell J, Chartrand P, Singer RH, Warner JR. The odyssey of a regulated transcript. RNA 2000;6:1773-80.

17. Ebbole DJ, Jin Y, Thon M, et al. Gene discovery and gene expression in the rice blast fungus, Magnaporthe grisea: analysis of expressed sequence tags. Mol Plant Microbe Interact 2004;17:1337-47.
18. Escalante R, Moreno N, Sastre L. Dictyostelium discoideum developmentally regulated genes whose expression is dependent on MADS box transcription factor SrfA. Eukaryot Cell 2003; 2:1327-35.
19. Muhia DK. Multiple splice variants encode a novel adenylyl cyclase of possible plastid origin expressed in the sexual stage of the malaria parasite Plasmodium falciparum. J Biol Chem 2003; 278:22014-22.
20. Kim E, Goren A, Ast G. Alternative splicing: current perspectives. Bioessays 2008;30: 38-47.
21. Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet 2010;11:345-55.
22. Guo YL. Gene family evolution in green plants with emphasis on the origination and evolution of Arabidopsis thaliana genes. Plant J 2012 Dec 6. [Epub ahead of print]
23. Kopelman NM, Lancet D, Yanai I. Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. Nat Genet 2005; 37:588-9.
24. Iida K, Seki M, Sakurai T, et al. Genome-wide analysis of alternative pre-mRNA splicing in Arabidopsis thaliana based on full-length cDNA sequences. Nucleic Acids Res 2004;32:5096-103.
25. Ner-Gaon H, Halachmi R, Savaldi-Goldstein S, et al. Intron retention is a major phenomenon in alternative splicing in Arabidopsis. Plant J 2004;39:877-85.
26. Kim E, Magen A, Ast G. Different levels of alternative splicing among eukaryotes. Nucleic Acids Res 2007; 35:125-31.
27. McCullough AJ, Berget SM. G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. Mol Cell Biol 1997;17:4562-71.
28. Gontijo AM, Miguela V, Whiting MF, et al. Intron retention in the Drosophila melanogaster Rieske Iron Sulphur Protein gene generated a new protein. Nat Commun 2011;2:323.
29. Irimia M, Rukov JL, Penny D, et al. Origin of introns by 'intronization' of exonic sequences. Trends Genet 2008;24:378-81.
30. Iwamoto M, Maekawa M, Saito A, et al. Evolutionary relationship of plant catalase genes inferred from intron-exon structures: isozyme divergence after the separation of monocots and dicots. Theor Appl Genet 1998;97:9-19.
31. Iwamoto M, Nagashima H, Nagamine T, et al. p-SINE1-like intron of the CatA catalase homologs and phylogenetic relationships among AA-genome Oryza and related species. Theor Appl Genet 1999;98:853-61.
32. Roy SW, Gilbert W. The evolution of spliceosomal introns: pattern, puzzles and progress. Nat Rev Genet 2006;7:211-21.
33. Stamm S, Zhu J, Nakai K, et al. An alternative-exon database and its statistical analysis. DNA Cell Biol 2000;19:739-56.
34. Zheng CL, Fu XD, Gribskov M. Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. RNA 2005;11:1777-87.
35. Romano M, Marcucci R, Baralle FE. Splicing of constitutive upstream introns is essential for the recognition of intra-exonic suboptimal splice sites in the thrombopoietin gene. Nucleic Acids Res 2001;29:886-94.
36. Dirksen WP, Sun Q, Rottman FM. Multiple splicing signals control alternative intron retention of bovine growth hormone pre-mRNA. J Biol Chem 1995;270:5346-52.
37. Marinescu V, Loomis PA, Ehmann S, et al. Regulation of retention of FosB intron 4 by PTB. PLoS ONE 2007;2:e828.
38. Yap K, Lim ZQ, Khandelia P, et al. Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. Genes Dev 2012;26:1209-23.
39. Zhou HL, Hinman MN, Barron VA, et al. Hu proteins regulate alternative splicing by inducing localized histone hyperacetylation in an RNA-dependent manner. Proc Natl Acad Sci USA 2011;108:627-35.
40. Galante PA, Sakabe NJ, Kirschbaum-Slager N, de Souza SJ. Detection and evaluation of intron retention events in the human transcriptome. RNA 2004;10:757-65.
41. Li Y, Bor YC, Misawa Y, et al. An intron with a constitutive transport element is retained in a Tap messenger RNA. Nature 2006;443:234-7.
42. Ding W, Lin L, Xiao Z, et al. Multiple sequence elements determine the intron retention in histamine H3 receptors in rats and mice. Int J Biochem Cell Biol 2009;41:2281-86.
43. Bell TJ, Miyashiro KY, Sul JY, et al. Intron retention facilitates splice variant diversity in calcium-activated big potassium channel populations. Proc Natl Acad Sci USA 2010;107:21152-7.
44. Garcia-Blanco MA, Baraniak AP, Lasda EL. Alternative splicing in disease and therapy. Nat Biotechnol 2004;22:535-46.
45. Nelson ME, Thurmes PJ, Hoyer JD, Steensma DP. A novel 5' ATRX mutation with splicing consequences in acquired alpha thalassemia-myelodysplastic syndrome. Haematologica 2005;90:1463-70.
46. Cartault F, Nava C, Malbrunot AC, et al. A new XPC gene splicing mutation has lead to the highest worldwide prevalence of xeroderma pigmentosum in black Mahori patients. DNA Repair 2011;10:577-85.
47. Saredi S, Ardissone A, Ruggieri A, et al. Novel POMGNT1 point mutations and intragenic rearrangements associated with muscle-eye-brain disease. J Neurol Sci 2012;318:45-50.
48. Forrest ST, Barringhaus KG, Perlegas D, et al. Intron retention generates a novel Id3 isoform that inhibits vascular lesion formation. J Biol Chem 2004;279:32897-903.