

Sequence analysis

Complementary intron sequence motifs associated with human exon repetition: a role for intragenic, inter-transcript interactions in gene expression

Richard J. Dixon^{1,*}, Ian C. Eperon² and Nilesh J. Samani¹¹Department of Cardiovascular Sciences and ²Department of Biochemistry, University of Leicester, Leicester, LE3 9Q, UK

Received on September 5, 2006; revised on October 13, 2006; accepted on November 8, 2006

Advance Access publication November 14, 2006

Associate Editor: Martin Bishop

ABSTRACT

Motivation: Exon repetition describes the presence of tandemly repeated exons in mRNA in the absence of duplications in the genome. The regulation of this process is not fully understood. We therefore investigated the entire flanking intronic sequences of exons involved in exon repetition for common sequence elements.

Results: A computational analysis of 48 human single exon repetition events identified two common sequence motifs. One of these motifs is pyrimidine-rich and is more common in the upstream intron, whilst the other motif is highly enriched in purines and is more common in the downstream intron. As the two motifs are complementary to each other, they support a model by which exon repetition occurs as a result of *trans*-splicing between separate pre-mRNA transcripts from the same gene that are brought together during transcription by complementary intronic sequences. The majority of the motif instances overlap with the locations of mobile elements such as Alu elements. We explore the potential importance of complementary intron sequences in a rat gene that undertakes natural exon repetition in a strain specific manner. The possibility that distant complementary sequences can stimulate inter-transcript splicing during transcription suggests an unsuspected new role for potential secondary structures in endogenous genes.

Availability:**Contact:** rd67@le.ac.uk**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

The majority of human genes are now known to be involved in mRNA alternative splicing suggesting that alternative splicing is one of the most significant processes contributing to the functional complexity of the human genome (Johnson *et al.*, 2003; Maniatis and Tasic, 2002; Modrek and Lee, 2002). Most alternative splicing research to date has focussed on alternative *cis*-splicing, in which exons located within an individual pre-mRNA are differentially joined to generate mature mRNAs (Black, 2003). This involves joining the exons to be included in the mature transcript in a linear contiguous manner, 5'–3', in the order they are found in the genome. However, the discovery of tandemly repeated exons

in mammalian mRNA in the absence of duplications in the genome, a phenomenon termed exon repetition, (or non-linear mRNA) suggested that there is an additional level of alternative splicing complexity in mammals (Caudevilla *et al.*, 1998; Frantz *et al.*, 1999). Exon repetition is allele-specific and operates strictly in *cis*, meaning that in heterozygotes only mRNA from the susceptible allele contains the repetition (Rigatti *et al.*, 2004). Until recently <10 genes have been discovered, mostly through serendipity, to exhibit exon repetition in mammals. However, our recent genome wide survey of candidate exon repetition events in expressed sequences from multiple species has yielded evidence for the occurrence of this phenomenon in at least 1% of mammalian genes suggesting that this process could contribute to phenotypic variation (Dixon *et al.*, 2005).

Despite the thousands of mRNA alternative splicing events identified to date, the regulation of alternative splicing is still not fully understood (Maniatis and Tasic, 2002). Regulatory sequence elements that direct alternative splicing have been characterized but it is still unclear how the majority of alternative splicing events are regulated (Ladd and Cooper, 2002; Matlin *et al.*, 2005). Superimposed upon this complexity of splicing regulatory elements is the role of the promoter architecture and the rate of RNA transcription procession (Proudfoot *et al.*, 2002). One aspect of uncertain significance is the influence of pre-mRNA secondary structures on splicing activity (Buratti and Baralle, 2004; Eperon *et al.*, 1988).

To explore the possibility that exon repetition is regulated by specific sequences in the exons and their flanking introns, we have analysed a set of 48 human single exon repetition events. These are events that comprise repetition of a single exon sequence (e.g. exon 2-2) within an expressed sequence from GenBank. Computational analysis of the entire exons and intronic sequences flanking these events was undertaken with algorithms that identify common motifs in unaligned sequences. This approach has revealed two motifs, a pyrimidine rich motif that is more predominant in the upstream introns and a purine rich motif that is more predominant in the downstream introns. Interestingly, these sequence motifs are complementary suggesting that base pairing interactions between these motifs could contribute to an RNA secondary structure that is conducive to the phenomenon of exon repetition taking place. This hypothesis is investigated by sequence analysis of the rat *Sa* gene in two different strains, one that exhibits exon

*To whom correspondence should be addressed.

repetition and one that does not (Frantz *et al.*, 1999; Rigatti *et al.*, 2004).

2 METHODS

2.1 Detection of candidate exon repetition events in expressed sequences

The dataset of single exon repetition events used in this study were derived from a computational genome wide survey as implemented and described in detail previously (Dixon *et al.*, 2005). The current survey in human was implemented exactly as mentioned previously (Dixon *et al.*, 2005) except for the use of more recent EST and mRNA sequence archives. Human mRNA sequences were downloaded from the UCSC genome browser (UCSC hg17, May 2004, NCBI 35 assembly), (<http://hgdownload.cse.ucsc.edu/goldenPath/hg17/bigZips/>). Human EST sequences were downloaded from NCBI (February 2006), (<ftp://ftp.ncbi.nih.gov/blast/db/FASTA/>).

2.2 Intron and exon sequence retrieval

The intron and exon sequences from the 48 human single exon repetition events that were used in this study were obtained from Ensembl (v29.35b, NCBI 35 assembly), (<http://www.ensembl.org/>). The flanking intron sequences were obtained by using a series of Perl scripts incorporating the BioPerl module 'intron.pm' (<http://doc.bioperl.org/releases/bioperl-current/bioperl-live/Bio/Seq/Feature/Gene/Intron.html>) that utilized the Ensembl Perl API. As a background intron sequence data set we used the above Perl programs to obtain the entire upstream and downstream introns of a random set of 1038 exons selected from a list of all Ensembl human exons, which had been filtered to remove those genes for which we have evidence for involvement in exon repetition. Random sampling was undertaken using a Perl program with a random number generator.

2.3 Identification of common motifs and sequence analysis

Two separate programs that search for common sequence motifs in unaligned sequences were used, MEME (<http://meme.sdsc.edu/meme/intro.html>) (Bailey and Elkan, 1994) and the Gibbs motif sampler (<http://bayesweb.wadsworth.org/gibbs/gibbs.html>) (Lawrence *et al.*, 1993). These programs were run on the exon sequences, the entire upstream and downstream intron sequences on a local Linux machine, with the default parameters when the length of the searched motif was limited to 20 nt, incorporated with a background dataset of 1038 upstream and downstream introns from a random set of exons that are not currently associated with exon repetition. The identified motifs in MEME were searched in the different datasets using the program MAST (<http://meme.sdsc.edu/meme/mast-intro.html>) (Bailey and Gribskov, 1998). The identified motifs were represented as weight matrices and presented in a graphical format with the Pictogram program (<http://genes.mit.edu/pictogram.html>) (Lim and Burge, 2001). The program Einverted from the EMBOSS package (version 3.0) was used to search for inverted repeats sequences in intronic sequences with default parameters (<http://emboss.sourceforge.net/>) (Rice *et al.*, 2000). The program Needle from the EMBOSS package (version 3.0) was used to undertake global sequence alignments with default parameters (<http://emboss.sourceforge.net/>) (Rice *et al.*, 2000). The program RNAcofold from the Vienna RNA package (version 1.6.1) was used with default parameters to predict the RNA secondary structure of two complementary RNA sequences (<http://www.tbi.univie.ac.at/~ivo/RNA/>) (Bernhart *et al.*, 2006). The locations and sequences of transposable elements in the human genome was obtained from the RepeatMasker track of the UCSC genome browser (Karolchik *et al.*, 2003). These are the output from RepeatMasker [<http://www.repeatmasker.org>]. The Rat *Sa* gene sequences were also analysed for transposable elements using the RepeatMasker web server.

Table 1. The exon and intron sizes of the control and exon repetition data sets

	Control exons	ExRep exons	<i>P</i> -value
Exon lengths	156 (\pm 21)	330 (\pm 85)	0.008
Upstream intron lengths	3356 (\pm 739)	48247 (\pm 7588)	9.5×10^{-7}
Downstream intron lengths	3209 (\pm 857)	38797 (\pm 6198)	3.1×10^{-7}

Numbers shown are the Mean length (\pm the standard error of the mean) of the exons and introns from the control (1038 exons) and exon repetition (ExRep) data sets (48 exons). An independent samples *t*-test was used to calculate the *P*-values for the average sizes between these two samples with a 95% confidence interval.

3 RESULTS

3.1 Detection of candidate exon repetition events in expressed sequences

To maximize the number of exon repetition events available for analysis, we extended our previous genome wide survey (Dixon *et al.*, 2005) for candidate single exon repetition events in human expressed sequences from Genbank. From this search a total of 54 human expressed sequences from 48 human genes were identified as candidate single exon repetition events (Supplementary Table A). We chose to utilize only the single exon repetition events for sequence motif analyses as these present a more simplified system of exon repetition to analyse than multiple exon repetition events.

3.2 Features of the single exon repetition events and the flanking intron sequences

We examined the nucleotide composition of the repeated exons and their flanking introns, to elucidate whether they deviate from the control sequences (exons and introns that are not associated with exon repetition). As shown in Supplementary Table 1, the composition of the repeated exons was not significantly different from that of the control exons. In both repeated and control exons, guanine and cytosine residues were over-represented. The introns flanking the repeated exons were not significantly different from the control introns, both sets being enriched in adenine and thymine residues. The lengths of the repeated exons and their flanking introns were compared as shown in Table 1. Our null hypothesis that the lengths of the introns from the control and exon repetition groups were not different was rejected. Also, the exons involved in exon repetition are significantly longer than the control exons.

3.3 Intron and exon sequence motif search

We examined the exons and flanking introns of the 48 human single exon repetition events that we identified in the genome wide survey described above. The common motifs identified by MEME in the exon sequences were of low informational value (<10 bits), compared to the identified motifs in both the upstream and downstream intron sequences that were of high informational content (>30 bits; Fig. 1). We cannot rule out the existence of exonic motifs that regulate exon repetition, however they are probably much shorter than the 20 base window of our search and may be less well defined. It has been shown previously (Miriami *et al.*, 2003) that using a 20 base window with MEME is appropriate for detecting intron sequence motifs associated with alternative splicing. Therefore we decided to concentrate on the intronic motifs that we discovered.

a single instance of each motif in the flanking introns. Of these 15 events, 9 are the result of Alu insertions in both flanking introns, to produce an inverted Alu pair. Interestingly, 4 of the 43 exon repetition events in Supplementary Figure 1 (designated with 'w') have single instances of the motifs in both flanking introns in which one of them is a weak motif (15–17 out of 20 bases match with the consensus motif), yet for 3 out of these 4 events the sequence complementarities between the motifs is maintained (>90%). These weak motif instances do not coincide with the location of a transposable element. Moreover, for the 15 exon repetition events that display a single instance of each motif in the flanking introns, 14 of them exhibit >90% sequence complementarity between the pair of motifs.

3.5 The potential role of intron motifs in exon repetition

The most apparent pattern that we observed from the distribution of our two intron motifs (Supplementary Figure 1) was that, of the 43 single exon repetition events that contained a >75% sequence match to both consensus motif sequences, all contained at least one instance of each motif in the flanking introns. A number of these events contained only one instance of the motifs in their entire intron sequence, suggesting that they are not involved in causing stem-loop structures within the same intron. Also, in instances where there are multiple copies of both motifs in the flanking introns, the order and the distance from the exon of the pyrimidine-rich and purine-rich motifs is not conserved suggesting that these motifs are not involved in creating a stem-loop structure with both flanking introns to form a loop containing the exon. Indeed, we found no evidence within the Alternative Splicing Database (Stamm *et al.*, 2006) that any of these 43 exons are skipped in linear mRNA spliceforms transcribed from their respective genes. These observations suggest a model in which the complementary intron sequence motifs could be involved in base-pairing between separate pre-mRNA transcripts. In Figure 2, a model for the possible role of these intron motifs in exon repetition is illustrated for a gene that exhibits non-linear splicing of exon 2, producing a duplicated exon 2 in the resultant mRNA. Folding these two intron sequence motifs by the RNAfold program reveals a predicted RNA duplex secondary structure with free energy of -42 kcal/mol.

To investigate the potential role of complementary intron sequences in exon repetition we analysed the rat *Sa* gene sequence in two different strains, the WKY strain which exhibits exon repetition isoforms 2-2 and 2-3-4-2-3-4 and the SHR strain which does not exhibit exon repetition at all. The sequences of cloned *Sa* genes from WKY and SHR rats are available from GenBank with accession nos AY456695 (WKY) and AY455861 (SHR). A previous detailed comparative analysis of these sequences (Rigatti *et al.*, 2004) has shown that the major sequence difference between the strains in this gene is a LINE element of 1.4 kb in the SHR intron 1 that is absent in WKY. The program Einverted was used to detect complementary sequences between the introns 1 and 4 in both strains. We found the same complementary sequences between introns 1 and 4 for both strains (Fig. 3a) suggesting that these could be involved in base pairing between separate transcripts. A global alignment of the two sequences using the EMBOSS needle program showed that there was only one single nucleotide difference between the strains in the intronic complementary regions

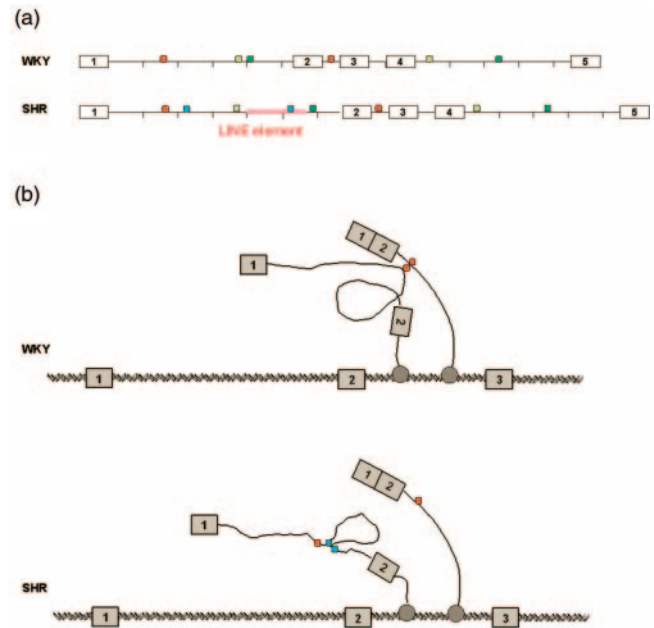


Fig. 3. Analysis of the rat *Sa* gene for the involvement of complementary intron base pairing in exon repetition. (a) The diagram depicts the identified complementary intron sequences between the first four introns of the *Sa* gene in the WKY and SHR rat strains. Complementary sequences (>20 bp) were identified using the Einverted program. Identified complementary regions were analysed with the RNAfold program to determine the free energy of the potential RNA duplex. The introns are shown as lines with vertical bars indicating the length in increments of 1 kb. The exons are shown as boxes. The purple line within the intron 1 of the SHR strain only represents the 1.4 kb LINE element insertion. The red square represents a complementary sequence between introns 1 and 2, that consists of 24 bases and a predicted RNA duplex structure with a free energy of -17.3 kcal/mol. The yellow square represents a complementary sequence between introns 1 and 4 that consists of 27 bases and a predicted RNA duplex structure with a free energy of -11.6 kcal/mol. The green square represents a complementary sequence between introns 1 and 4 that consists of 52 bases and a predicted RNA duplex structure with a free energy of -85.6 kcal/mol. The turquoise square represents a complementary sequence (inverted repeat) between two regions in intron 1 of the SHR strain only that consists of 130 bases and a predicted RNA duplex structure with a free energy of -89 kcal/mol. (b) A diagram that depicts the possible mechanism by which exon repetition of exon 2 occurs in the WKY strain and a prediction of why this is absent in the SHR strain. The 24 base complementary sequence (red square) between introns 1 and 2 enables the base pairing between the separate transcripts that are both tethered to separate RNA polymerase II molecules within intron 2. This RNA secondary structure involving separate transcripts enables *trans*-splicing to occur as shown in Figure 2. However, the presence of the LINE element in the intron 1 of the SHR strain introduces a large stem-loop structure that is stabilized by the inverted repeat sequences (turquoise square). This stem-loop interferes with the base pairing potential between the separate transcripts and therefore prevents *trans*-splicing occurring. This model can be applied similarly to the production of exon repetition isoforms 2-3-4-2-3-4 in the WKY strain and the absence of this non-linear isoform in the SHR strain. The grey circle on the DNA strand represents the RNA polymerase II molecule.

identified in Figure 3a. However, the LINE element insertion in the SHR intron 1 sequence introduces an extra potential base pairing interaction within intron 1, that would cause a large stem-loop structure within intron 1 (free energy of -89 kcal/mol) (Fig. 3a).

This intron 1 stem-loop structure would not be present in the WKY strain. We hypothesize (Fig. 3b) that this major sequence difference causes a major secondary structure difference that prevents the intronic base pairing interactions that take place in the WKY strain and are pivotal in enabling exon repetition to take place.

The contribution made by rodent mobile elements to the complementary sequences depicted in Figure 3a were investigated by scanning the rat *Sa* gene sequences with RepeatMasker to identify the locations of transposable elements. Only the complementary sequences depicted as a yellow square in Figure 3a coincided with the location of a mobile element, with the yellow square in intron 1 matching a SINE/ID element and the complementary yellow square sequence in intron 4 matching a SINE/Alu element.

4 DISCUSSION

Our computational analysis has enabled us to identify two intron sequence motifs that are associated with human exon repetition events. We found no major motif within the exons of these events. Moreover, it is unlikely that exon repetition is regulated by exon sequences as studies on the rat *COT* gene (2-2 isoform) have shown that the exon 2 involved is exactly the same sequence in strains that exhibit exon repetition and in those that do not. The distribution and complementary features of the sequence motifs in the flanking introns of single and multiple exon repetition events suggests that partial base pairing between the introns of two separate precursor mRNAs is involved in the natural mechanisms of this phenomenon. This model of mammalian *trans*-splicing between pre-mRNAs from the same gene has been proposed previously from studies on individual mammalian genes (Eul et al., 1995; Takahara et al., 2000, 2005). Our investigations with the largest collection of human exon repetition events from our genome wide survey support this model. It is intriguing also that studies on the human *Sp1* gene (Takahara et al., 2005) have shown that the presence of a long intron promotes *trans*-splicing, as our collection of exon repetition events are also characterized by long flanking introns. The presence of long introns flanking an exon would increase the time taken for synthesis of the introns by RNA polymerase II and therefore increase the likelihood of the introns from separate pre-mRNA molecules interacting in a RNA-RNA secondary structure. The major difference from conventional *trans*-splicing is that highly efficient exon repetition in endogenous genes is determined in *cis*, meaning that transcripts from two alleles do not splice to each other (Rigatti et al., 2004). Instead, it is likely that the splicing takes place between two RNA molecules while they are attached to the same gene (Rigatti et al., 2004). It is consistent with this model that a role for transcriptional pausing (Proudfoot et al., 2002) is evident, as was shown for the human *Sp1* gene (Takahara et al., 2005) and the rat *Sa* gene (J. -H. Jia and A. Sidorov, unpublished work). Other factors such as transcriptional pausing are likely to be involved and could explain why 5 of the 48 human single exon repetition events used in this study did not contain instances of the intron motifs. It has also been shown previously (Konarska et al., 1985) that the efficiency of *trans*-splicing between pre-mRNA molecules *in vitro* can be increased by the introduction of complementary intronic sequences to enable a short RNA duplex to occur between the separate molecules. Researchers have also utilized the *trans*-splicing process as a tool for gene therapy (Puttaraju et al., 1999) to enable the repair of genetic defects such as cystic fibrosis (Liu et al., 2002) or to

reprogram gene expression. A crucial step in this process is the introduction of an intronic complementary sequence of 20–150 nt to enable the targeted binding of a synthetic pre-mRNA to the intron region of a target pre-mRNA molecule (Garcia-Blanco, 2003; Mansfield et al., 2003). These studies suggest that complementary intron sequences would contribute to exon repetition via intragenic *trans*-splicing.

The observation that the majority of the human intron motifs coincide with the locations of Alu elements is intriguing. The role of transposable elements such as Alu and LINE elements in the shaping of eukaryotic genomes is becoming more recognized (Batzer and Deininger, 2002). For example, Alu element insertions can alter the transcription or the open reading frame of a gene (Batzer and Deininger, 2002) and affect the splicing of a gene by introducing new splice sites (Lev-Maor et al., 2003; Sorek et al., 2002). Not surprisingly, Alu insertions have been found to account for several human genetic disorders (Batzer and Deininger, 2002). Also, various inherited disorders have been caused by Alu-mediated homologous recombination events between dispersed Alu elements in the genome (Batzer and Deininger, 2002). There is also some evidence that Alu elements that insert into an inverted orientation are more prone than others to illegitimate recombination (Lobachev et al., 2000; Stenger et al., 2001), consequently there is a genomic depletion of inverted Alu repeats that are <1 kb apart due to evolutionary negative selection. Interestingly, research into RNA (adenosine to inosine) editing (Kim et al., 2004), has shown that Alu sequences within mRNA transcripts are major targets for RNA editing. This is likely to be due to the nature of the ADAR (Adenosine Deaminase that Act on RNA) family of RNA editing enzymes, to recognize and bind any dsRNA structure. An explanation would be that the inverted Alu pairs within the pre-mRNA transcript form extended RNA duplexes either within the same transcript or in *trans*. Our results would suggest that the insertion of mobile elements into the genome that create inverted repeat sequences across different introns (at distances >1 kb) from the same gene could increase the potential of a gene to undertake exon repetition via *trans*-splicing between separate transcripts from the same gene.

The functional significance of exon repetition has yet to be conclusively established. The absence of exon repetition in *Sa* mRNA from one strain of rat suggests that it is not essential to the viability of the animal and the functional role of the *Sa* non-linear transcripts are still not understood (Frantz et al., 1999; Rigatti et al., 2004). More detailed research needs to be undertaken on individual examples of exon repetition to elucidate the effect of this alternative splicing phenomenon on the biological function of the gene. Nevertheless, it is likely that exon repetition is subject to selection in some genes with its subsequent effects contributing to the phenotypic variation of a species. It is also likely that exon repetition is a ubiquitous possibility that is selected against in most genes due to the potential deleterious effects of non-linear spliceforms within the cell. As the number of aberrant splicing processes causing human disease is growing (Buratti et al., 2006), it is likely that some human diseases will in the future be linked to deviant exon repetition events. In this study we have added some further insight into the elucidation of the mechanisms of exon repetition. Our observations should help experimentalists to focus on the involvement of RNA secondary structures in the mechanisms of exon repetition. Only by understanding in further detail how this splicing phenomenon occurs to produce non-linear mRNA transcripts and

how most genes seemed to have evolved to avoid it will we understand the fundamentals of gene expression. This line of research could be pivotal in our understanding of how alternative splicing evolved to contribute to the complexity of multi-cellular life.

ACKNOWLEDGEMENTS

This work was supported by the Wellcome Trust Functional Genomics Initiative in Cardiovascular Genetics and a MRC Cooperative grant on variability, instability and pathology of the human genome.

Conflict of Interest: None declared.

REFERENCES

- Bailey,T.L. and Elkan,C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **2**, 28–36.
- Bailey,T.L. and Gribskov,M. (1998) Combining evidence using p-values: application to sequence homology searches. *Bioinformatics*, **14**, 48–54.
- Batzer,M.A. and Deininger,P.L. (2002) Alu repeats and human genomic diversity. *Nat. Rev. Genet.*, **3**, 370–379.
- Bernhart,S. *et al.* (2006) Partition function and base pairing probabilities of RNA heterodimers. *Algor Mol Biol.*, **1**, 3.
- Black,D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.*, **72**, 291–336.
- Buratti,E. and Baralle,F.E. (2004) Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol. Cell. Biol.*, **24**, 10505–10514.
- Buratti,E. *et al.* (2006) Defective splicing, disease and therapy: searching for master checkpoints in exon definition. *Nucleic Acids Res.*, **34**, 3494–3510.
- Caudevilla,C. *et al.* (1998) Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver. *Proc. Natl Acad. Sci. USA*, **95**, 12185–12190.
- Dixon,R.J. *et al.* (2005) A genome-wide survey demonstrates widespread non-linear mRNA in expressed sequences from multiple species. *Nucleic Acids Res.*, **33**, 5904–5913.
- Eperon,L.P. *et al.* (1988) Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? *Cell*, **54**, 393–401.
- Eul,J. *et al.* (1995) Experimental evidence for RNA trans-splicing in mammalian cells. *EMBO J.*, **14**, 3226–3235.
- Frantz,S.A. *et al.* (1999) Exon repetition in mRNA. *Proc. Natl Acad. Sci. USA*, **96**, 5400–5405.
- Garcia-Blanco,M.A. (2003) Messenger RNA reprogramming by spliceosome-mediated RNA trans-splicing. *J. Clin. Invest.*, **112**, 474–480.
- Johnson,J.M. *et al.* (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science*, **302**, 2141–2144.
- Karolchik,D. *et al.* (2003) The UCSC Genome Browser Database. *Nucleic Acids Res.*, **31**, 51–54.
- Kim,D.D. *et al.* (2004) Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res.*, **14**, 1719–1725.
- Konarska,M.M. *et al.* (1985) Trans splicing of mRNA precursors *in vitro*. *Cell*, **42**, 165–171.
- Ladd,A.N. and Cooper,T.A. (2002) Finding signals that regulate alternative splicing in the post-genomic era. *Genome Biol.*, **3**, reviews0008.
- Lawrence,C.E. *et al.* (1993) Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science*, **262**, 208–214.
- Lev-Maor,G. *et al.* (2003) The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. *Science*, **300**, 1288–1291.
- Lim,L.P. and Burge,C.B. (2001) A computational analysis of sequence features involved in recognition of short introns. *Proc. Natl Acad. Sci. USA*, **98**, 11193–11198.
- Liu,X. *et al.* (2002) Partial correction of endogenous DeltaF508 CFTR in human cystic fibrosis airway epithelia by spliceosome-mediated RNA trans-splicing. *Nat. Biotechnol.*, **20**, 47–52.
- Lobachev,K.S. *et al.* (2000) Inverted Alu repeats unstable in yeast are excluded from the human genome. *EMBO J.*, **19**, 3822–3830.
- Maniatis,T. and Tasic,B. (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature*, **418**, 236–243.
- Mansfield,S.G. *et al.* (2003) 5' exon replacement and repair by spliceosome-mediated RNA trans-splicing. *RNA*, **9**, 1290–1297.
- Matlin,A.J. *et al.* (2005) Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell. Biol.*, **6**, 386–398.
- Miriama,E. *et al.* (2003) Conserved sequence elements associated with exon skipping. *Nucleic Acids Res.*, **31**, 1974–1983.
- Modrek,B. and Lee,C. (2002) A genomic view of alternative splicing. *Nat. Genet.*, **30**, 13–19.
- Proudfoot,N.J. *et al.* (2002) Integrating mRNA processing with transcription. *Cell*, **108**, 501–512.
- Puttaraju,M. *et al.* (1999) Spliceosome-mediated RNA trans-splicing as a tool for gene therapy. *Nat. Biotechnol.*, **17**, 246–252.
- Rice,P. *et al.* (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
- Rigatti,R. *et al.* (2004) Exon repetition: a major pathway for processing mRNA of some genes is allele-specific. *Nucleic Acids Res.*, **32**, 441–446.
- Sorek,R. *et al.* (2002) Alu-containing exons are alternatively spliced. *Genome Res.*, **12**, 1060–1067.
- Stamm,S. *et al.* (2006) ASD: a bioinformatics resource on alternative splicing. *Nucleic Acids Res.*, **34**, D46–D55.
- Stenger,J.E. *et al.* (2001) Biased distribution of inverted and direct Alus in the human genome: implications for insertion, exclusion, and genome stability. *Genome Res.*, **11**, 12–27.
- Takahara,T. *et al.* (2000) Heterogeneous Sp1 mRNAs in human HepG2 cells include a product of homotypic trans-splicing. *J. Biol. Chem.*, **275**, 38067–38072.
- Takahara,T. *et al.* (2005) Delay in synthesis of the 3' splice site promotes trans-splicing of the preceding 5' splice site. *Mol Cell*, **18**, 245–251.