

Disease-Causing Mutations Improving the Branch Site and Polypyrimidine Tract: Pseudoexon Activation of LINE-2 and Antisense *Alu* Lacking the Poly(T)-Tail

David Meili,¹ Jana Kralovicova,² Julian Zagalak,¹ Luisa Bonafé,³ Laura Fiori,⁴ Nenad Blau,¹ Beat Thöny,^{1*} and Igor Vorechovsky^{2*}

¹Division of Clinical Chemistry and Biochemistry, University Children's Hospital Zürich, Switzerland; ²University of Southampton, School of Medicine, Southampton, United Kingdom; ³Division de Pédiatrie Moléculaire, Centre Hôpitalier Universitaire Vaudois, Lausanne, Switzerland; ⁴Department of Pediatrics, San Paolo Hospital, University of Milan, Milan, Italy

Communicated by Haig H. Kazanian, Jr.

Received 3 October 2008; accepted revised manuscript 8 December 2008.

Published online 20 January 2009 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/humu.20969

ABSTRACT: Cryptic exons or pseudoexons are typically activated by point mutations that create GT or AG dinucleotides of new 5' or 3' splice sites in introns, often in repetitive elements. Here we describe two cases of tetrahydrobiopterin deficiency caused by mutations improving the branch point sequence and polypyrimidine tracts of repeat-containing pseudoexons in the *PTS* gene. In the first case, we demonstrate a novel pathway of antisense *Alu* exonization, resulting from an intronic deletion that removed the poly(T)-tail of antisense *AluSq*. The deletion brought a favorable branch point sequence within proximity of the pseudoexon 3' splice site and removed an upstream AG dinucleotide required for the 3' splice site repression on normal alleles. New *Alu* exons can thus arise in the absence of poly(T)-tails that facilitated inclusion of most transposed elements in mRNAs by serving as polypyrimidine tracts, highlighting extraordinary flexibility of *Alu* repeats in shaping intron–exon structure. In the other case, a *PTS* pseudoexon was activated by an A>T substitution 9 nt upstream of its 3' splice site in a LINE-2 sequence, providing the first example of a disease-causing exonization of the most ancient interspersed repeat. These observations expand the spectrum of mutational mechanisms that introduce repetitive sequences in mature transcripts and illustrate the importance of intronic mutations in alternative splicing and phenotypic variability of hereditary disorders.

Hum Mutat 30:823–831, 2009. © 2009 Wiley-Liss, Inc.

KEY WORDS: long interspersed nuclear element; *Alu*; hyperphenylalaninemia; tetrahydrobiopterin; *PTS*

Introduction

Inclusion of intronic sequences in messenger RNA by aberrant splicing is a common cause of genetic disease. This usually occurs through mutation-induced activation of aberrant splice sites in the vicinity of authentic splice sites (reviewed in [Buratti et al., 2007a; Vorechovsky, 2006]). Less frequently, more distant intronic sequences are incorporated in mature transcripts through activation of decoy 3' or 5' splice sites close to newly created splice sites, leading to cryptic exons or pseudoexons [Buratti et al., 2006; Kralovicova and Vorechovsky, 2007]. The same events played an important role in gene evolution by modifying exon–intron structure, and were prominent in species with a large number of long introns that contained many repetitive sequences, including short and long interspersed nuclear elements (SINEs and LINEs). In particular, recently acquired exons have been strongly associated with *Alu* repeats, the most prevalent SINEs in primates, representing over 10% of the human genome [Krull et al., 2005; Makalowski et al., 1994; Sela et al., 2007; Sorek, 2007; Sorek et al., 2002; Zhang and Chasin, 2006]. The majority of *Alu* exons originated by alternative splicing of precursor mRNA (pre-mRNA) in an evolutionary process termed “exonization” [Sorek, 2007], largely through point mutations at several *Alu* hotspots that created new 3' [Lev-Maor et al., 2003] or 5' [Sorek et al., 2004] splice sites in a favorable sequence context recognized by the spliceosome machinery. In contrast to a high number of *Alus* partially included in mRNAs, *Alu* repeats exonized by constitutional splicing have been reported only occasionally [Knebelmann et al., 1995; Mitchell et al., 1991; Vervoort et al., 1998], presumably because of deleterious effects of intronic sequences fully incorporated in mature transcripts. Apart from splice sites activated by point mutations at key *Alu* positions [Lev-Maor et al., 2003; Sorek et al., 2004], new *Alu* exons employed additional splice sites both in the sense and antisense orientation [Sela et al., 2007], but molecular processes leading to their utilization are unknown.

Tetrahydrobiopterin (BH₄) is an essential cofactor required for tyrosine hydroxylase, tryptophan hydroxylase, phenylalanine hydroxylase, and nitric oxide synthase [Thöny et al., 2000]. These enzymes produce important neurotransmitters that are necessary for many aspects of brain function, including catecholamines and serotonin. De novo BH₄ synthesis from GTP involves GTP cyclohydrolase I, sepiapterin reductase, and 6-pyruvoyl-tetrahydropterin synthase (PTPS). The most common form of human

Additional Supporting Information may be found in the online version of this article.

The first two authors contributed equally to this work.

*Correspondence to: Beat Thöny, Division of Clinical Chemistry and Biochemistry, Department of Pediatrics, University of Zürich, Steinwiesstrasse 75, CH-8032 Zürich, Switzerland. E-mail: beat.thony@kispi.uzh.ch; or Igor Vorechovsky, Division of Human Genetics, School of Medicine, University of Southampton, Tremona Road, Southampton SO16 6YD, United Kingdom. E-mail: igvo@soton.ac.uk

BH₄ deficiency is due to a lack of PTPS (EC 4.6.1.10), which results from autosomal recessive mutations in the *PTS* gene (MIM# 261640). Over 45 mutated *PTS* alleles reported so far have been associated with a conspicuous phenotypic variability, ranging from mild hyperphenylalaninaemia to severe neurotransmitter deficiency and psychomotor retardation [Thony and Blau, 2006], but cellular determinants of the variable disease expressivity have been poorly understood.

In this work, we present two cases of pseudoexon activation identified in patients with severe and mild BH₄ deficiency. In the former case, we describe a new type of *Alu* exonization by an intronic deletion of the antisense *Alu*Sq poly(T)-tail, which normally facilitates exonization of *Alu* sequences. Instead, this deletion brought a new branch point sequence within the optimal distance from the pseudoexon 3' splice site and removed an upstream AG dinucleotide that repressed this splice site on normal chromosomes. The other case was due to the inclusion of a LINE-2 repeat in the *PTS* mRNA as a result of a point mutation that improved the polypyrimidine tract of the cryptic exon. In addition, we have defined *cis*-acting sequences required for activation of each pseudoexon and examined putative interactions between candidate *trans*-acting factors and each mutation.

Materials and methods

Patients

Patient MD130 was identified by newborn screening for hyperphenylalaninaemia. He was treated with BH₄ (5 mg/kg/day), which was followed by L-Dopa/Carbidopa, 5-OH-tryptophan, and folinic acid substitution therapy. Because he developed tremor and mild hypertonia of the limbs, the dosage was increased to BH₄ (6 mg/kg/day), L-Dopa/Carbidopa (10 mg/kg/day), and 5-OH-tryptophan (6 mg/kg/day). Patient MD96, a girl born to the first cousins of Moroccan origin, was also found by neonatal screening. She was subsequently treated with L-Dopa/Carbidopa (10 mg/kg/day), 5-OH-tryptophan (10 mg/kg/day), and BH₄ (1 mg/kg/day), and hyperphenylalaninemia has been controlled by a low dose of BH₄ (2 mg/kg/day). Laboratory findings in both patients are summarized in Table 1.

Identification of Cryptic Exons in the *PTS* Gene

Mutation analysis was carried out with cDNA and genomic DNA samples prepared from primary dermal fibroblasts and/or whole blood essentially as described [Thony and Blau, 2006]. Cycloheximide (0.75 mg/ml) was added to primary dermal fibroblasts 4 hr before harvest. Total RNA was isolated using the RNeasy Mini Kit (Qiagen, Chatsworth, CA) and the first-strand cDNA was synthesized with the AMV Reverse Transcription System (Promega, Madison, WI) and primer PTPS-53 (5'-TTT CAA TAA ATA GGC ACT CC). Amplification of the *PTS* cDNA was carried out with primers PTPS-53 and 5'-ATG AGC ACG GAA GGT GGT G.

Immunoblotting and Enzyme Assays

PTPS enzyme activity and Western blot analyses of crude extracts from cultured dermal fibroblasts obtained from both patients and from parents of patient MD96 were carried out as described [Leitner et al., 2003; Thony and Blau, 2006]. Western blot analysis with siRNA-transfected cells was performed using the

Table 1. Phenotypes and Genotypes of Two Patients with BH₄ Deficiency

Patient ID (BIODEF ^a number)	Phe (μM)	BH ₄ Test ^b	Neopterin (U) mmol/mol cr.	Neopterin (U) mmol/mol cr.	Neopterin (U) mmol/mol cr.	Bioppterin (U) mmol/mol cr.	Bioppterin (U) mmol/mol cr.	Neopterin (U) mmol/mol cr.	Bioppterin (U) mmol/mol cr.	5-HIAA (CSF) nM	HVA (CSF) nM	PTPS (RBC) μU/g Hb	PTPS (FB) μU/mg Protein	Mutation Designation in DNA (cDNA) ^c
MD130 (ID 363)	1,694	+	38.0	38.0	<0.01	69.0	8.05	31	81	0.34	0.21	Allele 1: g.3760_3816del57 (c.163_164ins45)		
MD96 (ID 503)	1,100	N.D.	5.04	5.04	<0.01	10.1	34.7	527	1023	0.41	0.05	Allele 2: g.7901delA (c.395delA)		
Reference ^d	<120	—	1.40–18.71	1.40–18.71	0.89–7.60	15–35	20–70	150–800	310–1100	11–29	0.4–1.6	g.2666A>T (c.[84–322A>T] + [84–322A>T])		

^aInternational Database of Tetrahydrobiopterin Deficiencies (BIODEF) is available at <http://www.bh4.org/BH4DatabasesBiodef.asp>

^bLoading test was with 20 mg BH₄/kg of body weight.

^cMutation designation is according to the official mutation nomenclature (<http://www.hgvs.org/mutnomen/>). Reference accession numbers for normal alleles: L76259.1 for human genomic DNA; M97655.1 for human cDNA (cDNA numbering starts with 1 at the ATG start codon). The deletion allele in patient MD130 was 55 nt shorter than the reference human genome assembly contig (NT_033899.7 and NW_925173.1), consistent with a length polymorphism of *Alu* poly(T)-tails.

^dReferences values in unaffected controls (BIODEF). Phe, phenylalanine levels in blood; cr., creatinine; 5-HIAA, 5-hydroxyindoleacetic acid; HVA, homovanillic acid; PTPS, 6-pyruvoyl-tetrahydropterin synthase; U, urine; CSF, cerebrospinal fluid; RBC, red blood cells; FB, skin fibroblasts; Hb, hemoglobin; N.D., not determined.

Enhanced Chemiluminescence ECL (Amersham, Lake Forest, IL) detection system and antibodies specific for human PUF60 (provided by Professor Adrian Krainer, Cold Spring Harbor Laboratory), PTB (provided by Professor Christopher Smith, University of Cambridge), U2AF⁶⁵ (Sigma, St. Louis, MO, catalog number U4758) and tubulin (Sigma, catalog number T8535).

In Silico Analysis of *PTS* Sequences

The intrinsic strength of *PTS* splice sites was estimated by the maximum entropy (ME) model [Yeo and Burge, 2004], which gave the best discrimination between authentic and aberrant splice sites [Buratti et al., 2007a; Vorechovsky, 2006]. Auxiliary splicing sequences were analyzed using publicly available prediction algorithms for RESCUE-ESEs (<http://genes.mit.edu/burgelab/rescue-ese/>) [Fairbrother et al., 2002], FAS-ESSs (<http://genes.mit.edu/fas-ess/>) [Wang et al., 2004], PESXs (<http://cubweb.biology.columbia.edu/pesx/>) [Zhang and Chasin, 2004] and exonic splicing regulatory sequences identified through conservation levels of wobble positions and the overabundance of sequence motifs between human and mouse orthologous exons [Goren et al., 2006]. Repetitive sequences were identified using Repeat Masker (<http://www.repeatmasker.org/>) and verified in Repbase (<http://girinst.org>). Intronic *Alus* were aligned with the Clustal-Wallis algorithm (<http://www.ebi.ac.uk>, v. 2).

Splicing Reporter Plasmids

DNA fragments encompassing *PTS* exons 2 through 4 in patient MD130 and exon 1 and 2 in patient MD96 were amplified from patients' and control DNA using a mixture of *Taq* and *Pfu* polymerases (Promega) and cloned into pCR3.1 (Invitrogen, Carlsbad, CA). The former minigene was cloned using primers 5'-ATA TGC TAG CAC TTG TGT CAT GCT GAC TTT T and 5'-ATA CTC TAG ACC CAA CAT GCC ATT ACC TC, the latter reporter was cloned using primers 5'-ATA TGC TAG CGG GAA GAT GAG CAC GGA A and 5'-ATC CTC TAG AAG GGC TGA AAT GTC ATC AGT. Cloning, sequencing, and mutagenesis by overlap-extension PCR was carried out essentially as described [Lei et al., 2005].

Transfection Experiments

Reporter constructs were transfected into 293 T and HeLa cells using siPORT *XP-1* (Ambion, Austin, TX) according to the manufacturer's recommendations. The cells were harvested after 24 hr as described [Kralovicova and Vorechovsky, 2007; Lei et al., 2005]. Total RNA was extracted using TRI Reagent (Ambion) according to manufacturer's recommendation and reverse transcribed with an oligo-d(T) primer and Moloney murine virus reverse transcriptase (Promega). The cDNA was amplified with vector-specific primers PL3 and PL4 as described [Kralovicova and Vorechovsky, 2007; Lei et al., 2005]. PCR products were separated on agarose and polyacrylamide gels and sequenced to confirm their identity.

For the siRNA experiments, 12-well plates were seeded with 10⁵ of 293 T cells 24 hr prior to transfection. Each well was transfected with 100 (PTB) or 150 (U2AF65, PUF60, and control siRNAs) pmol of duplex siRNAs using the HiPerFect reagent (Qiagen) as recommended by the manufacturer. On day 3, the cells were trypsinized and transferred to a six-well plate. On day 4, the cells were retransfected with siRNAs as on day 2. The following day, the cells were transfected with 750 ng of reporter plasmid using 2 μ l of siPORT-*XP* and harvested on day 6 or 7 for RNA isolation and Western blotting. siRNA sequences were r(GCA GAU GAA CUC

GGU GAU G)dTdT (PUF60), r(GCA AGU ACG GGC UUG UCA A)dTdT (U2AF65) [Hastings et al., 2007], r(AAC UUC CAU CAU UCC AGA GAA)dTdT (PTB) [Wollerton et al., 2004]; r(AGG UAG UGU AAU CGC CUU G)dTdT (47%) and r(UGC GCU AGG CCU CGG UUG C)dTdT (68%). All siRNAs were synthesized by the MWG Biotech (Germany). Plasmids expressing SR proteins and other factors were described previously [Kralovicova et al., 2004]. Coexpression of SR proteins and exogenous reporters was performed essentially as described [Kralovicova et al., 2004].

Mutation Nomenclature

Mutation designation is according to the official mutation nomenclature (<http://www.hgvs.org/mutnomen/>). Reference accession numbers for normal alleles: L76259.1 for human genomic DNA; M97655.1 for human cDNA (cDNA numbering starts with 1 at A at the ATG start codon).

Results

Intronic Deletion of the *Alu* poly(T)-Tail Improves 3' Splice Site Organization of a Disease-Causing Pseudoexon

Both *PTS* pseudoexons were identified in patients with an autosomal recessive form of hyperphenylalaninaemia with concomitant monoamine neurotransmitter deficiency (Table 1). In the first case (patient MD130), we found a reduced amount of correctly spliced transcripts and a larger cDNA fragment containing an extra 45-nucleotide (nt) of the open reading frame contributed by intron 2 sequences (Fig. 1A). The ratio of correctly spliced and aberrant fragments was not altered in cycloheximide-treated cells, indicating that the mutant transcript was not subject to degradation by nonsense-mediated mRNA decay (data not shown). The aberrant RNA product identified in the patient was, however, accompanied by a lack of PTPS on immunoblots (Fig. 1B). Nucleotide sequencing of flanking introns showed that the extra exon was flanked by canonical splice sites and revealed a maternal deletion located 20 nt upstream of the pseudoexon (Fig. 1C–E). Analysis of repetitive and splicing consensus sequences showed that the genomic deletion removed the poly(T)-tail from the right arm of antisense *Alu*Sq and introduced a favorable branch point sequence(s) in the optimal distance from the pseudoexon 3' splice site (Fig. 1D–E). Apart from the *Alu*Sq sequence, the mutation deleted a short upstream segment containing AG dinucleotides that may repress the pseudoexon 3' splice site on normal alleles (reviewed in [Reed, 1989]).

To test this prediction, to identify sequence motifs important for exonization of this *Alu* and to confirm that the mutation is disease-causing, we prepared wild-type (WT) and mutated splicing reporter constructs and analyzed their splicing pattern following transient transfection into embryonal kidney 293 T and HeLa cells (Figs. 2 and 3). In addition to correctly spliced transcripts, the WT constructs generated RNA products lacking 23-nt exon 3 (Fig. 3, lane 2), consistent with its poor inclusion in the *PTS* mRNA, as observed previously for some cell types [Leitner et al., 2003]. Moreover, ~5% of the WT mRNAs were spliced to a proximal cryptic 3' splice site between *Alu*Sq positions 281 and 280 (Fig. 1E and Fig. 2) that corresponded to 3' splice sites used by ~23% of exonized *Alus* [Lev-Maor et al., 2003; Sela et al., 2007]. In contrast to the WT minigene, reporter constructs with the deletion produced a larger transcript with the pseudoexon and no correctly spliced mRNAs, recapitulating the aberrant RNA product found in the heterozygous patient (Fig. 3, lane 3).

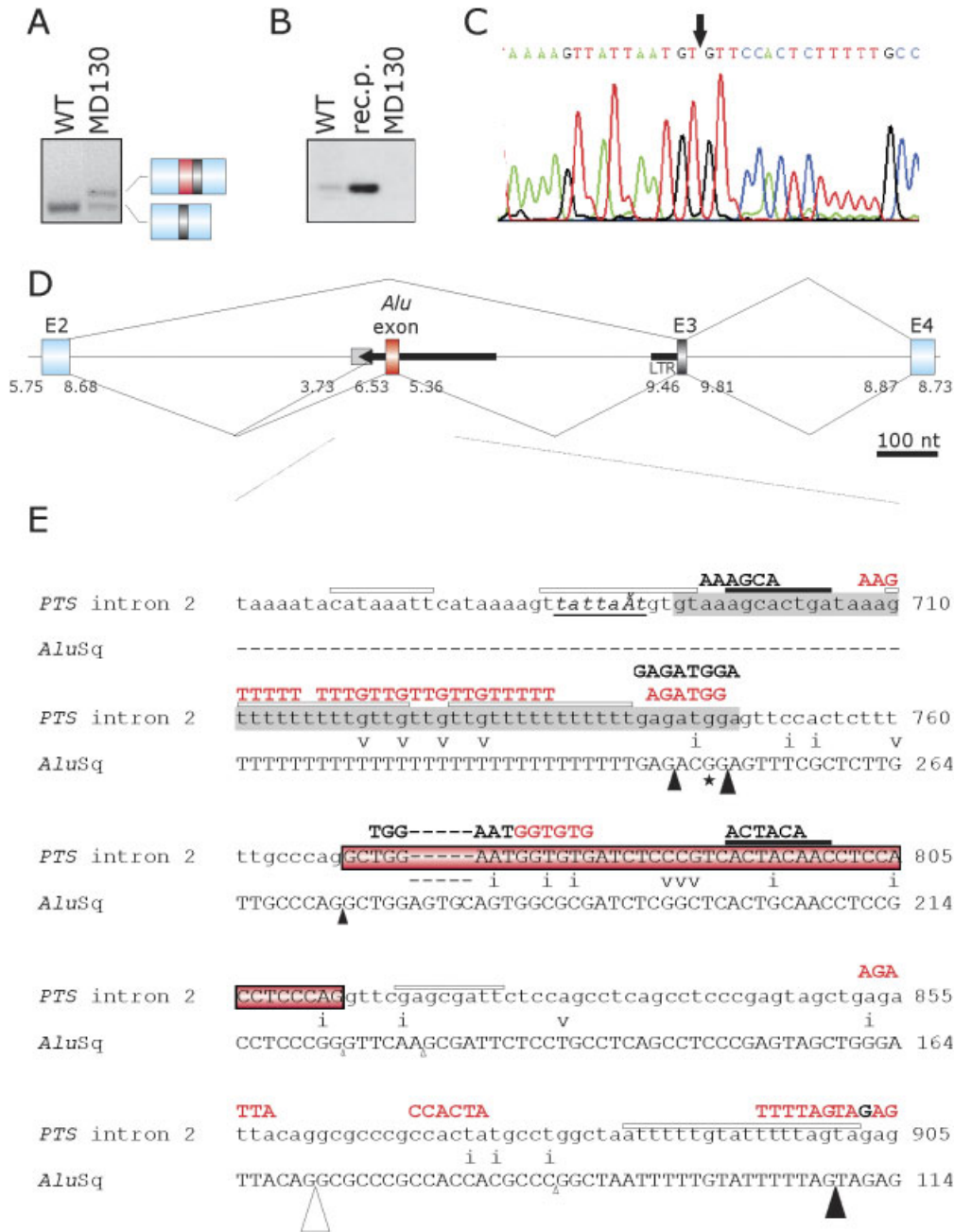


Figure 1. Activation of antisense *AluSx* exon upon deletion of the poly(T)-tail. **(A)** Reverse-transcription (RT)-PCR analysis with *PTS* cDNAs prepared from dermal fibroblasts derived from patient MD130 and an unaffected control (WT). The identity of RNA products is shown on the right. *Alu* pseudoexon is denoted by a red box, the remaining exons are in blue or black as in panel D. Control PCRs without cDNA or reverse transcriptase performed in parallel did not produce any detectable signal. **(B)** Immunodetection of PTPS in primary dermal fibroblasts derived from patient MD130 and a control. Rec.p., recombinant protein as a positive control. **(C)** Nucleotide sequence across the deletion breakpoint (arrow). The deletion was not found by screening 100 control DNA samples of blood donors. **(D)** Schematics of the splicing pattern/reporter. *PTS* exons are shown as boxes, introns as lines. Genomic deletion is denoted by a gray rectangle and antisense *AluSx* by a black arrow. The splicing pattern of the WT and mutated allele is schematically shown above and below the primary transcript, respectively. The maximum entropy (ME) scores are shown below each splice site. **(E)** Auxiliary splicing sequences in *PTS* intron 2. Intronic sequences are in lower case. Genomic deletion is highlighted in gray. The cryptic exon, which corresponds to *Alu* consensus positions 206–255, is in upper case in a red box. The predicted branch point sequence on the deletion allele is in italics and underlined, with the branch point denoted by Ä. Antisense *AluSx* consensus is aligned below *PTS* intron 2. The alignment was performed with RepeatMasker; mismatched nucleotides are shown as transitions (i) or transversions (v). Nucleotide positions of intron 2 (*upper sequence*; reference human genome assembly of chromosome 11; NT_033899.7) and of the *Alu* consensus (*lower sequence*) is shown on the right. Predicted auxiliary splicing sequences are shown above intron 2 in bold capitals. RESCUE-ESEs [Fairbrother et al., 2002] are in black, FAS-ESEs (hex2 set) [Wang et al., 2004] are in red, PESSs [Zhang and Chasin, 2004] are denoted by open rectangles and PESEs [Zhang and Chasin, 2004] by closed rectangles above the *PTS* sequence. The 3' splice site (closed arrows) and 5' splice site (open arrows) of naturally exonized *Alus* [Sela et al., 2007] are shown below the *AluSx* consensus. The size of each arrow reflects the number of natural *Alu* exons, as ascertained previously [Sela et al., 2007]. Star denotes guanosine in *AluSx* sequences compared to adenosine in younger *Alus*, creating important 3'AG of exonized elements [Lev-Maor et al., 2003].



Figure 2. Nucleotide sequences of mutated splicing reporter constructs. Genomic deletion in patient MD130 is highlighted in gray and the *Alu* pseudoexon is boxed. Mutations introduced in plasmid constructs are underlined in red. The plasmid DNA template used for mutagenesis is shown in parentheses. WT, plasmid representing the wild-type allele; Δ , plasmid with the disease-causing genomic deletion. K in clones M19 and M20 denotes the IUB code for T or G. Arrows show proximal and distal 3' splice sites in exonized *Alus* [Lev-Maor et al., 2003; Sorek et al., 2004]. $\Delta 3'$ AG denotes 3' splice site that was activated in the Δ construct. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

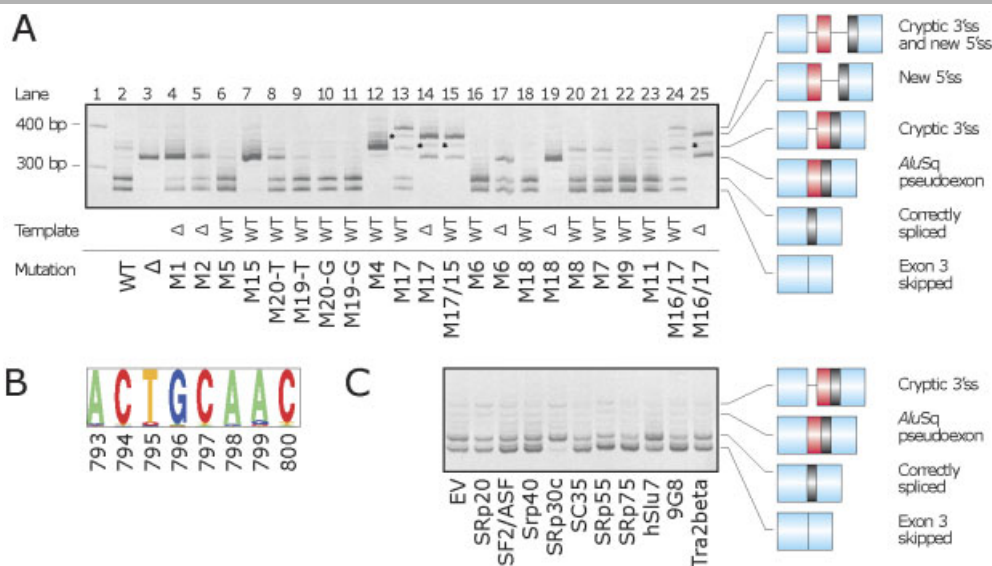


Figure 3. Sequence determinants of the *AluSq* pseudoexon activation. **(A)** The splicing pattern of WT and mutated reporter plasmids following transfection into 293 T cells. Spliced products are shown schematically on the right, with *Alu* exon in red and exon 3 in black. The identity of all fragments was confirmed by sequencing. Heteroduplexes are denoted by stars. Mutations are shown in Figure 2. HeLa cells exhibited a similar exogenous splicing pattern (data not shown). **(B)** Alignment of a predicted AC-rich splicing regulatory signal in *AluSq* sequences. Numbers indicate *Alu* positions and correspond to the sequence shown in Figure 1E. **(C)** Splicing pattern of the WT reporter construct in cells overexpressing SR proteins and human *SLU7*. EV, empty vector as a control. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Progressive mutations of predicted branch point adenosines in the deletion constructs increased natural mRNAs and reduced the pseudoexon inclusion, but did not eliminate it completely, suggesting that alternative, multiple or noncanonical branch points were used (lanes 4 and 5). A transition improving a putative branch site in the WT construct at a similar distance from pseudoexon 3' splice site as on the deletion allele did not noticeably enhance the pseudoexon inclusion (lane 6), implicating other factors in the 3' splice site repression or another branch site further upstream. Indeed, optimization of the upstream branch site to a heptamer TTCTGAT coupled with elimination of an AG dinucleotide upstream of pseudoexon 3' splice site fully activated

the *Alu* exon in the WT construct, mimicking the splicing pattern of the deletion reporter (lane 7). Replacement of the adenosine in the first upstream AG with uracil, which generally improves the polypyrimidine tract [Bouck et al., 1995; Roscigno et al., 1993], was sufficient for the pseudoexon inclusion (lanes 8 and 9), but not with guanosines (Gs) (lanes 10 and 11). This result is consistent with a newly created G-quartet acting as a repressor of the proximal 3' splice site and the pseudoexon. Finally, mutation optimizing a proximal 3'AG (GAG to CAG) in the WT clone led to exclusive use of this site and elimination of correctly spliced products, providing experimental evidence that the constitutive inclusion of *AluSq* can be achieved by a single mutation (lane 12).

Taken together, despite removing the poly(T)-tail, the intronic deletion exonized the *PTS AluS_q* by creating an AG exclusion zone upstream of the 3' pseudoexon splice site and downstream of the new branch site(s). The new branch site per se was sufficient for alternative splicing, but not for constitutive splicing of the *Alu*-derived exon (lanes 4 and 5).

Determinants of the Pseudoexon 5' Splice Site Activation

The pseudoexon 5' splice site was between *AluS_q* positions 205 and 206 (Fig. 1E), the same *Alu* positions as 5' splice sites of disease-causing *Alu* pseudoexons activated by mutations in *COL4A3* intron 5 [Knebelmann et al., 1995] and *GUSB* intron 8 [Vervoort et al., 1998]. In contrast, most natural *Alu* exons were exonized through 5' splice sites between *Alu* positions 157 and 158 [Sorek et al., 2004]. The corresponding decoy 5' splice site in *PTS* is intrinsically weaker than the pseudoexon 5' splice site (Supp. Table S1), because it contains cytosines (C) at *Alu* positions 156 and 153 (intron positions +2 and +5). We inspected sequences of natural *Alu* exons [Sorek et al., 2004], but found only *Alus* with a C either at position 156 or 153, whereas none of them had Cs at both positions (data not shown), suggesting that C+2 and C+5 at the *AluS_q* decoy 5' splice site consensus were key nucleotides that prevented its activation. In agreement with this prediction, a C>T mutation at position +2 was sufficient for substantial use of this 5' splice site and *Alu* exon inclusion, both in the WT and deletion constructs (lanes 13 and 14), and when accompanied by deletion of AGs in the WT (lane 15).

Contribution of Auxiliary Splicing Sequences to the Inclusion of the *AluS_q* Pseudoexon in mRNA

Mutation-induced pseudoexons have higher densities of splicing enhancers (ESEs) than average introns and lower densities of splicing silencers (ESSs) compared to constitutive exons, contributing to pseudoexon activation in vivo [Kralovicova and Vorechovsky, 2007]. To experimentally test whether this applies to a disease-causing *AluS_q* exon, we introduced point mutations in predicted auxiliary elements in both the WT and deletion minigenes. Figure 3 (lanes 16–25) shows that most mutations that eliminated predicted ESEs/ESSs had anticipated effects. Although abolition of a predicted ESE by an A>G mutation reduced pseudoexon inclusion in the mRNA (lanes 16 and 17), a 5-nt insertion eliminating the same element was exonization-neutral (lanes 18 and 19). This pentamer was absent in human and chimpanzee *PTS*, but present in orangutan and Rhesus monkey homologs, suggesting that this deletion occurred ~6–14 million years ago (Fig. 1E, and data not shown). Multiple alignment of 100 randomly selected *AluS_q* sequences revealed the absence of this pentamer at two loci on different chromosomes (317NT_033898, 2375NT_017568). Mutations eliminating additional regulatory elements altered pseudoexon inclusion to a minor extent (lanes 20–22). A mutation in clone M8, which emulates the *AluS_q* consensus (Fig. 3B), removes a predicted PESE and RESCUE-ESE and alters an AC-rich exonic regulatory sequence TACAACCTC [Goren et al., 2006; Ladd and Cooper, 2002], increased utilization of proximal 3' splice site (lane 20). However, a double mutation eliminating the former two elements reduced use of this site (lane 22). A mutation creating a stretch of overlapping FAS-ESSs between cryptic and pseudoexon 3' ss had virtually no effect (lane 23). Finally, a mutation in a predicted ESS at *Alu* positions 853–858, which creates an *AluS_q* consensus and a G triplet, inhibited intron-proximal 5' splice site (cf. lanes 24, 25,

and 13). Thus, most mutations in predicted auxiliary splicing sequences influenced *Alu* inclusion levels.

Splicing of Weak *PTS* Exon 3 Is Controlled by SR Proteins

Exonic splicing regulatory sequences interact with *trans*-acting factors, most notably with highly conserved serine/arginine-rich (SR) proteins that may positively or negatively influence splicing [Graveley et al., 1999; Manley and Tacke, 1996]. In addition to the RNA-recognition domain, which binds short elements in the pre-mRNAs, these SR proteins contain conserved RS domains involved in additional interactions. We examined the inclusion of the *PTS AluS_q* pseudoexon in cells overexpressing a series of SR or SR-like proteins and *SLU7*, a factor that regulates 3' splice site selection [Chua and Reed, 1999] (Fig. 3C). Although we observed no or only minor changes in the proportion of transcripts containing the *AluS_q* pseudoexon, the inclusion of preexisting exon 3 was dramatically affected by a subset of SR proteins, with SRp30c strongly inhibiting and most of the remaining SR proteins promoting exon skipping. This result suggests that the regulatory role of SR proteins is particularly important for short or weakly spliced exons (Fig. 1D), in agreement with similar observations for suboptimal splice sites, even in unicellular organisms [Ram and Ast, 2007; Webb et al., 2005]. Interestingly, multiple alignment of *PTS* exon 3 and flanking intronic sequences in mammals revealed absolute conservation of a strong branch site in a long terminal repeat (LTR; Fig. 1D and Supp. Fig. S1). In contrast, both 5' and 3' splice sites of this exon, the upstream polypyrimidine tract and the exon itself sustained numerous nucleotide substitutions during evolution, raising a hypothesis that the conserved branch site is important for the observed SR-mediated sensitivity to exon inclusion and that such branch sites provide compensatory help to weakly included exons, as reported for other splicing signals [Ke et al., 2008]. A direct interaction between an ESE-bound RS domain of ASF/SF2 and a branch site was previously proposed to enhance splicing by stabilizing base-pairing between small nuclear RNAs and primary transcripts [Shen and Green, 2006; Shen et al., 2004]. Finally, ASF/SF2 and SC35 did not exhibit opposite effects on the inclusion of exon 3 (Fig. 3C), as was reported previously for some but not other exons [Expert-Bezancon et al., 2004; Gallego et al., 1997; Kralovicova et al., 2004; Solis et al., 2008], consistent with a unique assembly of cooperative interactions that define each coding unit.

Pseudoexon Activation in a LINE-2 Repeat Leading to BH₄ Deficiency

The second patient had a milder form of BH₄ deficiency (MD96, Table 1). Our cDNA-based mutation analysis revealed an extra 79-nt fragment between *PTS* exon 1 and 2, with residual levels of correctly spliced transcripts in dermal fibroblasts (Fig. 4A). The ratio of aberrant and correctly spliced transcripts in cycloheximide-treated cells was increased approximately two- to threefold compared to untreated cells (data not shown), consistent with degradation of aberrant transcripts with premature termination codons by nonsense-mediated mRNA decay. Intron sequencing showed that the pseudoexon was flanked by canonical splice-site consensus sequences and revealed a single homozygous A>T transversion at position –9 relative to the pseudoexon 3' splice site (Fig. 4C). Analysis of repetitive sequences with RepeatMasker/RepBase showed that this mutation was located at the 3' end of the LINE-2A subfamily (Fig. 4C–D). Multiple alignment of exonized LINE-2 elements revealed that sequences corresponding to the

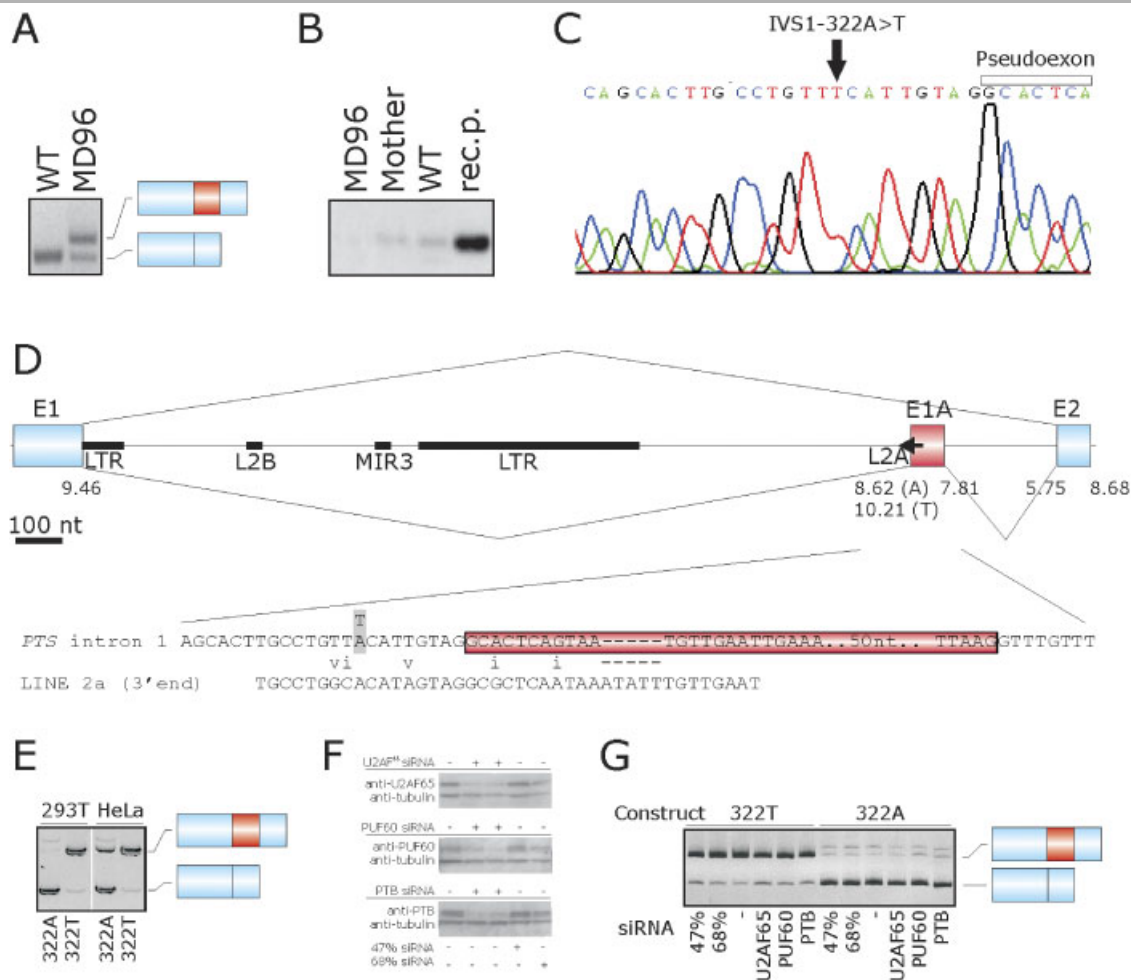


Figure 4. An intronic transition improves the polypyrimidine tract of a LINE-2 pseudoexon and results in BH₄ deficiency. **(A)** RT-PCR analysis of RNA samples extracted from dermal fibroblasts of patient MD96 and an unaffected control (WT). **(B)** Western blot analysis of lysates obtained from the patient, mother, and WT. Rec.p., signal from a recombinant protein as a positive control. **(C)** Identification of a homozygous point mutation in *PTS* intron 1. **(D)** Schematic representation of the splicing pattern/reporter (for legend, see Fig. 1D). LINE-2 (L2) repeat is shown by a horizontal arrow, additional repeats in this intron are denoted by thick lines. LTR, long terminal repeat. Lower panel shows the alignment of the 3' end of the LINE-2 A consensus with *PTS* intron 1. Pseudoexon sequence is highlighted in a red box; mutation is in grey box. **(E)** Splicing pattern of allele-specific minigenes in two cell lines. **(F)** siRNA-mediated depletion of putative *trans*-factors interacting with the homozygous transition (IVS1-322A>T) in 293T cells, as determined by Western blot analysis. **(G)** Pseudoexon inclusion in cells depleted for U2AF⁶⁵, PUF60, and PTB. 47%, 68%; nonspecific siRNA controls with the indicated CG contents. RT-PCR products are shown on the right.

pseudoexon 3' splice site were present in about a third of LINE-2 exons (Supp. Fig. S2).

To validate that this mutation resulted in abnormal RNA products, and to study requirements for 3' splice site recognition in the LINE-2 pseudoexon, we constructed splicing reporters containing exons 1 and 2 (Fig. 4D). The A>T mutation in the WT minigene recapitulated the pseudoexon activation upon transient transfection into both 293T and HeLa cells (Fig. 4E). In contrast to 293T cells, however, we observed a substantially higher inclusion of the LINE-2 exon in mRNAs transcribed from the WT clone in HeLa cells, which express lower amounts of PTPS than kidney cells, consistent with a differential expression of regulator(s) that control inclusion of the *PTS* pseudoexon.

In an attempt to identify interactions affected by the A>T mutation, we examined the pseudoexon inclusion levels in cells lacking proteins that bind to the polypyrimidine tract or polypyrimidine-rich RNA sequences, including a large subunit of auxiliary factor of U2 snRNP (U2AF⁶⁵) [Zamore and Green, 1989], PUF60 [Page-McCaw et al., 1999], and PTB [Garcia-Blanco et al., 1989]. However, substantial siRNA-mediated downregula-

tion of these factors in 293T cells, as documented by Western blot analysis (Fig. 4F), failed to alter relative ratios of transcripts with and without the pseudoexon (Fig. 4G).

Discussion

We have shown new examples of disease-causing pseudoexons activated in human repetitive sequences and defined requirements for their inclusion in mature transcripts at the nucleotide level. Repetitive elements constitute about a half of the human genome [Lander et al., 2001], but as much as 90% of recently acquired exons, suggesting that repeats are more likely to be exonized than random sequences [Sela et al., 2007; Sorek, 2007]. However, this increase is largely attributable to primate-specific *Alu*, with over 0.2% of their intronic copies present in human expressed sequence tags (ESTs) [Sela et al., 2007]. The majority if not all *Alu* exons are alternatively spliced [Sorek et al., 2002], but disease-causing *Alu*-derived exons are usually constitutively spliced, producing less natural transcripts and resulting in more profound phenotypic consequences. Unlike the LINE2 exon (Fig. 4), the *Alu* exon did not introduce any

premature termination codons (Fig. 1E) and the relative abundance of pseudoexon-containing transcripts was unaltered in cycloheximide-treated cells, suggesting that they were translated *in vivo*. Based on the conserved structure of PTPS (reviewed in [Thony et al., 2000]), the insertion of extra 15 amino acids (GWNGVISR-HYNLHLP) between codons 54 and 55 of the mutant PTPS would be predicted to destroy the second β -sheet, abrogate zinc binding and severely diminish folding stability. Although the deleterious nature of disease-causing, constitutively spliced *Alu* sequences may explain the overall rarity of such cases, a bias of current mutation screening strategies toward DNA rather than RNA and toward coding sequences is expected to play a major role.

In contrast to *Alus*, there was no apparent overrepresentation of LINE-2 among other exonized repetitive elements, both in human and mouse EST datasets [Sela et al., 2007]. Of all transposed elements, LINE-2 sequences exhibit the greatest divergence and are considered the most ancient repeats [Lander et al., 2001; Sela et al., 2007; Smit, 1996]. LINE-2 show sequence similarities to the 3' end of mammalian interspersed repeats (MIRs), which represent ~2% of the human genome mass [Lander et al., 2001; Smit, 1996]. Only about a hundred of LINE-2 have been identified in human exons, with some functionally contributing putative phosphorylation and myristoylation sites [Sela et al., 2007]. Compared to the MIR consensus, LINE-2 repeats have an A>G mutation at position -2 relative to the 3' splice site of *PTS* pseudoexon (Fig. 4D, and data not shown), thus creating a common 3' splice site signal TAG/G. This signal per se is unlikely to be sufficient for exonization, but the exonized *PTS* LINE-2 had additional three uracils in the polypyrimidine tract compared to the LINE-2 consensus, which was probably a prerequisite for significant activation of the pseudoexon 3' splice site (Fig. 4D). The ratio of transcripts with and without LINE-2 was not altered in cells individually depleted for three polypyrimidine-binding proteins, suggesting that their residual levels were still sufficient for efficient use of the 3' splice site and/or other RNA-interacting factors take part in the improved recognition (Fig. 4G).

In contrast to LINE elements, *Alus* and other SINES were exonized more often in antisense than sense orientation. This has been attributed to their poly(T)-tails and a higher number of splice site-like signals in antisense sequences [Lev-Maor et al., 2003; Makalowski et al., 1994; Sela et al., 2007; Sorek, 2007; Sorek et al., 2002]. Poly(T)-tails served as the polypyrimidine tracts not only for recent *Alu* exons [Lev-Maor et al., 2003; Sorek, 2007], but also for exons derived from other vertebrate repeats [Krull et al., 2007; Wang and Kirkness, 2005]. Most exonized antisense *Alu* employed 3' splice sites at positions 277/281 and 118/119 [Lev-Maor et al., 2003; Makalowski et al., 1994; Sela et al., 2007] (Fig. 1E). By contrast, the 3' splice site of *PTS Alu* exon was at position 256/255, which was used just by over a hundred of human *Alu* exons [Sela et al., 2007]. The 5' splice site of the *PTS* pseudoexon was employed by 16 *Alu* exons [Sela et al., 2007]. The 3' splice site of the *PTS* pseudoexon is comparable to the consensus sequence of exonized *Alus*, but its 5' splice site underwent an A>G transition at intron position +5 (*Alu* position 201; Fig. 1E). Guanosine at this position is a key nucleotide for cryptic 5' splice site activation [Buratti et al., 2007a], and was likely to be essential both for the constitutive splicing of the *AluSq* pseudoexon and the resulting severe BH₄ deficiency.

Alu exonization has been shown to require only one [Lev-Maor et al., 2003; Sorek et al., 2004] (Fig. 3A) or only a few [Lei et al., 2005; Singer et al., 2004] mutations. The new type of exonization described here was caused by a single deletion that removed intronic elements promoting (poly(T)-tail and branch site) or

repressing (AG dinucleotide) splice site selection and created a more favorable 3' splice site organization. A key event was the removal of AG dinucleotide located 6 nt downstream of the most likely WT branch site TTTTGAG (Fig. 1E), which was also the strongest branch site predicted computationally (data not shown). This distance was sufficient for significant repression of downstream 3' splice sites in some but not all pre-mRNAs tested [Kralovicova et al., 2005; Lei and Vorechovsky, 2005]. Shortening this distance in the WT even by a few nucleotides would be predicted to result in substantial pseudoexon inclusion [Kralovicova et al., 2005; Lei and Vorechovsky, 2005]. Activation of the *AluSq* pseudoexon was further facilitated by juxtaposition of a new branch site within the proximity of the pseudoexon 3' splice site, highlighting the importance of sterical constraints for assembly of splicing complexes at genuine 3' splice sites. Our results expand the range of disease-causing exonization events and suggest that the repeat and copy number variability, which is prevalent in vertebrate genomes, may significantly interfere with accurate pre-mRNA processing and modify exon-intron structure. It also suggests that evolutionary pathways leading to exonization of intronic sequences could be more diverse than previously thought. Primate specificity of *Alus* and their high representation among recent cassette exons [Sela et al., 2007; Sorek et al., 2002; Zhang and Chasin, 2006] make these sequences strong candidates for ensuring extraordinary versatility of primate genomes and a high organismal complexity of the *Hominidae* family.

Finally, the observed correspondence between the levels of natural *PTS* transcripts and laboratory measures of disease severity at diagnosis (Figs. 1A, 2A and Table 1) reiterates the importance of alternative splicing as a global contributor to variable disease expressivity. Although we could not exclude that residual PTPS activity was influenced by other factors, including posttranslational modifications [Oppliger et al., 1995], our findings point to a crucial and underappreciated role of this process in phenotypic variability. This applies not only to recessive diseases, exemplified here by BH₄ deficiency, but also to complex traits where intronic variants have increasingly been implicated in genetic susceptibility [Sladek et al., 2007]. Because alternative splicing is influenced by a combinatorial control of cellular factors that constitute the "splicing code," including activity and expression levels of numerous splicing regulatory proteins and secondary RNA structure [Buratti et al., 2007b; Matlin et al., 2005], future studies should aim at dissecting the complicated network of these interactions in well-defined cases of mutation-induced aberrant splicing.

Acknowledgment

We thank A. Krainer and C. Smith for antibodies, and J. Neidhardt and A. Lauber-Biason for valuable support and discussion. This work was supported by grants from the Swiss National Foundation (310000-107500) to N.B. and B.T., and the Juvenile Diabetes Research Foundation International (1-2008-047) to I.V.

References

- Bouck J, Fu XD, Skalka AM, Katz RA. 1995. Genetic selection for balanced retroviral splicing: novel regulation involving the second step can be mediated by transitions in the polypyrimidine tract. *Mol Cell Biol* 15:2663-2671.
- Buratti E, Baralle M, Baralle FE. 2006. Defective splicing, disease and therapy: searching for master checkpoints in exon definition. *Nucleic Acids Res* 34:3494-3510.
- Buratti E, Chivers MC, Kralovicova J, Romano M, Baralle M, Krainer AR, Vorechovsky I. 2007a. Aberrant 5' splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic Acids Res* 35:4250-4263.

- Buratti E, Dhir A, Lewandowska E, Baralle FE. 2007b. RNA structure is a key regulatory element in pathological *ATM* and *CFTR* pseudoexon inclusion. *Nucleic Acids Res* 35:4369–4383.
- Chua K, Reed R. 1999. The RNA splicing factor hSlu7 is required for correct 3' splice-site choice. *Nature* 402:207–210.
- Expert-Bezancon A, Sureau A, Drosay P, Saless R, Groeneveld H, Lecaer JP, Marie J. 2004. hnRNP A1 and the SR proteins ASF/SF2 and SC35 have antagonistic functions in splicing of beta-tropomyosin exon 6B. *J Biol Chem* 279:38249–38259.
- Fairbrother WG, Yeh RF, Sharp PA, Burge CB. 2002. Predictive identification of exonic splicing enhancers in human genes. *Science* 297:1007–1013.
- Gallego ME, Gattoni R, Stevenin J, Marie J, Expert-Bezancon A. 1997. The SR splicing factors ASF/SF2 and SC35 have antagonistic effects on intronic enhancer-dependent splicing of the beta-tropomyosin alternative exon 6A. *EMBO J* 16:1772–1784.
- Garcia-Blanco MA, Jamison SF, Sharp PA. 1989. Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. *Genes Dev* 3:1874–1886.
- Goren A, Ram O, Amit M, Keren H, Lev-Maor G, Vig I, Pupko T, Ast G. 2006. Comparative analysis identifies exonic splicing regulatory sequences—The complex definition of enhancers and silencers. *Mol Cell* 22:769–781.
- Graveley BR, Hertel KJ, Maniatis T. 1999. SR proteins are “locators” of the RNA splicing machinery. *Curr Biol* 9:R6–R7.
- Hastings ML, Allemand E, Duelli DM, Myers MP, Krainer AR. 2007. Control of pre-mRNA splicing by the general splicing factors PUF60 and U2AF. *PLoS ONE* 2:e538.
- Ke S, Zhang XH, Chasin LA. 2008. Positive selection acting on splicing motifs reflects compensatory evolution. *Genome Res* 18:533–543.
- Knebelmann B, Forestier L, Drouot L, Quinones S, Chuet C, Benessy F, Saus J, Antignac C. 1995. Splice-mediated insertion of an Alu sequence in the *COL4A3* mRNA causing autosomal recessive Alport syndrome. *Hum Mol Genet* 4:675–679.
- Kralovicova J, Christensen MB, Vorechovsky I. 2005. Biased exon/intron distribution of cryptic and *de novo* 3' splice sites. *Nucleic Acids Res* 33:4882–4898.
- Kralovicova J, Houngrinou-Molango S, Kramer A, Vorechovsky I. 2004. Branch sites haplotypes that control alternative splicing. *Hum Mol Genet* 13:3189–3202.
- Kralovicova J, Vorechovsky I. 2007. Global control of aberrant splice site activation by auxiliary splicing sequences: evidence for a gradient in exon and intron definition. *Nucleic Acids Res* 35:6399–6413.
- Krull M, Brosius J, Schmitz J. 2005. Alu-SINE exonization: en route to protein-coding function. *Mol Biol Evol* 22:1702–1711.
- Krull M, Petrusma M, Makalowski W, Brosius J, Schmitz J. 2007. Functional persistence of exonized mammalian-wide interspersed repeat elements (MIRs). *Genome Res* 17:1139–1145.
- Ladd AN, Cooper TA. 2002. Finding signals that regulate alternative splicing in the post-genomic era. *Genome Biol* 3:reviews0008.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W and others. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
- Lei H, Day INM, Vorechovsky I. 2005. Exonization of AluYa5 in the human *ACE* gene requires mutations in both 3' and 5' splice sites and is facilitated by a conserved splicing enhancer. *Nucleic Acids Res* 33:3897–3906.
- Lei H, Vorechovsky I. 2005. Identification of splicing silencers and enhancers in sense *Alus*: a role for pseudo-acceptors in splice site repression. *Mol Cell Biol* 25:6912–6920.
- Leitner KL, Meyer M, Leimbacher W, Peterbauer A, Hofer S, Heufler C, Muller A, Heller R, Werner ER, Thony B, Werner-Felmayer G. 2003. Low tetrahydrobiopterin biosynthetic capacity of human monocytes is caused by exon skipping in 6-pyruvoyl tetrahydropterin synthase. *Biochem J* 373:681–688.
- Lev-Maor G, Sorek R, Shomron N, Ast G. 2003. The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. *Science* 300:1288–1291.
- Makalowski W, Mitchell GA, Labuda D. 1994. Alu sequences in the coding regions of mRNA: a source of protein variability. *Trends Genet* 10:188–193.
- Manley JL, Tacke R. 1996. SR proteins and splicing control. *Genes Dev* 10:1569–1579.
- Matlin AJ, Clark F, Smith CW. 2005. Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* 6:386–398.
- Mitchell GA, Labuda D, Fontaine G, Saudubray JM, Bonnefont JP, Lyonnet S, Brody LC, Steel G, Obie C, Valle D. 1991. Splice-mediated insertion of an Alu sequence inactivates ornithine delta-aminotransferase: a role for Alu elements in human mutation. *Proc Natl Acad Sci USA* 88:815–819.
- Oppliger T, Thony B, Nar H, Burgisser D, Huber R, Heizmann CW, Blau N. 1995. Structural and functional consequences of mutations in 6-pyruvoyltetrahydropterin synthase causing hyperphenylalaninemia in humans. Phosphorylation is a requirement for in vivo activity. *J Biol Chem* 270:29498–29506.
- Page-McCaw PS, Amonlirdviman K, Sharp PA. 1999. PUF60: a novel U2AF65-related splicing activity. *RNA* 5:1548–1560.
- Ram O, Ast G. 2007. SR proteins: a foot on the exon before the transition from intron to exon definition. *Trends Genet* 23:5–7.
- Reed R. 1989. The organization of 3' splice-site sequences in mammalian introns. *Genes Dev* 3:2113–2123.
- Roscigno RF, Weiner M, Garcia-Blanco MA. 1993. A mutational analysis of the polypyrimidine tract of introns. Effects of sequence differences in pyrimidine tracts on splicing. *J Biol Chem* 268:11222–11229.
- Sela N, Mersch B, Gal-Mark N, Lev-Maor G, Hotz-Wagenblatt A, Ast G. 2007. Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. *Genome Biol* 8:R127.
- Shen H, Green MR. 2006. RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. *Genes Dev* 20:1755–1765.
- Shen H, Kan JL, Green MR. 2004. Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote presplicingosome assembly. *Mol Cell* 13:367–376.
- Singer SS, Mannel DN, Hehlhans T, Brosius J, Schmitz J. 2004. From “junk” to gene: curriculum vitae of a primate receptor isoform gene. *J Mol Biol* 341:883–886.
- Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S and others. 2007. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445:881–885.
- Smit AF. 1996. The origin of interspersed repeats in the human genome. *Curr Opin Genet Dev* 6:743–748.
- Solis AS, Peng R, Crawford JB, Phillips 3rd JA, Patton JG. 2008. Growth hormone deficiency and splicing fidelity: two serine/arginine-rich proteins, ASF/SF2 and SC35, act antagonistically. *J Biol Chem* 283:23619–23626.
- Sorek R. 2007. The birth of new exons: mechanisms and evolutionary consequences. *RNA* 13:1–6.
- Sorek R, Ast G, Graur D. 2002. Alu-containing exons are alternatively spliced. *Genome Res* 12:1060–1067.
- Sorek R, Lev-Maor G, Reznik M, Dagan T, Belinky F, Graur D, Ast G. 2004. Minimal conditions for exonization of intronic sequences: 5' splice site formation in Alu exons. *Mol Cell* 14:221–231.
- Thony B, Auerbach G, Blau N. 2000. Tetrahydrobiopterin biosynthesis, regeneration and functions. *Biochem J* 347:1–16.
- Thony B, Blau N. 2006. Mutations in the BH4-metabolizing genes GTP cyclohydrolase I, 6-pyruvoyl-tetrahydropterin synthase, sepiapterin reductase, carbinolamine-4a-dehydratase, and dihydropteridine reductase. *Hum Mutat* 27:870–878.
- Vervoort R, Gitzelmann R, Lissens W, Liebaers I. 1998. A mutation (IVS8+0.6kdelTC) creating a new donor splice site activates a cryptic exon in an Alu element in intron 8 of the human beta-glucuronidase gene. *Hum Genet* 103:686–693.
- Vorechovsky I. 2006. Aberrant 3' splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic Acids Res* 34:4630–4641.
- Wang W, Kirkness EF. 2005. Short interspersed elements (SINEs) are a major source of canine genomic diversity. *Genome Res* 15:1798–1808.
- Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. 2004. Systematic identification and analysis of exonic splicing silencers. *Cell* 119:831–845.
- Webb CJ, Romfo CM, van Heeckeren WJ, Wise JA. 2005. Exonic splicing enhancers in fission yeast: functional conservation demonstrates an early evolutionary origin. *Genes Dev* 19:242–254.
- Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA, Smith CW. 2004. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol Cell* 13:91–100.
- Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 11:377–394.
- Zamore PD, Green MR. 1989. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci USA* 86:9243–9247.
- Zhang XH, Chasin LA. 2004. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev* 18:1241–1250.
- Zhang XH, Chasin LA. 2006. Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. *Proc Natl Acad Sci USA* 103:13427–13432.