Minigene reporter for identification and analysis of *cis* elements and *trans* factors affecting pre-mRNA splicing

Gopal Singh and Thomas A. Cooper

Baylor College of Medicine, Houston, TX, USA

All human genes contain a diverse array of cis-*acting elements within introns and exons that are required for correct and efficient precursor messenger RNA (pre-mRNA) splicing. Recent computational analyses predict that most human exons contain elements required for splicing coinciding with an appreciation for the high frequency with which mutations that disruption pre-mRNA splicing cause disease. Minigenes provide a means to directly determine whether disease-causing mutations or single nucleotide polymorphisms (SNPs) affect splicing efficiency. Minigenes have also been instrumental in investigations of alternative splicing to identify* cis *elements required for cell-specific splicing events, demonstrating regulation of individual splicing events by specific RNA binding proteins, and correlating binding of these splicing regulators with splicing regulation. Here we present a versatile minigene plasmid vector designed for rapid cloning and analysis of* cis *elements and* trans-*acting factors that influence splicing efficiency or regulate cell-specific splicing. Ubiquitous expression and unique restriction sites allow for straightforward replacement of a variety of gene segments to analyze the effects of nucleotide substitutions on splicing, to identify tissue-specific regulatory elements, or to determine responsiveness to coexpressed proteins or small molecules.*

INTRODUCTION

The vast majority of human genes contain introns that are removed from the precursor messenger RNA (pre-mRNA) during pre-mRNA splicing to generate mature mRNAs. Splicing requires conserved splice site consensus sequences at the intron/exon borders as well as diverse *cis*-acting elements within exons and introns (1,2). Disease-causing mutations within the splice sites or exonic elements frequently disrupt splicing, resulting in loss of function of the affected allele (3,4). In addition, single nucleotide polymorphisms (SNPs) have the potential to affect splicing efficiency and modify disease severity (5) or responsiveness and side effects to pharmaceuticals (6).

The most direct approach to determine whether disease-causing mutations or SNPs are associated with splicing is to perform a reverse transcription PCR (RT-PCR) analysis on RNA from the relevant tissue(s) of affected individuals. However, tissue samples are often not available. An alternative approach is to test the effects of the mutation on splicing using minigenes, in which the relevant genomic segment is cloned within a plasmid between an upstream ubiquitous transcriptional promoter and a downstream gene segment necessary for mRNA 3′ end formation.

In addition to their utility for analyzing the effects of mutations and allelic variants on splicing efficiency, minigenes have been a mainstay of investigations to identify the *cis* elements and *trans*-acting factors that regulate tissue-specific alternative splicing (7,8). Detailed strategies for the use of minigenes to investigate a variety of different questions involving constitutive and alternative splicing have been described (8,9).

Here we describe a versatile plasmid vector for the identification and characterization of *cis*-acting elements that affect splicing efficiency and splicing regulation. Genomic segments cloned into this vector can be used to: (*i*) test whether mutations genetically linked with disease affect the splicing efficiency of an adjacent exon or cause cryptic splicing; (*ii*) test whether SNPs located within or adjacent to an exon affects its splicing efficiency; (*iii*) perform functional analyses of computationally identified putative splicing elements; (*iv*) identify *cis*-acting elements that enhance or repress basal splicing efficiency of specific exons; (*v*) identify *cis*-acting elements required for cell-specific splicing regulation; and (*vi*) determine the responsiveness of alternative exons and flanking introns to overexpression and depletion of specific regulators and identify regions required for responsiveness.

MATERIALS AND METHODS

Plasmids

RHCglo (Figure 1A) was constructed by PCR using priming oligonucleotides containing restriction sites that are unique to the plasmid. The exon (Figure 1B) was derived from oligonucleotides that when annealed generated 5′ overhangs compatible with *Bam*HI and *Xho*I restriction sites. RtauWT, RtauN279K, and RtauΔ280K were made by replacing the RHCglo exon with three derivatives of *MAPT* exon 10 via the *Bam*HI/*Xho*I sites. Exons were PCR-amplified using forward primers containing either unmodified or modified sequence [tauE10wtf (5′-ATATATGGATCCG TGCAGATAATTAATAAGAA-3′), tauE10N279Kf (5′-ATATATGGATCC GTGCAGATAATTAAGAAGAAGCT-3′), tau10del280Kf (5′-TATATGGATC CGTGCAGATAATTAATAAGCTGG ATCTT-3′)], and a reverse primer with unmodified sequence tauE10wtr (5′-AT ATATCTCGAGACTGCCGCCTCCC GGGACGT-3′) amplified on commercially available human genomic DNA. The 5′ splice site was modified using a "doped" oligonucleotide [GLO5 (5′-GGCAGTCTCGAGGTTGKYATC AAGGTTA-3′) as the forward primer in a PCR with a reverse primer located downstream of the *Xba*I site (Figure 1A). The PCR product was used to replace the *Xho*I/*Xba*I fragment. The oligonucleotide generated four possible 5′ splice site sequences, and all four clones were identified by sequencing.

RTB300 contains a genomic segment from human cardiac troponin T (cTNT)
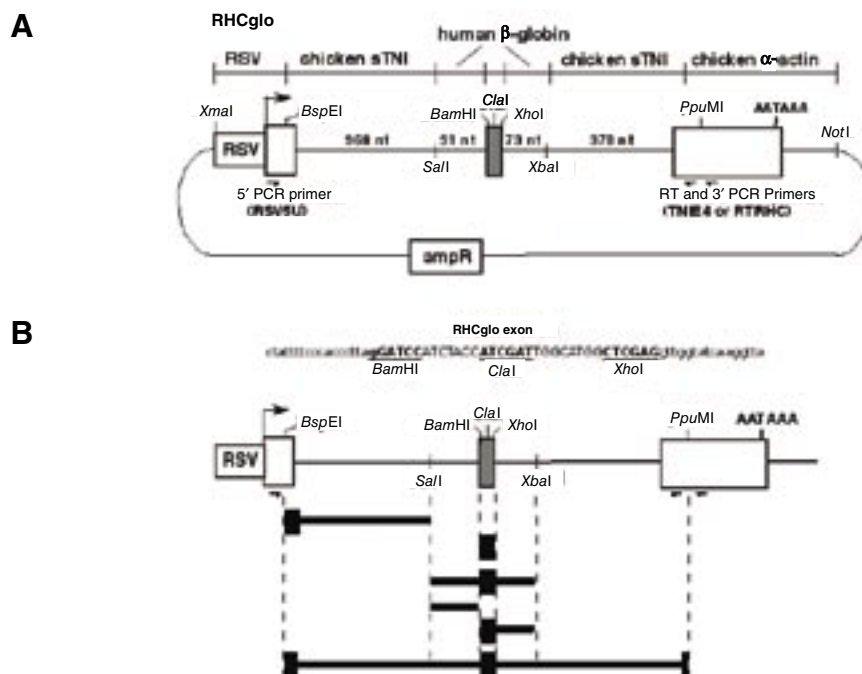
Short Technical Reports

Figure 1. RHCglo splicing reporter plasmid. (A) RHCglo showing all restriction sites that are unique to the plasmid. The middle exon is immediately flanked upstream and downstream by the last and first 91 and 73 nucleotides (nt) of human β-globin intron 1, respectively. The distal upstream segment of intron 1 contains introns 1 and 3 of chicken skeletal troponin I (sTNI), and the distal downstream region of intron 2 contains the last 364 nt of sTNI intron 3. RSV, rous sarcoma virus enhancer/promoter; RT, reverse transcription; ampR, ampicillin resistance gene. (B) Unique restriction sites within the minigene can be used to replace a variety of gene segments including segments within the artificial exon. Exon sequence is in uppercase, and intron sequence is in lowercase.

and was described previously (10). The flanking intronic segments from this minigene were PCR-amplified using oligonucleotides containing appropriate restriction sites and inserted between *Sal*I/*Bam*HI and *Xho*I/*Xba*I restriction sites to construct RF300 (see Figures 1A and 3). The four amino acid insertion (LYLQ) and 41-kDa protein isoforms of human CUG binding protein 1 (CUG-BP1) and muscleblind like 1 (MBNL1), respectively, were expressed with N-terminal FLAG® epitopes. Myotonic dystrophy protein kinase (DMPK) mRNAs containing 960 interrupted CUG repeats were expressed from a previously described minigene (11).

Cell Culture and Transfection

CosM6 cells [a derivative of the COS-7 African green monkey kidney fibroblast cell line (12)] were plated at 150,000 cells/60-mm culture dish in 3 mL growth medium [Dulbecco's modified Eagle medium (DMEM), high glucose supplemented with 10% fetal bovine serum (FBS), 1% penicillin-streptomycin, and 1% L-glutamine; all from GIBCO®; Invitrogen, Carlsbad, CA, USA]. Twenty-four hours after plating, cultures were transfected with 500 ng of minigene plasmid and 1 μg of a protein expression plasmid using FuGENE®6 (Roche Diagnostics, Indianapolis, IN, USA) according to the manufacturer's recommendations. Total RNA was extracted 40 h following transfection. C2C12 mouse myoblasts were plated at 400,000 cells/100-mm culture dish in 10 mL growth medium [DMEM, high glucose supplemented with 10% FBS, 1% penicillin-streptomycin, 1% minimum essential medium (MEM) nonessential amino acids]. Twenty-four hours after plating, cultures were transfected with 2 μg minigene plasmid using FuGENE 6. When the cells reached 80% confluency (usually the next day), cells were harvested for undifferentiated cultures. For the remaining cultures, the media was replaced with differentiation medium (DMEM, 2.5% horse serum, 1% penicillin-streptomycin,

1% MEM nonessential amino acids). Differentiation media was replaced every day until harvest on the fourth day following the switch to differentiation media.

RNA Extraction and RT-PCR

Total RNA was extracted using a modified version of the guanidinium thiocyanate protocol (13). Cells were rinsed with cold phosphate-buffered saline (PBS) and then harvested by the addition of 0.5 mL cold guanidinium thiocyanate solution per 60-mm plate and extracted twice with 1 volume phenol and 0.2 volume chloroform: isoamyl alcohol and one time with chloroform:isoamyl alcohol. RNA was precipitated with 1 volume isopropanol, and the pellet was washed with 70% ethanol and dissolved in 20 μL double-distilled water.

RT-PCR was performed using the downstream PCR primer to also prime reverse transcription. cDNA was generated in 20 μL containing 5 μL of the total RNA extracted (one-quarter of the RNA extracted from one dish), 50 ng $(dT)_{12-18}$ primer (Invitrogen), 5× reaction buffer [included with the avian myeloblastosis virus (AMV); Life Science, St. Petersburg, FL, USA], 0.01 M dithiothreitol (DTT), 0.5 mM dNTPs, and 10 U AMV (Life Science). To start the reverse transcription reaction, the RNA and primers were mixed, heated to 65°C for 10 min, chilled on ice, and then the remaining agents were added. The reaction mixture was incubated at 42°C for 1 h and at 65°C for 10 min and then either used immediately or stored at -20°C.

PCR including the ^{32}P kinase-labeled upstream primer was performed in 40 μL volume as described previously (14). The upstream primer was RSV5U (5′-CATTCACCACATTGGTGTGC-3′). One of two downstream primers was used for RT and PCR, either TNIE4 (5′-AGGTGCTGCCGCCGG GCGGTGGCTG-3′), which anneals upstream of the *Ppu*MI site, or RTRHC 5′-GGGCTTTGCAGCAACAGTAAC-3′), which anneals downstream of the *Ppu*MI (Figure 1A). For all results presented, RT-PCR was performed using RSV5U/TNIE4. Ratios of exon inclusion/exclusion were quantified
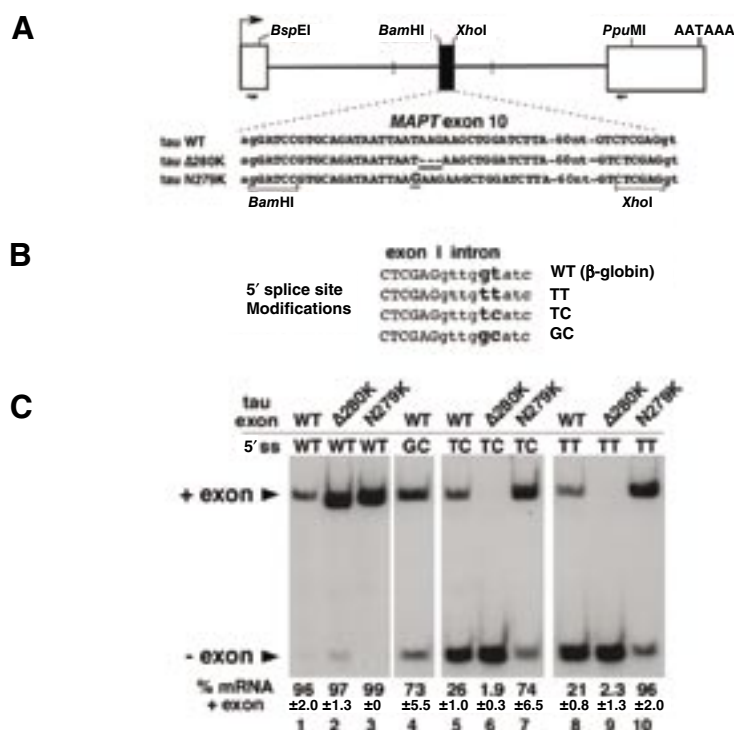
Figure 2. Tau exon 10 mutations that cause frontotemporal dementia with Parkinsonism linked to chromosome 17 (FTDP-17) reproduce splicing defects in the RHCglo splicing reporter plasmid. (A) Tau exon 10 variants replaced the RHCglo exon. Wild-type (WT), Δ280K, and N279K mutations were PCR-amplified from normal human genomic DNA using oligonucleotides containing wild-type or mutant sequences and *Bam*HI and *Xho*I flanking restriction sites. (B) Substitutions within the 5′ splice site of the downstream intron were performed in a separate PCR amplification of the *Xho*I/*Xba*I fragment using oligonucleotides containing substitutions within the indicated positions (bold larger font) of the β-globin 5′ splice site. (C) Reverse transcription PCR (RT-PCR) analysis of tau exon 10 alternative splicing. ss, single-stranded; mRNA, messenger RNA.

using a Molecular Dynamics® PhosphorImager SF (GE Healthcare, Piscataway, NJ, USA).

Western Blotting Analysis

Cell culture plates were rinsed with cold PBS, harvested in protein loading buffer, and the blots were prepared as described previously (11). Blots were probed with monoclonal anti-FLAG M2 peroxidase conjugate (Sigma-Aldrich, St. Louis, MO, USA) at a dilution of 1:5000.

RESULTS AND DISCUSSION

Description of the RHCglo Minigene Plasmid

A diagram of the RHCglo plasmid is shown in Figure 1A. The Rous sarcoma virus long-terminal repeat transcribes in a large variety of cell lines. mRNA 3′ end formation is directed by the chicken skeletal α actin last exon and 3′ flanking genomic region (15). The other gene segments have been shown to provide efficient splicing with minimal cryptic splice site usage (10,15,16). A key feature of this minigene is the presence of unique restriction sites within all three exons and both introns to allow straightforward "cut and paste" cloning of a large variety of genomic segments to test for splicing activity including segments within the middle exon (Figures 1, A and B). RT-PCR analysis is performed using oligonucleotides complementary to the first and last exons.

Test of Exon Mutations

To determine whether the RHCglo minigene can be used to identify disease-causing splicing mutations within exons, we tested two well-characterized splicing mutations within *MAPT* (tau) exon 10. Exon 10 is alter-

natively spliced to provide an appropriate ratio of tau protein isoforms containing three and four microtubule binding domains. Mutations within tau exon 10 or the surrounding introns cause frontotemporal dementia with Parkinsonism linked to chromosome 17 (FTDP-17) by disrupting the balanced expression of these isoforms (17). The Δ280K mutation (Figure 2A) disrupts an exonic splicing enhancer (ESE), causing decreased levels of exon 10 inclusion. N279K is a T to G substitution that results in increased ESE activity and a higher level of exon inclusion (18,19). Wild-type (WT), N279K, and Δ280K exons were amplified from human genomic DNA and inserted between the *Bam*HI and *Xho*I sites in RHCglo (Figure 2). The WT, N279K, and Δ280K minigenes were transiently transfected into COSM6 cells, and RT-PCR demonstrated that for all three minigenes, 95%–99% of the mRNAs contained *MAPT* exon 10 (Figure 2C, lanes 1–3).

We hypothesized that the efficient recognition of the exons was due to features outside of the exon and concealed differences in the effects of the different exon sequences on exon recognition. To reveal the potential effects of the FTDP-17 exon mutations on exon inclusion, we weakened the 5′ splice site by introducing mutations into the fifth and/or sixth positions of the consensus (AG|GTRAGT). Splice site strength can be evaluated using Splice Site Prediction by Neural Network available at www.fruitfly.org/seq_tools/splice.html (20). A doped oligonucleotide containing wild-type and mutant nucleotides in equal amounts at these positions was used to PCR amplify a *Xho*I/*Xba*I fragment containing the substitutions. The approach generates three mutants (plus the wild-type 5′ splice site) from one PCR cloning step and provided minigenes with a range of splicing efficiencies. Importantly, weakening the strength of the 5′ splice site revealed dramatic differences in the effects of the nucleotide changes within the exon. As observed in FTDP-17 tissues, the Δ280K deletion substantially reduced inclusion of the exon in the context of the TC or TT substitutions (Figure 2, compare lanes 5 and 6 and lanes 8 and 9). Similarly, the N279K
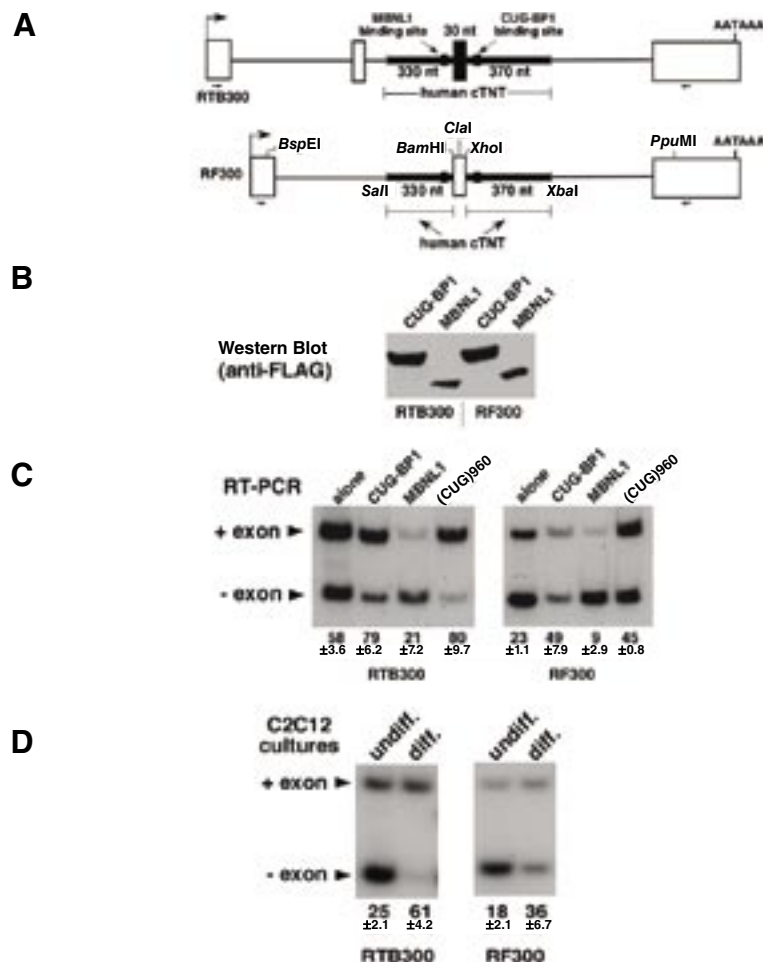
Figure 3. Regulation of the heterologous RHCglo exon by exogenous and endogenous proteins. (A) RF300 contains the flanking intron segments of human cardiac troponin T (cTNT) exon 5. (B) Regulation of the heterologous exon in RF300 by co-expression of FLAG-tagged CUG binding protein 1 (CUG-BP1) or muscleblind like 1 (MBNL1) of CUG repeat-containing RNA [(CUG)960] was comparable to that observed for RTB300. (C) Western blot analysis demonstrating comparable expression of FLAG-CUG-BP1 and FLAG-MBNL1 with RTB300 or RF300. (D) Responses of RTB300 and RF300 minigenes to endogenous regulators during differentiation of the C2C12 mouse skeletal muscle cell line. Reverse transcription PCR (RT-PCR) analysis of minigene RNA extracted from cultures in proliferation media (undiff.) and following 4 days in differentiation media (diff.). nt, nucleotides.

substitution increased inclusion (Figure 2, compare lanes 5 and 7 and lanes 8 and 10). These results demonstrate the utility of this minigene for testing exonic mutations as well as the importance of having a context in which the effects of mutations are not masked by other determinants of exon strength.

Test of Regulation of Alternative Splicing by *trans*-Acting Factors

cTNT exon 5 is antagonistically regulated by CUG-BP1 and MBNL1, which bind to separate intronic sites located downstream and upstream of the exon, respectively (Figure 3A) (11). The flanking intron segments from the complete cTNT genomic fragment (Figure 3A, RTB300) were placed into the RHCglo context (Figure 3A, RF300) to determine whether the regulatory elements retained responsiveness to splicing regulators. The 3′ and 5′ ends of introns 4 and 5 were PCR-amplified using oligonucleotides containing *Sal*I/*Bam*HI and *Xho*I/*Xba*I restriction sites, respectively, and inserted into RHCglo to generate RF300 (Figure 3A). Comparable levels of FLAG-tagged CUG-BP1 or MBNL1 proteins were co-expressed with the two minigenes as determined by Western blot analysis for the FLAG epitope (Figure 3B). We also tested whether the cTNT intronic segments retain responsiveness to

coexpressed CUG repeat-containing RNAs [(CUG)960] as demonstrated previously (10,11) (Figure 3).

The basal level of exon inclusion was lower for RF300. However, the RF300 minigene responded to CUG-BP1, MBNL1, and CUG repeat RNA comparably to RTB300 (Figure 3C). These results demonstrate that the intronic elements within the cTNT pre-mRNA retain their responsiveness to three different modulators of exon inclusion when placed in the RHCglo context.

Test of Regulation by Cell-Specific Elements

We next tested whether the *cis* elements within the contiguous genomic segment (RTB300) and in the RHCglo context (RF300) would respond to endogenous splicing regulators during differentiation of the C2C12 mouse myoblast cell line (21). C2C12 myoblasts proliferate in high serum medium and differentiate into skeletal muscle myotubes in low serum (22). RTB300 and RF300 plasmids were transfected, while the cells were approximately 40% confluent in growth (high serum) media. RNA from undifferentiated cultures was extracted 24 h after transfection, growth media was replaced by differentiation media (low serum), and RNA was extracted from differentiated cultures 4 days later. Analysis of splicing demonstrated a strong shift toward exon inclusion in differentiated cultures for the RTB300 minigene (Figure 3D). The response of the RF300 alternative exon to regulation during C2C12 differentiation is somewhat reduced compared with RTB300. Overall, however, the results demonstrate that a heterologous exon flanked by the two intron segments proximal to human cTNT exon 5 retained the majority of the regulatory response observed during C2C12 skeletal muscle differentiation in vitro.

Minigene splicing reporters are used to determine the effects of nucleotide substitutions on splicing efficiency or to identify *cis*-acting elements that respond to either differential regulatory environments, overexpression, and depletion of specific regulatory proteins, or to identify compounds that affect splice site selection. The design of the

RHCglo minigene splicing reporter facilitates rapid cloning of genomic segments for these analyses. Similarly, gain-of-function analyses can be readily performed by inserting putative regulatory elements into the heterologous RHCglo context. The caveats of minigene splicing reporters should also be considered when analyzing results. These include the unpredictable effects introduced by a heterologous genomic context, the possibility that important regulatory elements are missing from the genomic segment being tested, the effects of overexpressed pre-mRNA and protein, and the loss of potential coupling between splicing regulation and specific transcriptional promoters (8).

REFERENCES

1. **Fu, X.D.** 2004. Towards a splicing code. Cell *119*:736-738.
2. **Matlin, A.J., F. Clark, and C.W. Smith.** 2005. Understanding alternative splicing: towards a cellular code. Nat. Rev. Mol. Cell Biol. *6*:386-398.
3. **Cartegni, L., S.L. Chew, and A.R. Krainer.** 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat. Rev. Genet. *3*:285-298.
4. **Faustino, N.A. and T.A. Cooper.** 2003. Pre-mRNA splicing and human disease. Genes Dev. *17*:419-437.
5. **Nissim-Rafinia, M. and B. Kerem.** 2002. Splicing regulation as a potential genetic modifier. Trends Genet. *18*:123-127.
6. **Bracco, L. and J. Kearsey.** 2003. The relevance of alternative RNA splicing to pharmacogenomics. Trends Biotechnol. *21*:346-353.
7. **Black, D.L.** 2003. Mechanisms of alternative pre-messenger RNA splicing. Annu. Rev. Biochem. *27*:27-48.
8. **Cooper, T.A.** 2005. Use of minigene systems to dissect alternative splicing elements. Methods *37*:331-340.
9. **Cooper, T.A.** 1999. Strategies for defining pre-mRNA cis elements that regulate cell specific splicing, p. 391-403. *In* S.R. Haynes (Ed.), Methods in Molecular Biology. Humana Press, Totowa.
10. **Philips, A.V., L.T. Timchenko, and T.A. Cooper.** 1998. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. Science *280*:737-741.
11. **Ho, T.H., N. Charlet-B., M.G. Poulos, G. Singh, M.S. Swanson, and T.A. Cooper.** 2004. Muscleblind proteins regulate alternative splicing. EMBO J. *23*:3103-3112.
12. **Lovenberg, T.W., M.G. Erlander, B.M. Baron, M. Racke, A.L. Slone, B.W. Siegel, C.M. Craft, J.E. Burns, et al.** 1993. Molecular cloning and functional expression of 5-HT1E-like rat and human 5-hydroxytryptamine receptor genes. Proc. Natl. Acad. Sci. USA *90*:2184-2188.
13. **Chomczynski, P. and N. Sacchi.** 1987. Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. *162*:156-159.
14. **Ladd, A.N., N. Charlet-B., and T.A. Cooper.** 2001. The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. Mol. Cell. Biol. *21*:1285-1296.
15. **Ryan, K.J. and T.A. Cooper.** 1996. Muscle-specific splicing enhancers regulate inclusion of the cardiac troponin T alternative exon in embryonic skeletal muscle. Mol. Cell. Biol. *16*:4014-4023.
16. **Charlet-B, N., P. Logan, G. Singh, and T.A. Cooper.** 2002. Dynamic antagonism between ETR-3 and PTB regulates cell type-specific alternative splicing. Mol. Cell *9*:649-658.
17. **Ingram, E.M. and M.G. Spillantini.** 2002. Tau gene mutations: dissecting the pathogenesis of FTDP-17. Trends Mol. Med. *8*:555-562.
18. **Jiang, Z., H. Tang, N. Havlioglu, X. Zhang, S. Stamm, R. Yan, and J.Y. Wu.** 2003. Mutations in tau gene exon 10 associated with FTDP-17 alter the activity of an exonic splicing enhancer to interact with Tra2 beta. J. Biol. Chem. *278*:18997-19007.
19. **D'Souza, I., P. Poorkaj, M. Hong, D. Nochlin, V.M. Lee, T.D. Bird, and G.D. Schellenberg.** 1999. Missense and silent tau gene mutations cause frontotemporal dementia with parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements. Proc. Natl. Acad. Sci. USA *96*:5598-5603.
20. **Reese, M.G., F.H. Eeckman, D. Kulp, and D. Haussler.** 1997. Improved splice site detection in Genie. J. Comput. Biol. *4*:311-323.
21. **Yaffe, D. and O. Saxel.** 1977. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. Nature *270*:725-727.
22. **Neville, C., N. Rosenthal, M. McGrew, N. Bogdanova, and S. Hauschka.** 1997. Skeletal muscle cultures, p. 85-116. *In* C.P. Emerson and H.L. Sweeney (Eds.), Methods Cell Biology. Academic Press, San Diego.