Biased exon/intron distribution of cryptic and *de novo* 3′ splice sites

Jana Královičová, Mikkel B. Christensen and Igor Vořechovský*

University of Southampton School of Medicine, Division of Human Genetics, Southampton SO16 6YD, UK

ABSTRACT

We compiled sequences of previously published aberrant 3′ splice sites (3′ss) that were generated by mutations in human disease genes. Cryptic 3′ss, defined here as those resulting from a mutation of the 3′YAG consensus, were more frequent in exons than in introns. They clustered in ~20 nt region adjacent to authentic 3′ss, suggesting that their underrepresentation in introns is due to a depletion of AG dinucleotides in the polypyrimidine tract (PPT). In contrast, most aberrant 3′ss that were induced by mutations outside the 3′YAG consensus (designated '*de novo*') were in introns. The activation of intronic *de novo* 3′ss was largely due to AG-creating mutations in the PPT. In contrast, exonic *de novo* 3′ss were more often induced by mutations improving the PPT, branchpoint sequence (BPS) or distant auxiliary signals, rather than by direct AG creation. The Shapiro–Senapathy matrix scores had a good prognostic value for cryptic, but not *de novo* 3′ss. Finally, AG-creating mutations in the PPT that produced aberrant 3′ss upstream of the predicted BPS *in vivo* shared a similar 'BPS-new AG' distance. Reduction of this distance and/or the strength of the new AG PPT in splicing reporter pre-mRNAs improved utilization of authentic 3′ss, suggesting that AG-creating mutations that are located closer to the BPS and are preceded by weaker PPT may result in less severe splicing defects.

INTRODUCTION

The production of mature RNA in eukaryotes requires an accurate removal of intervening sequences or introns by splicing. Splicing of precursor (pre-)mRNA is facilitated by a large complex of small nuclear ribonucleoprotein particles (U1, U2, U4/U6 and U5 snRNPs) and a number of non-snRNP proteins that assemble on primary transcripts in a step-wise manner [reviewed in (1)]. Assembly of spliceosomal complexes requires the presence of conserved recognition sequences in the pre-mRNA: 5′ splice site (5′ ss; consensus MAG/GURAGU; M is A or C; R is purine), 3′ splice site (3′ss; consensus YAG/R; Y is pyrimidine), branchpoint sequence (BPS; mammalian consensus YNCURAY) and polypyrimidine tract (PPT). In addition to these signals, efficient intron removal often entails auxiliary sequences that repress or activate splicing, termed splicing silencers or enhancers, which function as binding sites for numerous factors, such as serine/arginine-rich (SR) proteins (2–5). Alterations in any of these *cis*-elements by mutations may severely impair pre-mRNA splicing and gene expression.

Mutations that affect splicing have been shown to account for up to half of disease-causing gene alterations (6,7). Since longer proteins are more likely to be involved in genetic disorders than shorter proteins and disease genes have, on average, a longer coding sequence and a higher number of introns than genes not causing recognizable phenotypes, it was hypothesized that splicing mutations may represent the most frequent cause of hereditary disease (8). The most common outcome of mutations affecting splice sites is exon skipping, followed by cryptic splice site activation (9,10). Cryptic 5′ ss are more common than cryptic 3′ss (9), but their unequal prevalence has not been understood.

Cryptic 5′ ss have a similar frequency distribution in exons and introns (11). Because recognition of 3′ss involves additional conserved elements in the intron (PPT and BPS), the distribution of aberrant 3′ss in exons and introns would be expected to reflect a more complex sequence context of the 3′ss. However, since the initial analysis of single-nucleotide substitutions in splice junctions (10) and a survey of 15 cryptic 3′ss in 1994 (9), no reliable data have been available in the literature.

Here, we have compiled sequences of previously published aberrant 3′ss in human disease genes. Our analysis revealed a biased distribution of cryptic 3′ss generated by mutations in the 3′ YAG towards exons and *de novo* 3′ss towards introns. We propose that the former can be fully explained by a depletion of AG dinucleotides in the PPT, while the latter is due to a lack of pyrimidine stretches downstream of authentic 3′ss.

*To whom correspondence should be addressed. Tel: +44 2380 796425; Fax: +44 2380 794264; Email: igvo@soton.ac.uk

In addition, we have investigated a group of disease-causing mutations that create AG dinucleotides in the PPT and activate aberrant 3′ss upstream of BPS. We found that they shared a similar distance between predicted BPS and newly introduced AGs and show that reduction of this distance and/or the strength of the new PPT enhanced the expression of natural transcripts. These results improve prediction of aberrant 3′ss localization in human disease genes and suggest that inspection of single-nucleotide substitutions near the 3′ss in their sequence context may facilitate prediction of their splicing outcome.

MATERIALS AND METHODS

Compilation of cryptic and *de novo* 3′ss

Published reports of cryptic and *de novo* 3′ss were identified by searching PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi), locus specific mutation databases (http://archive.uwcm.ac.uk/uwcm/mg/docs/oth_mut.html) and home pages of genetics journals. The search was restricted to human genes with sequence-verified aberrant RNA products published before May 2005 that resulted from disease-associated mutations or variants. In the majority of cases, these alterations were not found in DNA samples from unaffected individuals and/or showed co-segregation with affected family members in the pedigrees. The aberrant 3′ss were manually validated by mapping the information in the literature to sequences in the Human Genome project databases. Sequences of authentic, mutant and aberrant 3′ss together with the Shapiro–Senapathy (S&S) scores are available in Supplementary Tables 1 (exonic 3′ss) and 2 (intronic 3′ss). The S&S consensus matrix scores were computed using an algorithm described previously (12,13). To assess the significance of score values, the non-parametric Mann–Whitney rank test was employed as described previously (11). Sequences containing AG dinucleotides between predicted BPS and authentic 3′AG (termed 'intervening AGs') were extracted from a collection of 46 807 constitutively spliced human introns (14). The number of nucleotides that preceded and followed intervening AGs was computed using perl (http://www.activestate.com/Products/ActivePerl/, v. 5.6.1) scripts available on request.

Splicing reporter constructs, cell culture and transfections

Oligonucleotide primers for cloning *LIPC*, *HEXB*, *FBN2*, *TH* and *TSC2* reporter minigenes (pCR3.1; Invitrogen) are shown in Supplementary Table 3. Site-directed mutagenesis was carried out as described previously (15). All wild-type and mutated constructs were validated by sequencing as described previously (16). Transient transfections were performed in 12-well plates using FuGENE 6 (Roche). Human embryonic kidney 293T cells were grown under standard conditions in RPMI1640 supplemented with 10% (v/v) fetal calf serum (Gibco BRL). The plating density was $\sim 10^5$ cells per well 17–24 h before transfection. Medium was changed 2 h before adding a DNA mixture prepared by combining 1.5 µl of FuGENE and 50 µl of serum-free medium, followed by addition of 0.5 µg purified plasmid DNA. For co-transfections, 0.5 µg of reporter DNA was mixed with 1 µg of plasmids expressing SR proteins obtained as described previously (15). Cells were harvested 48 h post-transfection.

Detection of mRNA products

Total RNA was extracted as described previously (15), treated with DNase I (Ambion) and reverse transcribed using oligo(dT)$_{15}$ primers and Moloney murine virus reverse transcriptase (Promega) according to the manufacturer's recommendations. Three microlitres of each cDNA reaction together with negative and positive controls were amplified with vector-specific PCR primers as described previously (15). The number of PCR cycles was 29 or lower to maintain approximately linear relationship between the RNA input and signal. PCR products were separated on polyacrylamide gels and stained with ethidium bromide. Transcript levels were measured with FluorImager 595 using FluorQuant and Phoretix software (Nonlinear Dynamics Inc.). To confirm the identity of each product, visualized fragments were extracted from the gels and sequenced as described previously (16).

RESULTS

Biased distribution of cryptic and *de novo* 3′ss

To maintain previous categorization of aberrant splice sites (11), cryptic 3′ss are defined as those that are only used when a mutation disrupts use of the 3′ss consensus YAG. In contrast, the term *de novo* refers here to all aberrant 3′ss that were induced by mutations elsewhere than in the 3′YAG (Table 1). They include newly formed 3′ss AGs that are used instead of the natural one, or result from mutations that improve BPS, PPT or auxiliary splicing signals of the new 3′ss. For simplicity, a small number of aberrant 3′ss upstream of BPS that was generated by AG-creating mutations in the PPT were also included in the latter category, although these mutations do not improve the intrinsic strength of the new 3′ss, but instead interfere with the correct utilization of the natural site.

Compilation of previously published human aberrant 3′ss that were determined by sequencing mutated transcripts identified 147 cryptic and *de novo* sites in 89 genes (80 in exons and 67 in introns; Tables 1–3). A total of 77 cryptic 3′ss were activated as a result of point mutations in the 3′ YAG (Table 1). Thirty-eight point mutations were in intron position −1, position −2 was mutated in 33 cases and position −3 in 6 cases. Cryptic 3′ss were more frequent in exons ($n = 59$; $P < 10^{-4}$) than in introns. Three mutations produced cryptic 3′ss both in introns and in exons (17–19). Distribution of distances between authentic and cryptic 3′ss (Figure 1A) showed that 49 of the 59 (83%) exonic cryptic 3′ss were located within 21 nt just downstream of the intron–exon boundaries. The region of the same size upstream of 3′ss, which corresponds to an experimentally determined optimal distance between the branch point (BP) adenosine and 3′ss (20), is depleted of AG dinucleotides in several species, including humans (21,22). When cryptic 3′ss between authentic 3′AG and position +21 in the downstream exon were disregarded, the distribution of the remaining cryptic 3′ss was no longer biased towards exons and resembled normal distribution (Figure 1B).

Seventy 3′ss resulted from mutations outside the 3′ YAG consensus (Table 1). Most of them were in introns ($n = 48$; $P < 0.005$), with 41 of the 48 (85%) located within 25 nt upstream of the authentic 3′ss (Figure 1C). The majority

Table 1. Summary of aberrant 3′ splice sites in human genes

Location of cryptic or *de novo* splice sites Mutation	Exon In 3′YAG (cryptic)	Outside 3′YAG ('*de novo*')	Intron In 3′YAG (cryptic)	Outside 3′YAG ('*de novo*')	Total
Number of genes	39	16	14	35	89
Number of cryptic/*de novo* 3′ss (% in exons and introns)	59 (74)	21 (26)	18 (27)	49 (73)	147
Number of unique 3′ss (%)	56 (75)	19 (25)	18 (27)	48 (73)	141
Reading frame					
0	18	7	5	19	49 (33.3%)
+1	23	5	8	15	51 (34.7%)
+2	18	9	5	15	47 (32.0%)
Average distance (nt) between authentic and cryptic 3′ss (SD)	43.8 (46.4)[b]	57.9 (29.9)	−71.6 (120.3)	−19.1 (33.7)	10.7 (67.2)[b]
Median distance (nt) between authentic and cryptic 3′ss	12	54	−44	−12	2
Number of terminal exons (%)	10 (17)	2 (10)	6 (33)	3 (6)	21 (14)
Average S&S score (SD)					
Authentic (A)	85.3 (6.1)	80.8 (10.0)	87.6 (4.9)	81.3 (6.6)	83.6 (7.2)
Mutated (M)	68.4 (8.0)	80.2 (9.7)	72.3 (5.3)	78.8 (8.4)	74.1 (9.5)
Cryptic/*de novo* (CR)	74.2 (8.2)	82.3 (7.6)	83.8 (6.0)	77.6 (7.9)	77.7 (8.5)
Average difference[a]					
A-M (*P*-value)	16.9 (2×10^{-17})	0.9 (N.S.)	15.1 (9×10^{-10})	2.5 (N.S.)	9.6 (10^{-18})
M-CR (*P*-value)	5.7 (3×10^{-4})	2.0 (N.S.)	11.4 (10^{-16})	−1.2 (N.S.)	−3.6 (6×10^{-4})
A-CR (*P*-value)	11.2 (10^{-18})	−1.2 (N.S.)	3.8 (0.07)	3.7 (0.04)	6.0 (10^{-19})

[a]Mann–Whitney rank test (SPSS, SPSS Inc., USA).
[b]Excluding an outlier of 1165 nt (48).
SD, standard deviation; NS, not statistically significant.

of newly created AGs in introns that were used in the catalytic step were preceded by a strong PPT, although there were several exceptions (23–26). In addition to a major frequency peak at ∼19 nt upstream of the authentic 3′ss, a second small peak was observed ∼58 nt downstream of authentic 3′ss (Figure 1C and Table 1), but the number of these 3′ss was low. The second peak is likely to result from a depletion of *de novo* sites in the first ∼25 nt of the exon, rather than to their absolute over-representation further downstream. Such depletion could be due to interference by complexes assembled at the authentic 3′ss, selection against codons carrying AGs in the 5′ end of internal exons (27) or a lack of suitable BP adenosine within an optimal distance from *de novo* 3′ss.

Although intronic *de novo* 3′ss were almost exclusively generated by AG-creating mutations, such mutations contributed much less to the formation of exonic *de novo* 3′ss. Apart from three examples of direct AG creation in exons (28–30), three *de novo* sites were produced by point mutations in position –3 (31–33), one in position –5 (34) and three in position –6 (35–37) relative to the new intron–exon junction. Three aberrant exonic 3′ss resulted from mutations of the predicted BP adenosine (38,39) or conserved uridine in position –2 relative to BP (40), known hot-spots of single-nucleotide substitutions in the human BPS (15). Several aberrant 3′ss in exons resulted from more distant mutations (39,41), highlighting the importance of PPT, BPS and distant auxiliary splicing signals for the activation of 3′ss in this category. Together, these results indicated that, in contrast to cryptic 3′ss, distribution of *de novo* acceptors was biased towards introns, particularly towards the PPT. Unlike intronic 3′ss, which largely resulted from AG-creating mutations, *de novo* 3′ss in exons were commonly generated by mutations

elsewhere than in the 3′YAG of the new intron–exon boundary.

Sequence alignments of aberrant 3′ss in each category (Figure 2A–D) revealed a higher purine content for cryptic 3′ss in exons (Figure 2A) as compared with all human 3′ss (42) or corresponding authentic 3′ss (Figure 2E). Adenosine in position –3 was over-represented among exonic cryptic 3′ss (Figure 2A), possibly just reflecting higher levels of purine residues in the sequences surrounding the new intron–exon junction. The number of aberrant 3′ss in the next two categories (Figure 2B and C) was low. Intronic *de novo* 3′ss (Figure 2D) had frequent uridine in position +1 of the new exon as well as pyrimidines in position –4 of the new intron. Finally, purine depletion observed for aberrant 3′ss in the new PPT was in similar PPT positions as in authentic 3′ss (Figure 2E) or all 3′ss in vertebrates (42). This points to similar requirements for interactions with poly(Y) binding proteins, such as the large subunit of U2 auxiliary factor (U2AF[65]), and is consistent with frequent manifestation of splicing phenotypes as a result of mutations or naturally occurring DNA variants in these PPT positions (15,43–47).

The median distance between authentic and aberrant 3′ss was 2 nt (Table 1, each distance is shown in Supplementary Tables 1 and 2), while the absolute median distance was 16 nt. The median distances from authentic sites to exonic cryptic 3′ss or intronic *de novo* 3′ss were similar (12 nt; Figure 1 and Table 1). The median distances between intronic cryptic 3′ss and exonic *de novo* sites were also comparable (Table 1). Occasionally, mutations activated a cryptic 3′ss in an exon further downstream, up to 1165 nt from the mutation (48). In this extreme case, the authentic 3′ss was preceded by a very strong PPT, separating the 3′ss and the first upstream adenosine by 64 nt (or a predicted distant BP by 68 nt),

Table 2. Cryptic and *de novo* 3′ splice sites in exons

Gene	Phenotype	Mutation	Location of cryptic 3′ss	Reference
ABCR (ABCA4)	Stargardt disease	*E16+1G>C*	E16+3	(31)
ACAT1	Mitochondrial acetoacetyl-CoA thiolase deficiency	*E5+46C>T*	E5+51	(35)
ALG8	Glycosylation deficiency	*IVS1-2A>G*	E2+11	(96)
ARSA	Metachromatic leukodystrophy	*E8+22C>T*	E8+27	(37)
ASS	Citrullinaemia	*IVS14-1G>C*	E15+7	(97,98)
ATM	Ataxia telangiectasia	IVS38-2A>C	E39+61	(6)
ATM	Ataxia telangiectasia	IVS64-1G>C	E65+13	(6)
BRCA2	Breast cancer	*IVS23-2A>G*	E24+7	(99)
BTD	Biotinidase deficiency	*E1+56G>A*	E1+57	(30)
CBFA2 (RUNX1)	Familial thrombocytopenia	*IVS20-1G>T*	E21+13	(100)
CDKL5	Rett syndrome	*IVS13-1G>A*	E14+1	(101)
CLN3	Batten disease	*IVS15-1G>T*	E16+5	(102)
COH1	Cohen syndrome	*IVS51-1G>T*	E52+16	(103)
COL17A1	Epidermolysis bullosa	*IVS31-1G>T*	E32+9	(17)
COL17A1	Epidermolysis bullosa simplex	*IVS21-2A>C*	E22+27	(104)
COL1A2	Ehlers-Danlos syndrome type VII	*IVS5-1G>C*	E6+15	(105)
COL1A2	Osteogenesis imperfecta	*IVS27-2A>G*	E28+46	(106)
COL2A1	Stickler syndrome	*IVS17-2A>G*	E18+16	(107)
COL5A1	Ehlers-Danlos syndrome type I	*IVS4-2A>G*	E5+12, E5+15	(108)
DAF	CD55 deficiency	*E5+18C>T*	E5+44	(40)
DMD	Dystrophinopathy	IVS20-2A>G	E21+7	(109)
DMD	Muscular dystrophy	*IVS74-2A>G*	E76+60	(48)
EPB42	Recessive hereditary spherocytosis	*E11+39G>T*	E11+41	(33)
F8	Haemophilia A	*E11+32G>T*	E11+36	(34)
F8	Haemophilia A	*E16+26G>A*	**E16+47**	(38)
F8	Haemophilia A	*IVS15+1G>T*	**E16+47**	(50)
FGB	Hypofibrinogenaemia	*E4+115T>A*	E4+116	(28)
G6PC	Glycogen storage disease type 1a	*E5+86G>T*	E5+91	(36)
G6PD	Glucose-6-phosphate dehydrogenase deficiency	*IVS10-2A>G*	E11+9	(110)
GH-1	Growth hormone deficiency	*IVS3del28-45*	E3+98	(111)
GLA	Fabry disease	*IVS3-1G>A*	E4+1	(112)
GLA	Fabry disease	*IVS6-1G>A*	E7+1	(113)
GPB	Henshaw antigen	*E5+65C>G*	E5+65	(114)
HEXB	Sandhoff disease	*E11+8C>T*	**E11+112**	(49)
HEXB	Sandhoff disease	*IVS10-17A>G*	**E11+112**	(19)
*HLA-B*3916*	Deficient expression of *HLA-B*	*E3+17G>C*	E3+19	(32)
HPRT1	Hypoxanthine–guanine phosphoribosyltransferase deficiency	*IVS1-2A>G*	E2+5	(115)
HPRT1	Hypoxanthine–guanine phosphoribosyltransferase deficiency	*IVS5-1G>A*	E6+1	(115)
HPRT1	Hypoxanthine–guanine phosphoribosyltransferase deficiency	*IVS7-1G>A*	E8+21	(115)
HPRT1	Hypoxanthine–guanine phosphoribosyltransferase deficiency	*IVS9-1G>A*	E10+17	(116)
HPRT1	Hypoxanthine–guanine phosphoribosyltransferase deficiency	*IVS9-2A>G*	E10+17	(117,118)
HPRT1	Hypoxanthine–guanine phosphoribosyltransferase deficiency	*IVS9-2A>T*	E10+17	(117,118)
INSR	Rabson-Mendenhall's syndrome	*IVS4-2A>G*	E5+12	(119)
ITGA2B	Glanzmann thrombasthenia	IVS3-3DEL13	E4+18	(120)
ITGB4	Epidermolysis bullosa	*IVS31-19T>A*	E32+38	(41)
KRT14	Recessive epidermolysis bullosa simplex	*IVS2-2A>C*	E3+10	(121)
LAMA2	Muscular dystrophy	*IVS28-1G>C*	E29+69	(122)
LAMC2	Junctional epidermolysis bullosa	*IVS3-1G>A*	E3+2	(123)
LDLR	Familial hypercholesterolaemia	*IVS1-1G>C*	E2+10	(124)
LDLR	Familial hypercholesterolaemia	*IVS7-1G>C*	E8+17	(125)
LDLR	Familial hypercholesterolaemia	*IVS9-1G>A*	E10+7	(126)
LDLR	Familial hypercholesterolaemia	*IVS9-30 GTGCTGATG>CGGCT*	E10+54	(127)
LHX4	Syndromic short stature	*IVS4-1G>C*	E5+12, E5+20	(128)
MANBA	Beta-mannosidosis	IVS15-2A>G	E16+172	(81)
NF1	Neurofibromatosis type 1	*IVS27b-2A>T*	E28+293	(80)
NIS (SLC5A5)	Congenital hypothyroidism	*E13+67C>G*	E13+67	(29)
OAS1	Oligoadenylate synthase activity	*IVS6-1A>G*	E7+1, E7+137	(129)
OTC	Ornithine transcarbamylase deficiency	*IVS4-2A>T*	E5+12	(130)
PDE6B	Autosomal recessive retinitis pigmentosa	*IVS2-1G>T*	E3+12	(131)
PFKM	Muscle phosphofructokinase deficiency	*IVS6-2A>C*	E7+5, E7+12	(132)
PKLR	Pyruvate kinase deficiency	*IVS3-2A>T*	E4+6	(133)
PTPS	Pyruvoyl-tetrahydropterin synthase deficiency	*IVS1-3C>G*	E2+12	(134)
SCNN1G	Pseudohypoaldosteronism type 1	*IVS2-1G>A*	E3+6	(135)
SPG4	Spastic paraplegia	*IVS6-1G>A*	E7+8	(136)
SPINK5	Netherton syndrome	*IVS20-1G>A*	E21+1	(137)
TNFSF5 (HIGM1)	X-linked hyper-IgM syndrome	*IVS4-2A>G*	E5+8	(138)
TP53	Li-Fraumeni syndrome	*IVS3-1G>A*	E4+19	(139,140)
TP53	Li-Fraumeni syndrome	IVS5-11DEL11	E6+17	(141)
TP53	Li-Fraumeni syndrome	*IVS5-1G>A*	E6+1	(139)

Table 2. (*Continued*)

Gene	Phenotype	Mutation	Location of cryptic 3′ss	Reference
TP53	Lung cancer	*IVS3-1G>C*	E4+19	(142)
TSC2	Tuberous sclerosis	*IVS38-18G>A*	E39+74	(39)
TSC2	Tuberous sclerosis	*IVS9-15G>A*	E10+56	(39)
TSC2	Tuberous sclerosis	*IVS9-3C>G*	E10+56	(39)
UGT1A1	Crigler-Najjar syndrome type 1	*IVS3-2A>G*	E4+107	(143)
UGT1A1	Crigler-Najjar syndrome type 1	*IVS4-1G>A*	E5+7	(144)
XPA	Xeroderma pigmentosum group A	*IVS3-1G>C*	E4+2	(145)

which might have contributed to cryptic 3′ss selection such a long distance from authentic 3′ss. We also found that several identical aberrant 3′ss in exons were activated by distinct mutations, such as *HEXB* E11+112 by E11+8C>T and IVS10-17A>G (19,49), or *F8* E16+47 by a mutation at IVS15-1 and an exon 16 mutation (38,50) (Table 2), providing clear evidence that mutations in very diverse positions may result in the same splicing defect.

As with the cryptic 5′ ss (11), neither cryptic nor *de novo* 3′ss showed any bias towards a particular reading-frame phase relative to the position of authentic 3′ss (Table 1). This may reflect only a partial elimination of mRNAs with premature termination codons by nonsense-mediated mRNA decay or even a complete unresponsiveness of mutated transcripts to RNA surveillance mechanisms, as reported for *HBB* (51).

Finally, similar to the cryptic 5′ ss (11), pair-wise comparisons of the average S&S matrix scores for authentic 3′ ss and their mutated and cryptic counterparts showed significant differences, with the identical A>CR>M score hierarchy for both exonic and intronic cryptic 3′ ss (Table 1 and Supplementary Tables 1 and 2). In contrast, the average S&S matrix scores for *de novo* 3′ ss were not significantly higher than corresponding mutated sites and were not lower than corresponding authentic 3′ ss, except for a borderline significance of intronic *de novo* 3′ ss (Table 1). Thus, the predictive value of the S&S matrix scores was evident only for cryptic, but not for *de novo* 3′ss.

AG-creating mutations in the PPT that suppress authentic 3′ss and activate aberrant 3′ss: a role for 'BPS-new AG' distance

The majority of AG-creating mutations in the PPT that resulted in the activation of aberrant 3′ss used the newly introduced AGs in the second step of splicing (Figure 1C). However, there were several exceptions. Newly created AGs in *LIPC* (52), *HEXB* (19) and *AR* (23) were not efficiently used for exon ligation, but led to the activation of aberrant 3′ss upstream of putative BPS, while suppressing authentic 3′ss. A similar repression of authentic 3′ss was observed for *FBN2*, where the introduction of AG, which was not used in the catalytic step *in vivo*, resulted in exon skipping (53). In this small group of mutations, recognition of new AGs was sufficient to suppress utilization of authentic 3′ss, but was insufficient for exon joining. Sequence inspection of these cases revealed that even though the distance between newly created AGs and authentic 3′ss was variable, the distance between predicted BPS and the mutation was similar, ranging from 11 to 15 nt (Figure 3A).

To investigate the distance requirements for the activation of aberrant 3′ss, we constructed three-exon splicing reporters for *LIPC*, *HEXB* and *FBN2* (Figure 3B). We refer to these 3′ss as upstream (U) and mutated (M) (Figure 3A) to avoid ambiguous distinction between cryptic and *de novo* 3′ss in these cases. As each germline mutation was an A>G transition (19,52,53), which are characteristic of BP substitutions (15), we first attempted to determine the BP. However, none of the pre-mRNA substrates could be spliced *in vitro* with varying concentrations of nuclear extracts and Mg^{2+} (data not shown). As A>G substitutions in the BP impart a particularly strong block of splicing *in vitro* (54) and *in vivo*, which may facilitate BP determination (J. Králováčová, H. Lei and I. Vořechovský, manuscript submitted), we mutated the putative BP adenosines into G, C and T in each wild-type construct (Figure 3C). Examination of RNA products after transfection showed that only the G-containing clones led to aberrant splicing. In *FBN2*, exon skipping observed *in vivo* in a patient with contractural arachnodactyly (53) was replicated in 293T cells where it was accompanied by a partial utilization of the newly introduced AG (Figure 3C and D). In *LIPC*, we found two aberrant 3′ss, as reported earlier (52). The first 3′ss was utilized in ~6% at the site of mutation (IVS1-13) and the other was upstream of the predicted BPS (IVS1-78; Figure 3B, C and E). The ratios of unspliced/-13/-78 RNA products described previously (1/2.9/0.8, respectively) (52) was altered in favour of the upstream cryptic 3′ss in our system (Figure 3C). This was most likely due to a better BPS consensus in our construct, because a previously used minigene (52) had intron truncation just upstream of IVS1-78 and lacked a suitable alternative BPS (data not shown). Finally, the *HEXB* transition (Figure 3A) also activated two aberrant 3′ss. One was 37 nt upstream of exon 11 and the other was in exon 11, 112 nt downstream of the authentic site (Figure 3B and F). We could not exclude that IVS10-17A is the BP of the central exon as suggested earlier (19), but improved splicing of uridine- and, to a lesser extent, cytosine-containing pre-mRNAs (Figure 3C, middle panel) suggested that this mutation is in the PPT since uridines are the preferred PPT nucleotides (55,56). As *LIPC* and *FBN2* mutations were even closer to the 3′ss than the *HEXB* mutation (Figure 3A), their BPs are likely to map upstream of newly introduced AGs as well.

To test whether the aberrant 3′ss were activated only by AGs and not by other dinucleotides, we changed BP-1 adenosines to guanosines in each wild-type and mutated *FBN2/LIPC* minigene to create additional dinucleotides GA, GC, GT and GG (Figure 3G). Of the eight dinucleotides (Figure 3C and G), only AG-containing reporters maintained the splicing defects. In *LIPC*, utilization of the upstream 3′ss IVS1-78 was

Table 3. Cryptic and *de novo* 3′ splice sites in introns

Gene	Phenotype	Mutation	Location of cryptic 3′ss	Reference
APOE	ApoE deficiency	IVS3-2A>G	IVS3-52	(73)
AR	Androgen insensitivity	IVS2-11T>A	IVS2-69	(23)
ATM	Ataxia telangiectasia	IVS32-12A>G	IVS32-11	(6)
ATM	Ataxia telangiectasia	IVS16-10T>G	IVS16-9	(6)
ATP7B	Wilson disease	IVS11-2A>G	IVS11-39	(146)
BRCA1	Breast cancer	IVS7-24del10	IVS7-59	(147)
BRCA1	Familial breast cancer	IVS5-12A>G	IVS5-11	(148)
CFTR	Cystic fibrosis	IVS17a-26A>G	IVS17a-25	(149,150)
CHIT1	Chitotriosidase deficiency	E10+20dupl24	E10+103	(151)
COL17A1	Benign epidermolysis bullosa	*IVS31-1G>T*	IVS31-69, IVS31-264	(17)
COL5A1	Ehlers-Danlos syndrome type II	IVS13-2A>G	IVS13-100	(94)
CPO	Hereditary coproporphyria	IVS1-15C>G	IVS1-14	(152)
CYBB	Chronic granulomatous disease	IVS4-15del36	IVS4-179	(153)
CYP21B	21-hydroxylase deficiency	IVS2-13C>G	IVS2-19, IVS2-33	(154)
DBT	Maple syrup urine disease	IVS4-17TTT>AAA	IVS4-17	(155)
DMD	Muscular dystrophy	IVS59-9T>A	IVS59-7	(156)
DMD	Muscular dystrophy	IVS8-15A>G	IVS8-14	(48)
ELN	Supravalvular aortic stenosis	IVS15-3C>G	IVS15-44	(93)
ERCC3	Xeroderma pigmentosum	IVS14-6C>A	IVS14-4	(157)
FANCA	Fanconi anaemia	IVS15-1G>T	IVS15-90	(158)
GCH1	Dystonia	IVS2-2A>G	IVS2-1	(159)
HBB	Beta-thalassaemia	IVS1-15T>G	IVS1-14	(160)
HBB	Beta-thalassaemia	IVS1-21G>A	IVS1-19	(161–163)
HBB	Beta-thalassaemia	IVS2-A>G	IVS2-271	(74)
HEXB	Sandhoff disease	IVS10-17A>G	IVS10-37	(19)
HEXB	Sandhoff disease	IVS12-26G>A	IVS12-24	(24,164)
HPRT1	Hypoxanthine–guanine phosphoribosyltransferase deficiency	IVS8-3T>G	IVS8-2	(165)
HPRT1	Hypoxanthine–guanine phosphoribosyltransferase deficiency	IVS8-16G>A	IVS8-14	(165,166)
ITGB2	Leukocyte adhesion deficiency	IVS6-14C>A	IVS6-12	(167)
L1CAM	X-linked hydrocephalus	IVS17-19A>C	IVS-69	(168)
L1CAM	X-linked hydrocephalus, MASA syndrome, spastic paraplegia	IVS26-12G>A	IVS26-10	(169)
LIPC	Hepatic lipase deficiency	IVS1-14A>G	IVS1-78, IVS1-13	(52)
MECP2	Rett syndrome	IVS1-6C>G	IVS1-5	(170)
MLYCD	Malonyl-CoA decarboxylase deficiency	IVS4-14A>G	IVS4-13	(171)
MPO	Myeloperoxidase deficiency	IVS11-2A>C	IVS11-109	(72)
MTM1	X-linked recessive myotubular myopathy	IVS12-10A>G	IVS12-9	(172)
MYBPC3	Hypertrophic cardiomyopathy	IVS14-13G>A	IVS14-11	(173)
NF1	Neurofibromatosis type I	IVS15-16A>G	IVS15-15	(7,25)
NF1	Neurofibromatosis type I	IVS15-15A>G	IVS15-14	(25)
NF1	Neurofibromatosis type I	IVS10a-9T>A	IVS10a-7	(7,25)
NF1	Neurofibromatosis type I	IVS39-12T>A	IVS39-10	(80)
NF1	Neurofibromatosis type I	IVS26-2A>T	IVS26-14, IVS26-17	(80)
NF1	Neurofibromatosis type I	IVS11-3C>G	IVS11-43	(7)
NF1	Neurofibromatosis type I	IVS15-12T>G	IVS15-11	(25)
OCA2	Type II oculocutaneous albinism	IVS5-19A>G	IVS5-18	(174)
PAH	Phenylketonuria	IVS8-7A>G	IVS8-6	(66)
PAH	Phenylketonuria	IVS10-11G>A	IVS10-9	(175)
RB1	Retinoblastoma	IVS22-8T>A	IVS22-6	(176)
SAG(S-ARRESTIN)	Retinitis pigmentosa	IVS10-25A>G	IVS1-24	(177)
SALL1	Townes-Brocks syndrome	IVS2-19T>A	IVS2-17	(178)
SERPINC1	Type I antithrombin deficiency	IVS4-14G>A	IVS4-12	(179)
SOD1	Amyotrophic lateral sclerosis	IVS4-10T>G	IVS4-9	(180)
TCF1 (HCF-1A)	Maturity-onset diabetes of the young	IVS4-2A>G	IVS4-202	(181)
TCF1 (HCF-1A)	Maturity-onset diabetes of the young	IVS7-6G>A	IVS7-4	(181)
TH	Extrapyramidal movement disorder	IVS11-24T>A	IVS11-36	(57)
TNFRSF1A	Periodic fever syndrome	IVS3-14G>A	IVS3-12	(26)
TP53	Li-Fraumeni syndrome	IVS9-1G>C	IVS9-44	(92)
TP53	Li-Fraumeni syndrome	IVS3-11C>G	IVS3-10	(139)
TPMT	Thiopurine methyltransferase deficiency	IVS9-1G>A	IVS9+1, IVS9-330	(18)
WRN	Werner's syndrome	IVS29-7T>A	IVS29-5	(182)
ZAP70	Severe combined immunodeficiency	IVS9-11G>A	IVS9-9	(183)

highest in pre-mRNAs containing guanosine -15 (IVS1-15G), followed by those containing A or C, and was lowest for pre-mRNAs with uridine at this position, further supporting the PPT location of the newly formed AG. To formally show that the predicted BPS itself can tolerate AG dinucleotides, we employed two splicing reporters that were derived from the *TH* and *TSC2* genes. Both mutated minigenes represented previously observed substitutions in the predicted BPS that resulted in genetic disease (39,57). Substitutions of the predicted BP adenosines to the remaining nucleotides
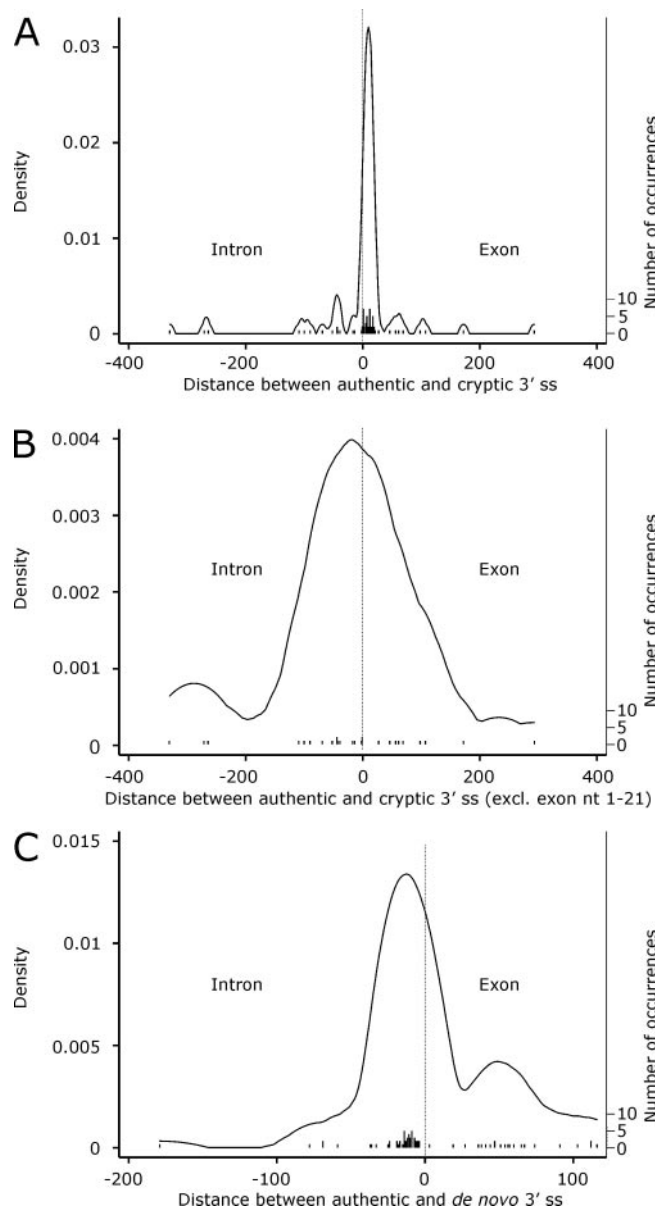
Figure 1. Distribution of the distances between authentic and aberrant 3′ss. (**A**) Cryptic splice sites resulting from mutations at the 3′YAG consensus. (**B**) Cryptic splice sites resulting from mutations of the 3′YAG, except for cryptic 3′ss in exon positions 1–21. (**C**) Aberrant 3′ss due to mutations outside the 3′YAG ('*de novo*'). The Stata statistical package (v. 8.2, StataCorp, TX) was used to fit kernel density plots to the distances between authentic and cryptic/*de novo* 3′ss. Positive and negative numbers correspond to aberrant 3′ss located in the downstream exon or the upstream intron, respectively. The number of occurrences of aberrant 3′ss is shown as short vertical bars for each distance (in nt). The corresponding scale is shown on the right side. A cryptic splice site that was found at a large distance from the authentic 3′ss (48) was omitted from the plot.

in both reporters dramatically increased exon skipping (J. Královičová, H. Lei and I. Vořechovský, manuscript submitted). Since predicted BP-As were preceded by Gs in each intron, we mutated BP-1Gs to As in the wild-type and mutated reporters, but the splicing pattern observed for the constructs carrying AA, AC, AT and AG dinucleotides was similar to those containing GA, GC, GT and GG (Figure 3H and data not shown). Taken together, these data suggested that activation of
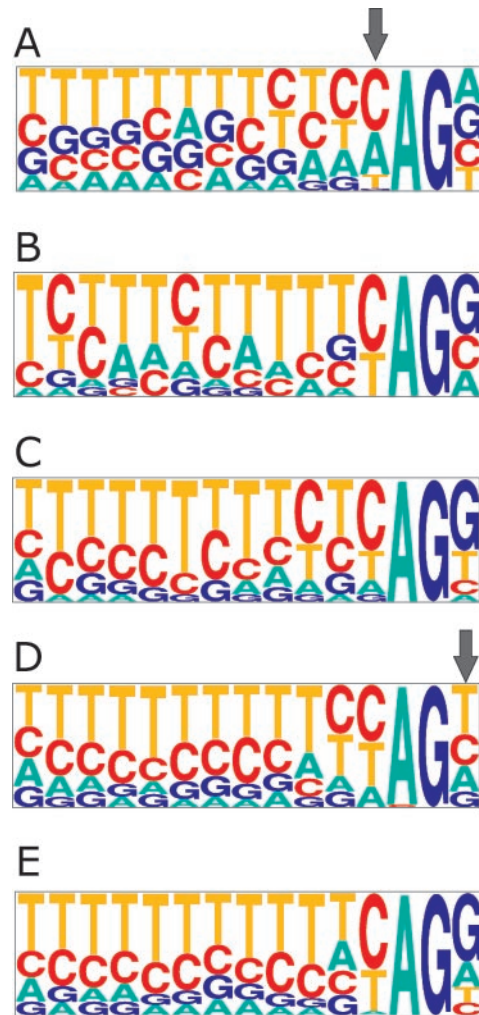

Figure 2. Consensus sequences of aberrant 3′ss. (**A**) Cryptic splice acceptors in exons resulting from mutations of the 3′ YAG; (**B**) 3′ss in exons due to mutations outside the 3′ YAG; (**C**) cryptic 3′ss in introns that resulted from mutations of the 3′ YAG; (**D**) aberrant 3′ss located in introns generated by mutations outside the 3′ YAG. (**E**) Consensus sequences of corresponding authentic 3′ss (*n* = 147). The relative nucleotide frequencies at each position were plotted with a pictogram utility (http://genes.mit.edu/pictogram.html). The height of each letter is proportional to the frequency of the corresponding base at the given position. Arrows indicate over-representation of adenine in position −3 (A) and of uridine in position +1 (D) of the new intron or exon, respectively.

3′ss upstream of the BPS and inhibition of authentic 3′ss is specific for AG dinucleotides in the PPT.

Reduction of the 'BPS-new AG' distance improves the expression of natural transcripts

To test a role of the 'BPS-newAG' distance in repression and activation of three competing 3′ss, we examined the splicing pattern of *LIPC* and *HEXB* constructs containing serial 3 nt deletions in this region (Figure 4A–C). In addition, we modified this distance in the *HLA-DQB1* reporter (allele **0602*), which contains a preexisting AG dinucleotide downstream of the BPS (Figure 4A and D) (15). The BP of *DQB1* exon 4 was determined by reverse transcription and mutagenesis (15) and corresponded to the best match to mammalian BPS in intron 3 (Figure 4A) and to a computationally predicted BPS (14).

In *LIPC*, deletions that reduced the distance between predicted BP and the newly created AG from 13 to 10, 7 and 4 nt, eliminated use of the mutated site (Figure 4A and B, lanes 1–5). Three- and six-nt deletions rescued normal splicing to 16 and 89%, respectively, and progressively inhibited upstream 3′ss (lanes 3 and 4). However, the largest deletion no longer improved splicing to the authentic 3′ss (lane 5), most likely by reducing the gap between the BP and authentic 3′ss to

only 17 nt, which is below the experimentally determined optimum of 19–23 nt (20). These results suggested that the minimum 'BP-new AG' distance required for engagement of the new AG in the catalytic step and for complete suppression of authentic 3′ss in this pre-mRNA was ~13 nt. Bringing this AG closer to the BPS was associated with a significant production of natural transcripts. Their increase was accompanied by a decrease of the S&S scores for the mutated sites (Figure 4A), suggesting that diminished recognition of the *de novo* site due to reduced pyrimidine content upstream of the mutated AG in deletion mutants weakens AG-induced inhibition of the authentic 3′ ss. Also, we could not exclude a beneficial effect of a reduced 'BP-authentic AG' distance on the expression of natural transcripts.

In *HEXB*, splicing to the authentic 3′ss was rescued only by the largest deletion, which decreased the 'BP-new AG' distance from 15 to 6 nt (Figure 4C). Reduction of the S&S score for the mutated 3′ ss in this deletion mutant was associated with increased use of authentic sites and decreased use of 3′ ss E+112. In *DQB1*, a preexisting AG is located 7 nt downstream of the BP adenosine (Figure 4A) and is not used in the second splicing step (15). Insertion of five uridine residues in front of the 3′ss CAG consensus to reach a BP-AG distance comparable with the remaining reporters (Figure 3A) was sufficient for ~9% utilization of this AG (Figure 4D, lane 2). The new 3′ss (Figure 4E) was markedly promoted further by extending the new PPT to $(T)_{10}$ with a concomitant improvement of exon inclusion, both in the presence (alleles *DQB1*02-05*) and absence (*DQB1*06*) of a polymorphic guanosine in position -14 (Figure 4A and D, lanes 3 and 6). Together, these results showed that moving the newly created AG closer to the BP while maintaining the distance between authentic and new AGs promoted selection of the authentic 3′ss. They also suggested that AG-creating mutations located closer to the BPS
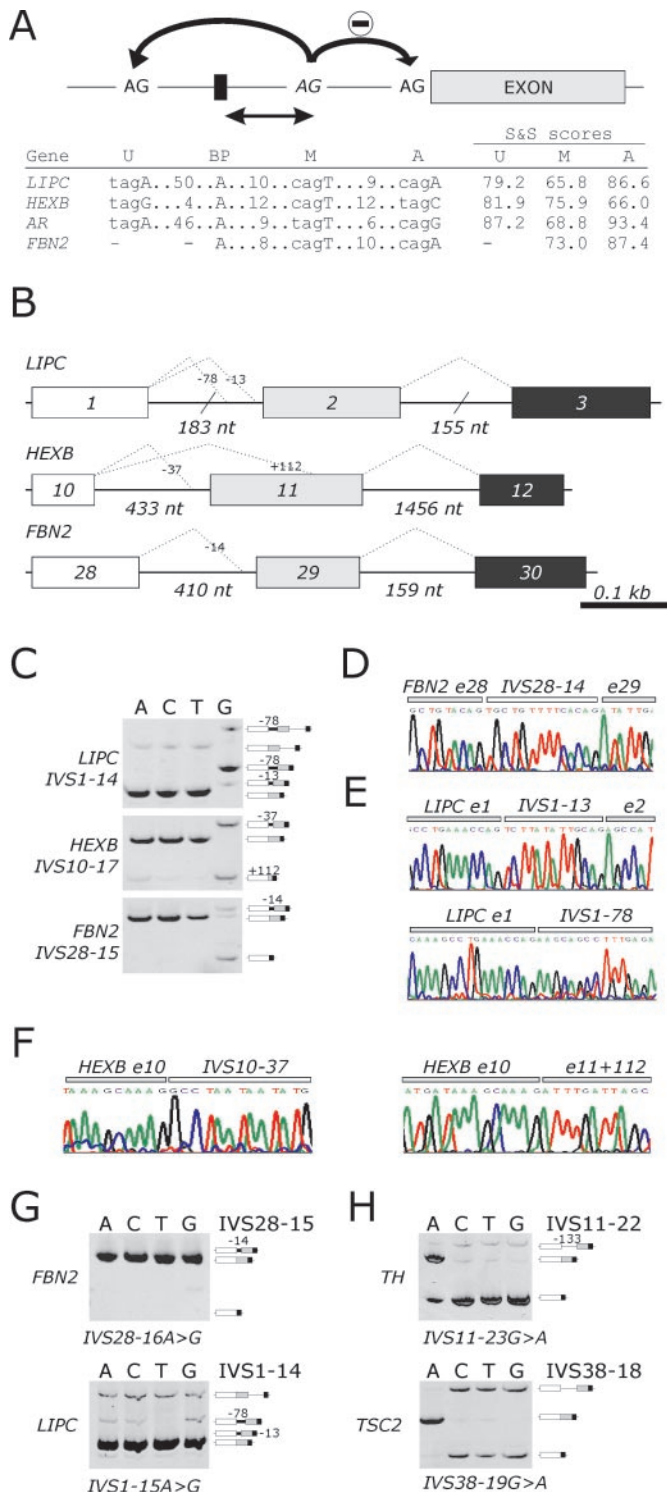

Figure 3. AG-creating mutations in the PPT that activate aberrant 3′ss upstream of predicted BPS. (**A**) Aberrant 3′ss activated by newly created AGs (in italics) that repress (minus sign) authentic 3′ss in the PPT (upper panel). Distances between upstream 3′ss (U) and predicted BP, between newly created AG (mutated, M) and BP (arrow) and between the mutated and authentic (A) 3′ss are in base pairs, bp (lower panel). The S&S scores were computed for U, M and A 3′ss using an algorithm as described previously (12,13). BPS is shown as a black rectangle. Disease-causing mutations were *LIPC* IVS1-14A>G (52), *HEXB* IVS10-17A>G (19), *AR* IVS2-11T>A (23) and *FBN2* IVS28-15A>G (53). Putative BPSs were GGCTAAG, GCCTAAT, TATCAAC and TGACAAT, respectively. (**B**) Schematic representation of minigene constructs. Exons are shown to scale (scale unit is 0.1 kb). The sizes of minigene introns (lines; not to scale) are shown below each construct. Intron truncations are indicated by a slash. Full *LIPC* introns 1 and 2 were 106.2 and 3.3 kb, respectively. Allele-specific *DQB1* minigenes were described previously (15). (**C**) Splicing pattern of mutated minigenes after transfection into 293T cells. RT–PCR products amplified with vector-specific primers PL3 and PL4. Wild-type minigenes containing predicted BP adenine in the indicated positions were mutated to C, T and G. RNA species were confirmed by sequencing and are schematically shown on the right side and in (B). The first, second and third exons are shown as white, grey and black boxes, respectively. Introns are shown as lines. Thick lines indicate partial intron retention due to activation of aberrant splice sites. (**D–F**) Nucleotide sequence of RT–PCR products bridging aberrant 3′ss in mutated constructs. Exons (e) are indicated by grey rectangles, introns (IVS, intervening sequence) by a white rectangle. Aberrant 3′ss are designated by a distance from the corresponding authentic splice site. (**G**) Activation of aberrant splice sites is specific for AG dinucleotides. Mutations removing AG dinucleotides are indicated at the top and bottom of each panel. (**H**) AG dinucleotides within predicted BPS do not activate cryptic splice sites. Mutations creating AG dinucleotides in the predicted BPS are indicated at the top and bottom of each panel.
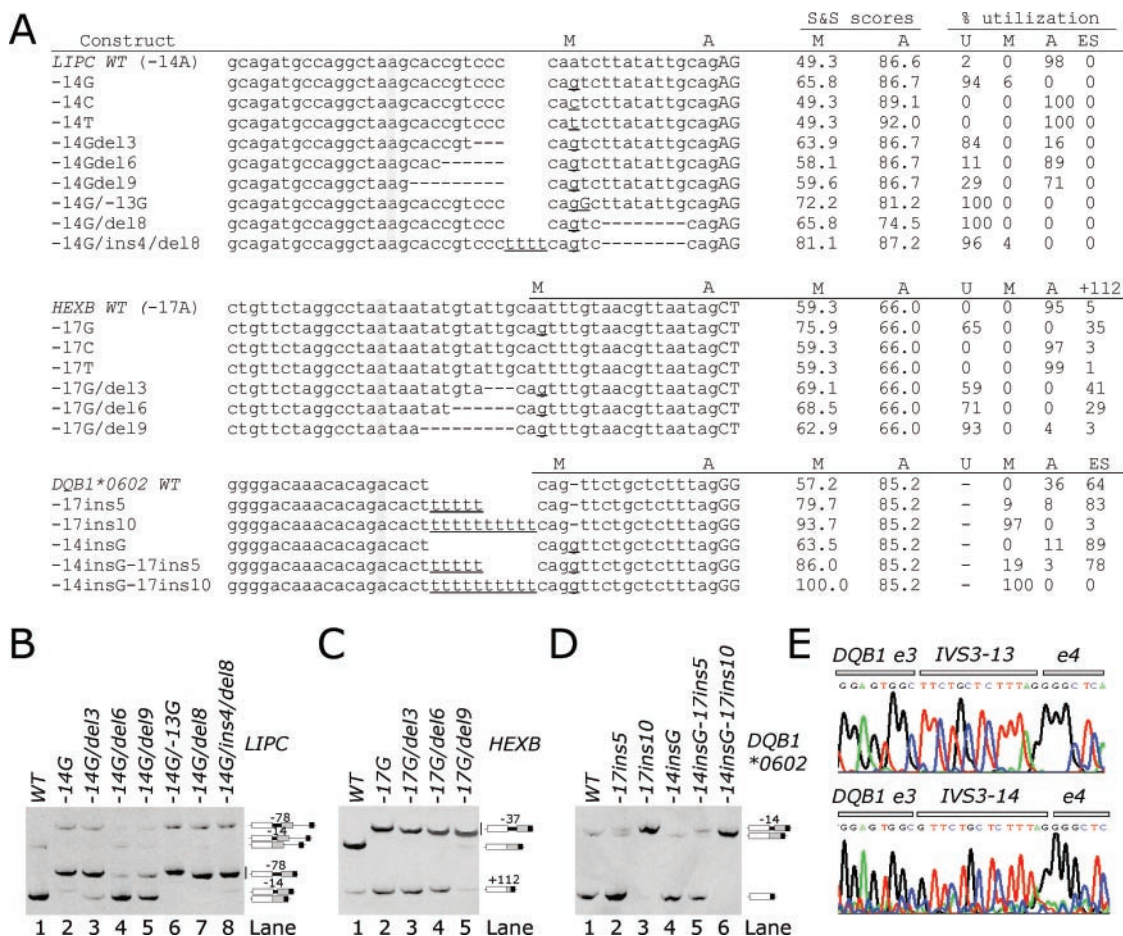
A

Construct				S&S scores M	A	% utilization U	M	A	ES
LIPC WT (−14A)	gcagatgccaggctaagcaccgtccc	caatcttatattgcagAG		49.3	86.6	2	0	98	0
−14G	gcagatgccaggctaagcaccgtccc	cagtcttatattgcagAG		65.8	86.7	94	6	0	0
−14C	gcagatgccaggctaagcaccgtccc	cactcttatattgcagAG		49.3	89.1	0	0	100	0
−14T	gcagatgccaggctaagcaccgtccc	cattcttatattgcagAG		49.3	92.0	0	0	100	0
−14Gdel3	gcagatgccaggctaagcaccgt---	cagtcttatattgcagAG		63.9	86.7	84	0	16	0
−14Gdel6	gcagatgccaggctaagcac------	cagtcttatattgcagAG		58.1	86.7	11	0	89	0
−14Gdel9	gcagatgccaggctaag---------	cagtcttatattgcagAG		59.6	86.7	29	0	71	0
−14G/−13G	gcagatgccaggctaagcaccgtccc	cagGcttatattgcagAG		72.2	81.2	100	0	0	0
−14G/del8	gcagatgccaggctaagcaccgtccc	cagtc-------cagAG		65.8	74.5	100	0	0	0
−14G/ins4/del8	gcagatgccaggctaagcaccgtccctttt	cagtc-------cagAG		81.1	87.2	96	4	0	0

				M	A	U	M	A	+112
HEXB WT (−17A)	ctgttctaggcctaataatatgtattgcaatttgtaacgttaatagCT			59.3	66.0	0	0	95	5
−17G	ctgttctaggcctaataatatgtattgcagtttgtaacgttaatagCT			75.9	66.0	65	0	0	35
−17C	ctgttctaggcctaataatatgtattgcactttgtaacgttaatagCT			59.3	66.0	0	0	97	3
−17T	ctgttctaggcctaataatatgtattgcattttgtaacgttaatagCT			59.3	66.0	0	0	99	1
−17G/del3	ctgttctaggcctaataaatatgta---cagtttgtaacgttaatagCT			69.1	66.0	59	0	0	41
−17G/del6	ctgttctaggcctaataatat------cagtttgtaacgttaatagCT			68.5	66.0	71	0	0	29
−17G/del9	ctgttctaggcctaataa---------cagtttgtaacgttaatagCT			62.9	66.0	93	0	4	3

				M	A	U	M	A	ES
*DQB1*0602* WT	ggggacaaacacagacact	cag-ttctgctctttagGG		57.2	85.2	−	0	36	64
−17ins5	ggggacaaacacagacactttttt	cag-ttctgctctttagGG		79.7	85.2	−	9	8	83
−17ins10	ggggacaaacacagacactttttttttttcag-ttctgctctttagGG			93.7	85.2	−	97	0	3
−14insG	ggggacaaacacagacact	caggttctgctctttagGG		63.5	85.2	−	0	11	89
−14insG-17ins5	ggggacaaacacagacactttttt	caggttctgctctttagGG		86.0	85.2	−	19	3	78
−14insG-17ins10	ggggacaaacacagacactttttttttttcaggttctgctctttagGG			100.0	85.2	−	100	0	0

Figure 4. A role for the BPS-new AG distance and/or the strength of the new PPT in upstream cryptic 3′ss activation. (**A**) Nucleotide sequences of splicing reporter constructs at the 3′ss are followed by the S&S matrix scores and by the percentage of splice site utilization of the indicated RNA products (means of duplicate transfections). Intronic sequences are in lower case, exonic sequences are in upper case. Putative BPs are shaded. U, aberrant 3′ss upstream of predicted BPS; M, newly created (preexisting in *DQB1*) or proximal AG between the authentic 3′ss and BP; A, authentic or distal 3′ss. ES, exon skipping; +112, % utilization of the splice site +112 by the *HEXB* pre-mRNAs site. The S&S scores were calculated according to the algorithm described previously (12,13). BPSs of *LIPC* exon 2 and of *DQB1* exon 4 were predicted by a branch site tool (http://ast.bioinfo.tau.ac.il/), with BPS scores 3.25 and 3.2, respectively. No BPS was predicted for *HEXB*. The *HEXB* IVS10-29A>G and IVS10-29A>T gave splicing patterns identical to the wild-type constructs (data not shown). RT–PCR products for the *LIPC*-14Y and *HEXB*-17Y mutations are shown in Figure 3C. (**B–D**) RNA products generated by wild-type and mutated constructs after transfection into 293T cells. The designation of splicing reporter constructs (top of each panel) corresponds to that in (A). RNA products were confirmed by sequencing and are schematically shown on the right side. (**E**) Nucleotide sequences of RT–PCR products illustrating aberrant splice sites in *DQB1* reporters.

and/or preceded by weaker PPTs may have less severe phenotypic consequences and provided support to the model that proposes a substrate-specific area of exclusion of AG dinucleotides downstream of the BP.

Prevalence and sequence context of intervening AG dinucleotides in authentic and aberrant 3′ss

The mutated AG was preceded by the pyrimidine and followed by the uridine in each case (Figure 3A). The occurrence of pyrimidines in the former position suggested that, similar to AGs in authentic 3′ss, cytosines and uridines may facilitate partial recognition of intervening AG dinucleotides. However, the presence of uridines that immediately followed each of these AGs was conspicuous, because uridine is the least frequent nucleotide in humans, mouse, zebrafish and fugu in this position (42). This could simply reflect a higher uridine content in the surrounding sequence (cf. Figures 2D and E); nevertheless, we analysed the effect of T>G alterations in this

position on *LIPC* and *DQB1* splicing. The *LIPC* IVS1-13T>G mutation, which improved a match to the 3′ss YAG/R, completely eliminated utilization of this site and promoted the upstream 3′ss (Figure 4B, lane 6). In contrast, an insertion of a polymorphic guanosine in position −14 of *DQB1* intron 3 further promoted the new 3′ss (Figure 4D, lanes 2 and 5).

AGs between BPS and authentic 3′AG are rare (21), but they are more common near the BPS (14) or closer to the 3′ss, often as 'tandem' (NAGNAG) acceptors (58) that effectively compete with each other. Inspection of 147 aberrant 3′ss revealed 15 (10%; 11 in exons, 4 in introns) intervening AGs in a 14 nt sequence upstream of authentic 3′ss (Supplementary Tables 1 and 2). In contrast, corresponding authentic 3′ss had only four AGs in this region ($P < 0.01$; Fisher exact test), suggesting that intervening AGs are more common in aberrant than in authentic 3′ss.

Use of 3′AG in splicing is influenced by the identity of the preceding nucleotide, with a hierarchy of competitiveness CAG ∼ UAG > AAG > GAG (59,60). We analysed the

frequency of each nucleotide that precedes intervening AGs as a function of distance from the BPs predicted through comparison of mouse and human introns (14). To maximize the probability of studying functional BPSs, we limited our analysis to 40 388 human introns, in which predicted BPs were located within a 40 nt distance from authentic 3'ss. Interestingly, cytosines were consistently over-represented in each position between 3 and ~21 nt downstream of the predicted BP adenosine (Figure 5A). A frequency peak in position 3 apparently reflects over-representation of cytosines in the last nucleotide of predicted BPS ($YNCURAY_{+1}A_{+2}G_{+3}$). The relative nucleotide frequencies varied little around 40% for cytosines and 20% for the remaining nucleotides along the whole distance. An exception was position 8 and 9 downstream of BP adenosine where purine depletion was even greater (Figure 5A). The non-random distribution of nucleotides that preceded intervening AGs ($P < 10^{-15}$ for positions 4–21) is consistent with the presence of splicing complexes covering this region and appears to support functionality of most predicted BPSs, although their average distance from authentic 3'ss (14) is longer than an experimentally determined optimum (20,61).
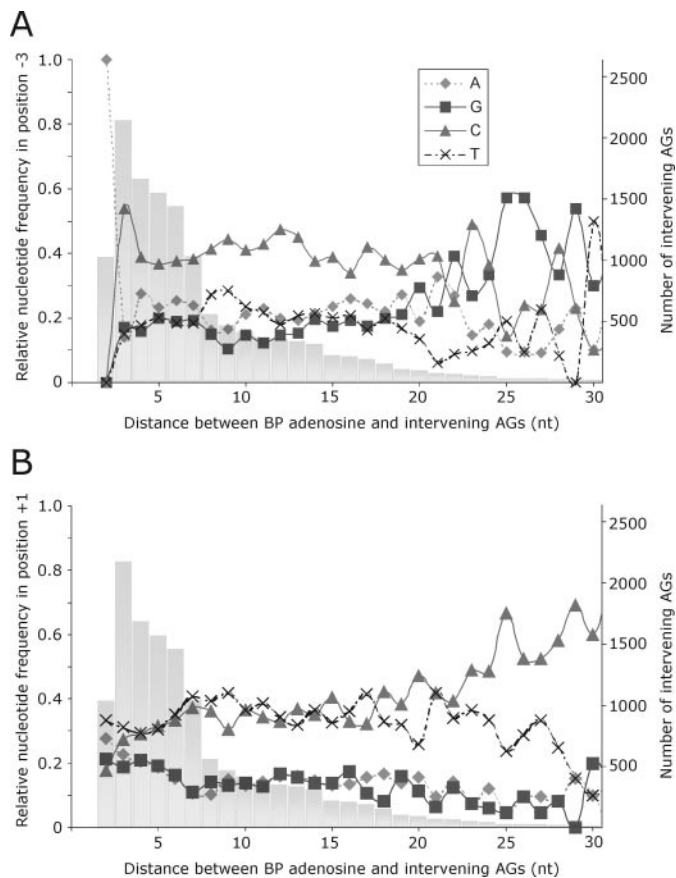
Figure 5. Sequence context of intervening AG dinucleotides in authentic 3'ss. Relative frequencies of nucleotides that immediately precede (**A**) or follow (**B**) AGs located 2–30 nt downstream of predicted BP adenosines. Corresponding numbers of intervening AGs are shown as grey columns. Distances between the predicted BP adenosine and downstream AG are in nucleotides (nt) as follows: 2 nt ($YNCURAA_{+1}G_{+2}$; BP is underlined), 3 nt ($YNCUR\underline{A}Y_{+1}A_{+2}G_{+3}$), 4 nt ($YNCUR\underline{A}Y_{+1}N_{+2}A_{+3}G_{+4}$), etc., up to 30.

Among nucleotides that followed intervening AGs, pyrimidines were consistently over-represented in positions 4–28 nt downstream of the predicted BP adenosine (Figure 5B, $P < 10^{-9}$), in sharp contrast to authentic 3'AGs. The number of intervening AGs was low for distances more than 20 nt downstream from the predicted BP adenosines, but 117 of the 208 (56%) AGs were followed by cytosine residues in this region. The distribution of nucleotides that precede and follow intervening AGs was very similar in 40 626 murine introns where the distance characterized by over-representation of cytosines that preceded intervening AGs had the same length (3–21 nt downstream of predicted BP adenosine; data not shown). Over-representation of pyrimidines in position +1 was also very similar in mouse introns to that observed in human introns (Figure 5 and data not shown). In a 14 nt sequence upstream of aberrant 3'ss, 5 of the 15 AGs were preceded by cytosines and 10 of the 15 AGs were followed by pyrimidines, which did not appear to be significantly different from intervening AGs of authentic 3'ss.

Finally, since closely spaced AG dinucleotides at the 3'ss (20,60–62) and in a splicing silencer (63) may interfere with each other, we employed the *LIPC* reporter to bring the newly created and authentic 3'AGs to proximity. We introduced an 8 nt deletion just upstream of the authentic 3'ss in the construct with the new AG to shorten this distance to just 5 nt while maintaining the BP-3'ss length within the previously determined optimum (20). The 5 nt distance was still sufficient for an interplay between closely spaced AGs (20,61,63). The 8 nt deletion repressed utilization of both competing AGs and resulted in exclusive splicing to upstream 3'ss (Figure 4B, lane 7). As with *DQB1* (Figure 4A and D), a small insertion of repetitive uridines between the BP and the newly created AG rescued splicing to this AG in the presence of the 8 nt deletion (Figure 4B, lane 8). However, this AG was used slightly less than for the -14A>G mutation, despite more optimal BP-AG distance and better PPT of the new exon, supporting a strong interference by the authentic 3'AG (Figure 4A and B, lanes 2 and 8).

The influence of SR proteins on utilization of aberrant 3' and 5' ss in *LIPC*

To illustrate the effect of known regulators of splicing on cryptic 3'ss utilization, we co-expressed the wild-type and mutated (IVS1-14A>G) *LIPC* reporter with SR proteins (Figure 6A). The co-transfection experiments revealed activation of two more aberrant splice sites (designated e2-24 and IVS2-94, Figure 6B) of the second minigene intron. In both mutated and wild-type reporters, activation of proximal 3'ss (IVS2-94) and distal 5' ss (e2-24) was promoted by ASF/SF2 and SRp40. Since the IVS2-94 splice site has a very strong PPT and is separated from the authentic 5' ss by only 61 nt, activation of the distal 5' ss in exon 2 may be due to restrictions on minimum intron size. A subset of SR proteins also promoted the use of upstream 3'ss IVS1-78 in both the wild-type and mutated minigenes and reduced the amount of correctly spliced products and those spliced to the mutated site IVS1-14 (Figure 6A). These results confirm that SR proteins may promote selection of proximal splice sites as described (15,64,65) and suggest that the observed activation of atypical distal 5' ss can be explained by intronic length constraints.
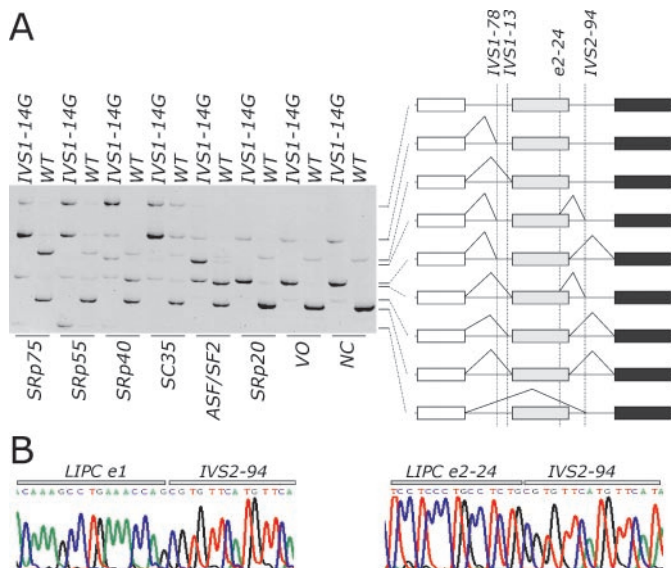
Figure 6. The influence of SR proteins on utilization of aberrant 3′ and 5′ ss in *LIPC*. (**A**) Wild-type and mutated splicing reporters were co-transfected with plasmids expressing the indicated SR proteins. Splicing reporters are shown at the top and SR proteins are indicated at the bottom. VO, vector only control, in which an empty pCG vector was co-transfected with the wild-type and mutated reporter constructs; NC, no co-transfection (reporter only) control. The corresponding *LIPC* isoforms are shown on the right side. (**B**) Nucleotide sequences surrounding aberrant splice sites induced by SR proteins. Exons (e) are shown as a grey rectangle, introns (IVS) are indicated by white rectangles.

DISCUSSION

In this study, we have undertaken the largest compilation of aberrant 3′ss in human disease genes to date. Although cryptic 3′ss are less frequent than cryptic 5′ ss (9), the size of the compiled dataset was comparable with that analysed earlier for cryptic 5′ ss (11). The overall number of aberrant 3′ss was marginally higher in exons than in introns (Table 1), but reports of AG-creating PPT mutations, in which RNA products were not sequenced and were thus not included in our study, appear to be more frequent in the literature (66–71) than those in exons. This suggests that aberrant 3′ss described in the literature are distributed in exons and introns with approximately equal frequencies, consistent with the observed low median distance between all aberrant and authentic 3′ss (Table 1). In contrast, if a mutation affects 3′YAG, cryptic 3′ss are ∼3 times more likely to occur in exons than in introns. Conversely, aberrant 3′ss due to mutations outside 3′YAG are ∼3 times more frequent in introns than in exons. The relative intronic depletion of the former can be accounted for by a lower prior probability of finding an alternative 3′AG in the PPT (Figure 1A and B), whereas intronic over-representation of the latter is due to a better 3′ss consensus created in the pyrimidine-rich sequence. Sequence constraints that limit the number of cryptic and *de novo* 3′ss also explain a lower total number of aberrant 3′ss in the literature as compared with aberrant 5′ ss. However, since recognition sequences at the 3′ ss (YAG, PPT and BPS) are more complex than at the 5′ss and spread over a longer and more variable distance, mutations affecting 3′ss may have multiple effects that make it difficult to categorize cryptic or *de novo* sites unambiguously. For example, apart from a direct AG creation in the PPT, pyrimidine to

purine substitutions may weaken the PPT of authentic 3′ss and thus contribute to their repression. Similarly, mutations of the 3′ YAG (such as IVS-1G>A, Supplementary Table 4) may potentially improve a match to the BPS consensus and promote cryptic 3′ss activation in the downstream exon if a suitable YAG is available within an optimal distance from the newly created BPS. We cannot entirely exclude that this can contribute to the observed bias (Figure 1A and C); nevertheless, it is likely that their distribution can be fully explained by a lack of AGs and pyrimidine residues upstream and downstream of authentic 3′ss, respectively.

Although the number of reported cryptic 3′ss in introns was low (*n* = 18), their proportion in terminal introns (33%) (7,18,72–74) appeared to be higher than in other categories of aberrant 3′ss (Table 1), pointing to possible involvement of 3′ end processing factors. Two *de novo* 3′ss observed in terminal exons were due to a transversion (36) and transition (37) in position –6 of the new intron creating uridine residues that improve PPT recognition. Purines are greatly underrepresented in this PPT position in several species (42) and both transitions and transversions of uridine –6 have been shown to increase exon skipping and/or retention of weakly spliced introns (15,46). These splicing defects are likely to result from diminished interaction of the PPT with the first RNA recognition motif of U2AF^{65} (75), which has been shown to promote 3′ end processing (76,77).

The exact mechanism by which the 3′ss is recognized is not well understood. Although the first AG downstream from the BPS is usually selected for exon ligation, it is unclear why it is sometimes not the case and how exactly this is achieved. A 'scanning mechanism' of 3′ss selection (59,60) postulated a unidirectional, linear search that initiates at the BPS and continues until a suitable AG is selected. This model is consistent with blocking the second step on the hairpin structures inserted between BP and 3′AG (60,78). However, closely spaced, tandemly arranged AGs can efficiently compete when placed 13–22 nt downstream of the *Saccharomyces cerevisiae* branchpoint (61) and 15–23 nt in a metazoan substrate (20), although not when placed at least 35 nt from a mammalian BP (78). Frequent cryptic 3′ss in exons would support the scanning model (79), but our finding that cryptic 3′ss cluster in ∼20 nt region just downstream of authentic 3′ss (Figure 1A) provides no support *per se* for this concept since their underrepresentation in introns can be explained by AG depletion in the PPT. Several distant exonic cryptic 3′ss, such as *NF1* E28+293 (293 nt from the start of exon 28) activated by the IVS27b-2A>T mutation (80) or *MANBA* E16+172 induced by the IVS15-2A>G transition (81), had a strong 3′ss upstream of mutated 3′AGs, but they were not used despite being much closer to the BPS than exonic cryptic 3′ss. These cases raise a speculation that exons could first be scanned for downstream AGs to completion before AGs upstream of authentic 3′ss are considered. The concomitant activation of cryptic 3′ss (Figure 3B and C) or 5′ ss (82) in both the intron and the exon indicates that inactivation of genuine splice sites eventually triggers a search for the most suitable splice sites in both directions, consistent with a general finding that more distant AGs compete less efficiently (Figure 1).

Using well-documented cases of AG-creating mutations in the PPT that repress authentic 3′ss and activate aberrant 3′ss upstream of BPS, we have illustrated the importance of the

'BPS-new AG' distance and/or the strength of the new PPT for the expression of correctly spliced mRNAs. Interestingly, this distance (Figure 3A) was similar to that reported previously for AGs that were outcompeted by a downstream AG (62,83) and is consistent with steric interference with a factor(s) bound to this region. Newly created AGs are likely to recruit spliceosome components that compete for interactions with authentic 3′ss, are partially recognized by essential splicing factors, such as U2 snRNP (84), but may not assemble fully functional splicing complexes (83,85). Utilization of such competing AGs has been shown to be affected by hSlu7 (83) and SPF45 (86). Future studies should examine interactions of human homologues of yeast Prp18 and Prp22 with the reporter pre-mRNAs described here, since these proteins are required only as the BP-AG distance increases (87,88). Unexpected promotion of upstream 3′ss observed for the *LIPC* -14G/-13G construct (Figure 4B, lane 6) might be addressed by examining contacts between the first nucleotide of the new exon and U2AF^{35} (89), U5 snRNA (90) or PRP8 (91).

As with cryptic 5′ ss (11), the average S&S scores of cryptic 3′ss were significantly lower than for corresponding authentic 3′ss and higher than for mutated sites (Table 1). This indicates that intrinsic differences in the consensus 3′ss sequences contribute to the cryptic 3′ss remaining completely silent in the presence of a wild-type authentic site. How such cryptic 3′ss are efficiently inhibited is poorly understood, but nucleotides that precede AGs and similar AG-AG measuring mechanisms may also play an important and more general role in auxiliary splicing sequences (63). In contrast to cryptic 3′ss, the S&S scores of *de novo* sites were similar to authentic sites (Table 1), suggesting that they lack a predictive value in these cases. To explore the significance of S&S scores in selection of cryptic 3′ss in exon versus intron, we paired S&S scores of intronic cryptic 3′ss with the best match to the 3′ss in adjacent exons (Supplementary Table 5), and *vice versa* (Supplementary Table 6). Each intronic cryptic 3′ss had a higher S&S score than the best corresponding exonic YAG (Supplementary Table 5). The average difference between intronic cryptic 3′ss and the best match to 3′ss in exons was high (14.0; $S\&S_i = 83.8$ versus $S\&S_{be} = 69.8$; where i stands for intron and be is the best S&S score in adjacent downstream exons). Several exons lacked the YAG consensus altogether, such as exon 10 of *TP53* (92), *APOE* exon 4 (73) and *ELN* exon 16 (93), or were very short, such as 27 nt exon 32 of *COL17A1* (17) or 57 nt exon 14 of *COL5A1* (94), and likely to be poorly defined (95). In contrast, as much as ∼45% of the putative 3′ ss in an adjacent intronic sequence had S&S scores higher than the corresponding cryptic 3′ss in exons (average $S\&S_e = 73.8$ versus $S\&S_{bi} = 71.1$; where e is exon and bi are the best scores in a 100 nt of the upstream intron), with the average difference only 2.7. The high differential between the S&S scores for intronic cryptic 3′ss and their putative exon counterparts, which apparently reflects a lack of Y-runs in exons, should be considered when using this and other algorithms for prediction of competing 3′ss.

In summary, our results improve prediction of aberrant 3′ss localization in human disease genes and provide a valuable resource for studying the mechanisms of 3′ss selection. They also suggest that inspection of genetic alterations in or near the 3′ss in their sequence context may in some cases facilitate prediction of their splicing and phenotypic outcomes.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Burge,C.B., Tuschl,T. and Sharp,P.A. (1999) Splicing of precursors to mRNAs by the spliceosome. In Gesteland,R.F., Cech,T.R. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, New York, pp. 525–560.
2. Fairbrother,W.G., Yeh,R.F., Sharp,P.A. and Burge,C.B. (2002) Predictive identification of exonic splicing enhancers in human genes. *Science*, **297**, 1007–1013.
3. Zhang,X.H. and Chasin,L.A. (2004) Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev*., **18**, 1241–1250.
4. Cartegni,L., Chew,S.L. and Krainer,A.R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Rev. Genet*., **3**, 285–298.
5. Wang,Z., Rolish,M.E., Yeo,G., Tung,V., Mawson,M. and Burge,C.B. (2004) Systematic identification and analysis of exonic splicing silencers. *Cell*, **119**, 831–845.
6. Teraoka,S.N., Telatar,M., Becker-Catania,S., Liang,T., Onengut,S., Tolun,A., Chessa,L., Sanal,O., Bernatowska,E., Gatti,R.A. *et al*. (1999) Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences. *Am. J. Hum. Genet*., **64**, 1617–1631.
7. Ars,E., Serra,E., Garcia,J., Kruyer,H., Gaona,A., Lazaro,C. and Estivill,X. (2000) Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. *Hum. Mol. Genet*., **9**, 237–247.
8. Lopez-Bigas,N., Audit,B., Ouzounis,C., Parra,G. and Guigo,R. (2005) Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett*., **579**, 1900–1903.
9. Nakai,K. and Sakamoto,H. (1994) Construction of a novel database containing aberrant splicing mutations of mammalian genes. *Gene*, **141**, 171–177.
10. Krawczak,M., Reiss,J. and Cooper,D.N. (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum. Genet*., **90**, 41–54.
11. Roca,X., Sachidanandam,R. and Krainer,A.R. (2003) Intrinsic differences between authentic and cryptic 5′ splice sites. *Nucleic Acids Res*., **31**, 6321–6333.
12. Shapiro,M.B. and Senapathy,P. (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res*., **15**, 7155–7174.
13. Senapathy,P., Shapiro,M.B. and Harris,N.L. (1990) Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol*., **183**, 252–278.
14. Kol,G., Lev-Maor,G. and Ast,G. (2005) Human-mouse comparative analysis reveals that branch-site plasticity contributes to splicing regulation. *Hum. Mol. Genet*., **14**, 1559–1568.

15. Královičová,J., Houngninou-Molango,S., Krämer,A. and Vořechovský,I. (2004) Branch sites haplotypes that control alternative splicing. *Hum. Mol. Genet.*, **13**, 3189–3202.

16. Vořechovský,I., Luo,L., Dyer,M.J., Catovsky,D., Amlot,P.L., Yaxley,J.C., Foroni,L., Hammarström,L., Webster,A.D. and Yuille,M.A. (1997) Clustering of missense mutations in the ataxia-telangiectasia gene in a sporadic T-cell leukaemia. *Nature Genet.*, **17**, 96–99.

17. Darling,T.N., Yee,C., Koh,B., McGrath,J.A., Bauer,J.W., Uitto,J., Hintner,H. and Yancey,K.B. (1998) Cycloheximide facilitates the identification of aberrant transcripts resulting from a novel splice-site mutation in *COL17A1* in a patient with generalized atrophic benign epidermolysis bullosa. *J. Invest. Dermatol.*, **110**, 165–169.

18. Otterness,D.M., Szumlanski,C.L., Wood,T.C. and Weinshilboum,R.M. (1998) Human thiopurine methyltransferase pharmacogenetics. Kindred with a terminal exon splice junction mutation that results in loss of activity. *J. Clin. Invest.*, **101**, 1036–1044.

19. Fujimaru,M., Tanaka,A., Choeh,K., Wakamatsu,N., Sakuraba,H. and Isshiki,G. (1998) Two mutations remote from an exon/intron junction in the β-hexosaminidase β-subunit gene affect 3′-splice site selection and cause Sandhoff disease. *Hum. Genet.*, **103**, 462–469.

20. Chua,K. and Reed,R. (2001) An upstream AG determines whether a downstream AG is selected during catalytic step II of splicing. *Mol. Cell. Biol.*, **21**, 1509–1514.

21. Mount,S.M. (1982) A catalogue of splice junction sequences. *Nucleic Acids Res.*, **10**, 459–472.

22. Hebsgaard,S.M., Korning,P.G., Tolstrup,N., Engelbrecht,J., Rouze,P. and Brunak,S. (1996) Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. *Nucleic Acids Res.*, **24**, 3439–3452.

23. Bruggenwirth,H.T., Boehmer,A.L., Ramnarain,S., Verleun-Mooijman,M.C., Satijn,D.P., Trapman,J., Grootegoed,J.A. and Brinkmann,A.O. (1997) Molecular analysis of the androgen-receptor gene in a family with receptor-positive partial androgen insensitivity: an unusual type of intronic mutation. *Am. J. Hum. Genet.*, **61**, 1067–1077.

24. Nakano,T. and Suzuki,K. (1989) Genetic cause of a juvenile form of Sandhoff disease. Abnormal splicing of β-hexosaminidase β chain gene transcript due to a point mutation within intron 12. *J. Biol. Chem.*, **264**, 5155–5158.

25. Ars,E., Kruyer,H., Morell,M., Pros,E., Serra,E., Ravella,A., Estivill,X. and Lazaro,C. (2003) Recurrent mutations in the *NF1* gene are common among neurofibromatosis type 1 patients. *J. Med. Genet.*, **40**, e82.

26. Aksentijevich,I., Galon,J., Soares,M., Mansfield,E., Hull,K., Oh,H.H., Goldbach-Mansky,R., Dean,J., Athreya,B., Reginato,A.J. *et al.* (2001) The tumor-necrosis-factor receptor-associated periodic syndrome: new mutations in *TNFRSF1A*, ancestral origins, genotype-phenotype studies, and evidence for further genetic heterogeneity of periodic fevers. *Am. J. Hum. Genet.*, **69**, 301–314.

27. Eskesen,S.T., Eskesen,F.N. and Ruvinsky,A. (2004) Natural selection affects frequencies of AG and GT dinucleotides at the 5′ and 3′ ends of exons. *Genetics*, **167**, 543–550.

28. Asselta,R., Duga,S., Spena,S., Peyvandi,F., Castaman,G., Malcovati,M., Mannucci,P.M. and Tenchini,M.L. (2004) Missense or splicing mutation? The case of a fibrinogen B^β-chain mutation causing severe hypofibrinogenemia *Blood*, **103**, 3051–3054.

29. Pohlenz,J., Rosenthal,I.M., Weiss,R.E., Jhiang,S.M., Burant,C. and Refetoff,S. (1998) Congenital hypothyroidism due to mutations in the sodium/iodide symporter. Identification of a nonsense mutation producing a downstream cryptic 3′ splice site. *J. Clin. Invest.*, **101**, 1028–1035.

30. Pomponio,R.J., Reynolds,T.R., Mandel,H., Admoni,O., Melone,P.D., Buck,G.A. and Wolf,B. (1997) Profound biotinidase deficiency caused by a point mutation that creates a downstream cryptic 3′ splice acceptor site within an exon of the human biotinidase gene. *Hum. Mol. Genet.*, **6**, 739–745.

31. Maugeri,A., van Driel,M.A., van de Pol,D.J., Klevering,B.J., van Haren,F.J., Tijmes,N., Bergen,A.A., Rohrschneider,K., Blankenagel,A., Pinckers,A.J. *et al.* (1999) The 2588G>C mutation in the *ABCR* gene is a mild frequent founder mutation in the Western European population and allows the classification of *ABCR* mutations in patients with Stargardt disease. *Am. J. Hum. Genet.*, **64**, 1024–1035.

32. Tamouza,R., El Kassar,N., Schaeffer,V., Carbonnelle,E., Tatari,Z., Marzais,F., Fortier,C., Poirier,J.C., Sadki,K., Bernaudin,F. *et al.* (2000) A novel *HLA*-B*39 allele (*HLA*-B*39*16*) due to a rare mutation causing cryptic splice site activation. *Hum. Immunol.*, **61**, 467–473.

33. Bruce,L.J., Ghosh,S., King,M.J., Layton,D.M., Mawby,W.J., Stewart,G.W., Oldenborg,P.A., Delaunay,J. and Tanner,M.J. (2002) Absence of CD47 in protein 4.2-deficient hereditary spherocytosis in man: an interaction between the Rh complex and the band 3 complex. *Blood*, **100**, 1878–1885.

34. Tavassoli,K., Eigel,A., Dworniczak,B., Valtseva,E. and Horst,J. (1998) Identification of four novel mutations in the factor VIII gene: three missense mutations (E1875G, G2088S, I2185T) and a 2-bp deletion (1780delTC). *Hum. Mutat.*, **Suppl. 1**, S260–S262.

35. Nakamura,K., Fukao,T., Perez-Cerda,C., Luque,C., Song,X.Q., Naiki,Y., Kohno,Y., Ugarte,M. and Kondo,N. (2001) A novel single-base substitution (380C>T) that activates a 5-base downstream cryptic splice-acceptor site within exon 5 in almost all transcripts in the human mitochondrial acetoacetyl-CoA thiolase gene. *Mol. Genet. Metab.*, **72**, 115–121.

36. Kajihara,S., Matsuhashi,S., Yamamoto,K., Kido,K., Tsuji,K., Tanae,A., Fujiyama,S., Itoh,T., Tanigawa,K., Uchida,M. *et al.* (1995) Exon redefinition by a point mutation within exon 5 of the glucose-6-phosphatase gene is the major cause of glycogen storage disease type 1a in Japan. *Am. J. Hum. Genet.*, **57**, 549–555.

37. Hasegawa,Y., Kawame,H., Ida,H., Ohashi,T. and Eto,Y. (1994) Single exon mutation in arylsulfatase A gene has two effects: loss of enzyme activity and aberrant splicing. *Hum. Genet.*, **93**, 415–420.

38. Tavassoli,K., Eigel,A., Pollmann,H. and Horst,J. (1997) Mutational analysis of ectopic factor VIII transcripts from hemophilia A patients: identification of cryptic splice site, exon skipping and novel point mutations. *Hum. Genet.*, **100**, 508–511.

39. Mayer,K., Ballhausen,W., Leistner,W. and Rott,H. (2000) Three novel types of splicing aberrations in the tuberous sclerosis *TSC2* gene caused by mutations apart from splice consensus sequences. *Biochim. Biophys. Acta*, **1502**, 495–507.

40. Lublin,D.M., Mallinson,G., Poole,J., Reid,M.E., Thompson,E.S., Ferdman,B.R., Telen,M.J., Anstee,D.J. and Tanner,M.J. (1994) Molecular basis of reduced or absent expression of decay-accelerating factor in Cromer blood group phenotypes. *Blood*, **84**, 1276–1282.

41. Chavanas,S., Gache,Y., Vailly,J., Kanitakis,J., Pulkkinen,L., Uitto,J., Ortonne,J. and Meneguzzi,G. (1999) Splicing modulation of integrin beta4 pre-mRNA carrying a branch point mutation underlies epidermolysis bullosa with pyloric atresia undergoing spontaneous amelioration with ageing. *Hum. Mol. Genet.*, **8**, 2097–2105.

42. Yeo,G., Hoon,S., Venkatesh,B. and Burge,C.B. (2004) Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc. Natl Acad. Sci. USA*, **101**, 15000–15005.

43. Hagstrom,S.A., North,M.A., Nishina,P.L., Berson,E.L. and Dryja,T.P. (1998) Recessive mutations in the gene encoding the tubby-like protein TULP1 in patients with retinitis pigmentosa. *Nature Genet.*, **18**, 174–176.

44. Gu,S., Lennon,A., Li,Y., Lorenz,B., Fossarello,M., North,M., Gal,A. and Wright,A. (1998) Tubby-like protein-1 mutations in autosomal recessive retinitis pigmentosa. *Lancet*, **351**, 1103–1104.

45. Lee,P.L., Barton,J.C., Brandhagen,D. and Beutler,E. (2004) Hemojuvelin (*HJV*) mutations in persons of European, African-American and Asian ancestry with adult onset haemochromatosis. *Br. J. Haematol.*, **127**, 224–229.

46. Claes,K., Vandesompele,J., Poppe,B., Dahan,K., Coene,I., De Paepe,A. and Messiaen,L. (2002) Pathological splice mutations outside the invariant AG/GT splice sites of *BRCA1* exon 5 increase alternative transcript levels in the 5′ end of the *BRCA1* gene. *Oncogene*, **21**, 4171–4175.

47. Vafiadaki,E., Cooper,A., Heptinstall,L.E., Hatton,C.E., Thornley,M. and Wraith,J.E. (1998) Mutation analysis in 57 unrelated patients with MPS II (Hunter's disease). *Arch. Dis. Child.*, **79**, 237–241.

48. Tuffery-Giraud,S., Chambert,S., Demaille,J. and Claustres,M. (1999) Point mutations in the dystrophin gene: evidence for frequent use of cryptic splice sites as a result of splicing defects. *Hum. Mutat.*, **14**, 359–368.

49. Wakamatsu,N., Kobayashi,H., Miyatake,T. and Tsuji,S. (1992) A novel exon mutation in the human beta-hexosaminidase beta subunit gene affects 3′ splice site selection. *J. Biol. Chem.*, **267**, 2406–2413.

50. Yenchitsomanus,P., Thanootarakul,P., Akkarapatumwong,V., Oranwiroon,S., Pung-Amritt,P., Veerakul,G. and Mahasandana,C. (2001) Mutation causing exon 15 skipping and partial exon 16 deletion in

factor VIII transcript, and a method for direct mutation detection. *Haemophilia*, **7**, 335–338.

51. Danckwardt,S., Neu-Yilik,G., Thermann,R., Frede,U., Hentze,M.W. and Kulozik,A.E. (2002) Abnormally spliced β-globin mRNAs: a single point mutation generates transcripts sensitive and insensitive to nonsense-mediated mRNA decay. *Blood*, **99**, 1811–1816.

52. Brand,K., Dugi,K.A., Brunzell,J.D., Nevin,D.N. and Santamarina-Fojo,S. (1996) A novel A>G mutation in intron I of the hepatic lipase gene leads to alternative splicing resulting in enzyme deficiency. *J. Lipid Res.*, **37**, 1213–1223.

53. Putnam,E.A., Park,E.S., Aalfs,C.M., Hennekam,R.C. and Milewicz,D.M. (1997) Parental somatic and germ-line mosaicism for a FBN2 mutation and analysis of FBN2 transcript levels in dermal fibroblasts. *Am. J. Hum. Genet.*, **60**, 818–827.

54. Query,C.C., Strobel,S.A. and Sharp,P.A. (1996) Three recognition events at the branch-site adenine. *EMBO J.*, **15**, 1392–1402.

55. Coolidge,C.J., Seely,R.J. and Patton,J.G. (1997) Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res.*, **25**, 888–896.

56. Bouck,J., Fu,X.D., Skalka,A.M. and Katz,R.A. (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.

57. Janssen,R.J., Wevers,R.A., Haussler,M., Luyten,J.A., Steenbergen-Spanjers,G.C., Hoffmann,G.F., Nagatsu,T. and Van den Heuvel,L.P. (2000) A branch site mutation leading to aberrant splicing of the human tyrosine hydroxylase gene in a child with a severe extrapyramidal movement disorder. *Ann. Hum. Genet.*, **64**, 375–382.

58. Hiller,M., Huse,K., Szafranski,K., Jahn,N., Hampe,J., Schreiber,S., Backofen,R. and Platzer,M. (2004) Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. *Nature Genet.*, **36**, 1255–1257.

59. Smith,C.W., Porro,E.B., Patton,J.G. and Nadal-Ginard,B. (1989) Scanning from an independently specified branch point defines the 3′ splice site of mammalian introns. *Nature*, **342**, 243–247.

60. Smith,C.W., Chu,T.T. and Nadal-Ginard,B. (1993) Scanning and competition between AGs are involved in 3′ splice site selection in mammalian introns. *Mol. Cell. Biol.*, **13**, 4939–4952.

61. Luukkonen,B.G. and Seraphin,B. (1997) The role of branchpoint-3′ splice site spacing and interaction between intron terminal nucleotides in 3′ splice site selection in *Saccharomyces cerevisiae*. *EMBO J.*, **16**, 779–792.

62. Chiara,M.D., Palandjian,L., Feld Kramer,R. and Reed,R. (1997) Evidence that U5 snRNP recognizes the 3′ splice site for catalytic step II in mammals. *EMBO J.*, **16**, 4746–4759.

63. Lei,H. and Vořechovský,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.

64. Fu,X.D., Mayeda,A., Maniatis,T. and Krainer,A.R. (1992) General splicing factors SF2 and SC35 have equivalent activities *in vitro*, and both affect alternative 5′ and 3′ splice site selection. *Proc. Natl Acad. Sci. USA*, **89**, 11224–11228.

65. Krainer,A.R., Conway,G.C. and Kozak,D. (1990) The essential pre-mRNA splicing factor SF2 influences 5′ splice site selection by activating proximal sites. *Cell*, **62**, 35–42.

66. Zschocke,J. and Hoffmann,G.F. (1999) Phenylketonuria mutations in Germany. *Hum. Genet.*, **104**, 390–398.

67. Amsellem,S., Briffaut,D., Carrie,A., Rabes,J.P., Girardet,J.P., Fredenrich,A., Moulin,P., Krempf,M., Reznik,Y., Vialettes,B. *et al.* (2002) Intronic mutations outside of Alu-repeat-rich domains of the LDL receptor gene are a cause of familial hypercholesterolemia. *Hum. Genet.*, **111**, 501–510.

68. Du,Y.Z., Srivastava,A.K. and Schwartz,C.E. (1998) Multiple exon screening using restriction endonuclease fingerprinting (REF): detection of six novel mutations in the L1 cell adhesion molecule (*L1CAM*) gene. *Hum. Mutat.*, **11**, 222–230.

69. Fahsold,R., Hoffmeyer,S., Mischung,C., Gille,C., Ehlers,C., Kucukceylan,N., Abdel-Nour,M., Gewies,A., Peters,H., Kaufmann,D. *et al.* (2000) Minor lesion mutational spectrum of the entire *NF1* gene does not explain its high mutability but points to a functional domain upstream of the GAP-related domain. *Am. J. Hum. Genet.*, **66**, 790–818.

70. Korkko,J., Ala-Kokko,L., De Paepe,A., Nuytinck,L., Earley,J. and Prockop,D.J. (1998) Analysis of the COL1A1 and COL1A2 genes by PCR amplification and scanning by conformation-sensitive gel electrophoresis identifies only *COL1A1* mutations in 15 patients with osteogenesis imperfecta type I: identification of common sequences of null-allele mutations. *Am. J. Hum. Genet.*, **62**, 98–110.

71. Bosco,P., Cali,F., Meli,C., Mollica,F., Zammarchi,E., Cerone,R., Vanni,C., Palillo,L., Greco,D. and Romano,V. (1998) Eight new mutations of the phenylalanine hydroxylase gene in Italian patients with hyperphenylalaninemia. *Hum. Mutat.*, **11**, 240–243.

72. Marchetti,C., Patriarca,P., Solero,G.P., Baralle,F.E. and Romano,M. (2004) Genetic studies on myeloperoxidase deficiency in Italy. *Jpn. J. Infect. Dis.*, **57**, S10–S12.

73. Cladaras,C., Hadzopoulou-Cladaras,M., Felber,B.K., Pavlakis,G. and Zannis,V.I. (1987) The molecular basis of a familial ApoE deficiency. An acceptor splice site mutation in the third intron of the deficient apoE gene. *J. Biol. Chem.*, **262**, 2310–2315.

74. Antonarakis,S.E., Irkin,S.H., Cheng,T.C., Scott,A.F., Sexton,J.P., Trusko,S.P., Charache,S. and Kazazian,H.H.Jr (1984) β-thalassemia in American blacks: novel mutations in the 'TATA' box and an acceptor splice site. *Proc. Natl Acad. Sci. USA*, **81**, 1154–1158.

75. Banerjee,H., Rahn,A., Davis,W. and Singh,R. (2003) Sex lethal and U2 small nuclear ribonucleoprotein auxiliary factor (U2AF^{65}) recognize polypyrimidine tracts using multiple modes of binding. *RNA*, **9**, 88–99.

76. Vagner,S., Vagner,C. and Mattaj,I.W. (2000) The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF^{65} to couple 3′-end processing and splicing. *Genes Dev.*, **14**, 403–413.

77. Millevoi,S., Geraghty,F., Idowu,B., Tam,J.L., Antoniou,M. and Vagner,S. (2002) A novel function for the U2AF^{65} splicing factor in promoting pre-mRNA 3′-end processing. *EMBO Rep.*, **3**, 869–874.

78. Chen,S., Anderson,K. and Moore,M.J. (2000) Evidence for a linear search in bimolecular 3′ splice site AG selection. *Proc. Natl Acad. Sci. USA*, **97**, 593–598.

79. Yamanaka,T., Yada,T., Takagi,T. and Nakai,K. (2001) *RECOMB2000. The Fourth Annual International Conference on Computational Molecular Biology April 8–11 2000*. Tokyo, Japan, pp. 82–83.

80. Messiaen,L.M., Callens,T., Mortier,G., Beysen,D., Vandenbroucke,I., Van Roy,N., Speleman,F. and Paepe,A.D. (2000) Exhaustive mutation analysis of the *NF1* gene allows identification of 95% of mutations and reveals a high frequency of unusual splicing defects. *Hum. Mutat.*, **15**, 541–555.

81. Alkhayat,A.H., Kraemer,S.A., Leipprandt,J.R., Macek,M., Kleijer,W.J. and Friderici,K.H. (1998) Human beta-mannosidase cDNA characterization and first identification of a mutation associated with human β-mannosidosis. *Hum. Mol. Genet.*, **7**, 75–83.

82. De Jonghe,C., Cruts,M., Rogaeva,E.A., Tysoe,C., Singleton,A., Vanderstichele,H., Meschino,W., Dermaut,B., Vanderhoeven,I., Backhovens,H. *et al.* (1999) Aberrant splicing in the presenilin-1 intron 4 mutation causes presenile Alzheimer's disease by increased Abeta42 secretion. *Hum. Mol. Genet.*, **8**, 1529–1540.

83. Chua,K. and Reed,R. (1999) The RNA splicing factor hSlu7 is required for correct 3′ splice-site choice. *Nature*, **402**, 207–210.

84. Kan,J.L. and Green,M.R. (1999) Pre-mRNA splicing of IgM exons M1 and M2 is directed by a juxtaposed splicing enhancer and inhibitor. *Genes Dev.*, **13**, 462–471.

85. Nemeroff,M.E., Utans,U., Krämer,A. and Krug,R.M. (1992) Identification of cis-acting intron and exon regions in influenza virus NS1 mRNA that inhibit splicing and cause the formation of aberrantly sedimenting presplicing complexes. *Mol. Cell. Biol.*, **12**, 962–970.

86. Lallena,M.J., Chalmers,K.J., Llamazares,S., Lamond,A.I. and Valcarcel,J. (2002) Splicing regulation at the second catalytic step by Sex-lethal involves 3′ splice site recognition by SPF45. *Cell*, **109**, 285–296.

87. Schwer,B. and Gross,C.H. (1998) Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *EMBO J.*, **17**, 2086–2094.

88. Zhang,X. and Schwer,B. (1997) Functional and physical interaction between the yeast splicing factors Slu7 and Prp18. *Nucleic Acids Res.*, **25**, 2146–2152.

89. Guth,S., Tange,T.O., Kellenberger,E. and Valcarcel,J. (2001) Dual function for U2AF(35) in AG-dependent pre-mRNA splicing. *Mol. Cell. Biol.*, **21**, 7673–7681.

90. Newman,A.J., Teigelkamp,S. and Beggs,J.D. (1995) snRNA interactions at 5′ and 3′ splice sites monitored by photoactivated crosslinking in yeast spliceosomes. *RNA*, **1**, 968–980.

91. Teigelkamp,S., Newman,A.J. and Beggs,J.D. (1995) Extensive interactions of PRP8 protein with the 5′ and 3′ splice sites during splicing

suggest a role in stabilization of exon alignment by U5 snRNA. *EMBO J.*, **14**, 2602–2612.

92. Verselis,S.J., Rheinwald,J.G., Fraumeni,J.F.Jr and Li,F.P. (2000) Novel *p53* splice site mutations in three families with Li-Fraumeni syndrome. *Oncogene*, **19**, 4230–4235.

93. Urban,Z., Michels,V.V., Thibodeau,S.N., Donis-Keller,H., Csiszar,K. and Boyd,C.D. (1999) Supravalvular aortic stenosis: a splice site mutation within the elastin gene results in reduced expression of two aberrantly spliced transcripts. *Hum. Genet.*, **104**, 135–142.

94. Bouma,P., Cabral,W.A., Cole,W.G. and Marini,J.C. (2001) *COL5A1* exon 14 splice acceptor mutation causes a functional null allele, haploinsufficiency of alpha 1(V) and abnormal heterotypic interstitial fibrils in Ehlers-Danlos syndrome II. *J. Biol. Chem.*, **276**, 13356–13364.

95. Berget,S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.*, **270**, 2411–2414.

96. Schollen,E., Frank,C.G., Keldermans,L., Reyntjens,R., Grubenmann,C.E., Clayton,P.T., Winchester,B.G., Smeitink,J., Wevers,R.A., Aebi,M. *et al.* (2004) Clinical and molecular features of three patients with congenital disorders of glycosylation type Ih (CDG-Ih) (*ALG8* deficiency). *J. Med. Genet.*, **41**, 550–556.

97. Kobayashi,K., Jackson,M.J., Tick,D.B., O'Brien,W.E. and Beaudet,A.L. (1990) Heterogeneity of mutations in argininosuccinate synthetase causing human citrullinemia. *J. Biol. Chem.*, **265**, 11361–11367.

98. Su,T.S. and Lin,L.H. (1990) Analysis of a splice acceptor site mutation which produces multiple splicing abnormalities in the human argininosuccinate synthetase locus. *J. Biol. Chem.*, **265**, 19716–19720.

99. Claes,K., Poppe,B., Machackova,E., Coene,I., Foretova,L., De Paepe,A. and Messiaen,L. (2003) Differentiating pathogenic mutations from polymorphic alterations in the splice sites of *BRCA1* and *BRCA2*. *Genes. Chromosomes Cancer*, **37**, 314–320.

100. Song,W.J., Sullivan,M.G., Legare,R.D., Hutchings,S., Tan,X., Kufrin,D., Ratajczak,J., Resende,I.C., Haworth,C., Hock,R. *et al.* (1999) Haploinsufficiency of *CBFA2* causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nature Genet.*, **23**, 166–175.

101. Weaving,L.S., Christodoulou,J., Williamson,S.L., Friend,K.L., McKenzie,O.L., Archer,H., Evans,J., Clarke,A., Pelka,G.J., Tam,P.P. *et al.* (2004) Mutations of *CDKL5* cause a severe neurodevelopmental disorder with infantile spasms and mental retardation. *Am. J. Hum. Genet.*, **75**, 1079–1093.

102. Munroe,P.B., Mitchison,H.M., O'Rawe,A.M., Anderson,J.W., Boustany,R.M., Lerner,T.J., Taschner,P.E., de Vos,N., Breuning,M.H., Gardiner,R.M. *et al.* (1997) Spectrum of mutations in the Batten disease gene, *CLN3*. *Am. J. Hum. Genet.*, **61**, 310–316.

103. Hennies,H.C., Rauch,A., Seifert,W., Schumi,C., Moser,E., Al-Taji,E., Tariverdian,G., Chrzanowska,K.H., Krajewska-Walasek,M., Rajab,A. *et al.* (2004) Allelic heterogeneity in the *COH1* gene explains clinical variability in Cohen syndrome. *Am. J. Hum. Genet.*, **75**, 138–145.

104. Pasmooij,A.M., van der Steege,G., Pas,H.H., Smitt,J.H., Nijenhuis,A.M., Zuiderveen,J. and Jonkman,M.F. (2004) Features of epidermolysis bullosa simplex due to mutations in the ectodomain of type XVII collagen. *Br. J. Dermatol.*, **151**, 669–674.

105. Chiodo,A.A., Hockey,A. and Cole,W.G. (1992) A base substitution at the splice acceptor site of intron 5 of the *COL1A2* gene activates a cryptic splice site within exon 6 and generates abnormal type I procollagen in a patient with Ehlers-Danlos syndrome type VII. *J. Biol. Chem.*, **267**, 6361–6369.

106. Tromp,G. and Prockop,D.J. (1988) Single base mutation in the pro alpha 2(I) collagen gene that causes efficient splicing of RNA from exon 27 to exon 29 and synthesis of a shortened but in-frame pro alpha 2(I) chain. *Proc. Natl Acad. Sci. USA*, **85**, 5254–5258.

107. Williams,C.J., Ganguly,A., Considine,E., McCarron,S., Prockop,D.J., Walsh-Vockley,C. and Michels,V.V. (1996) A-2→G transition at the 3′ acceptor splice site of IVS17 characterizes the *COL2A1* gene mutation in the original Stickler syndrome kindred. *Am. J. Med. Genet.*, **63**, 461–467.

108. Takahara,K., Schwarze,U., Imamura,Y., Hoffman,G.G., Toriello,H., Smith,L.T., Byers,P.H. and Greenspan,D.S. (2002) Order of intron removal influences multiple splice outcomes, including a two-exon skip, in a *COL5A1* acceptor-site mutation that results in abnormal pro-alpha1(V) N-propeptides and Ehlers-Danlos syndrome type I. *Am. J. Hum. Genet.*, **71**, 451–465.

109. Adachi,K., Takeshima,Y., Wada,H., Yagi,M., Nakamura,H. and Matsuo,M. (2003) Heterogous dystrophin mRNA produced by a novel splice acceptor site mutation in intermediate dystrophinopathy. *Pediatr. Res.*, **53**, 125–131.

110. Efferth,T., Bachli,E.B., Schwarzl,S.M., Goede,J.S., West,C., Smith,J.C. and Beutler,E. (2004) Glucose-6-phosphate dehydrogenase (G6PD) deficiency-type Zurich: a splice site mutation as an uncommon mechanism producing enzyme deficiency. *Blood*, **104**, 2608.

111. McCarthy,E.M. and Phillips,J.A.,3rd (1998) Characterization of an intron splice enhancer that regulates alternative splicing of human GH pre-mRNA. *Hum. Mol. Genet.*, **7**, 1491–1496.

112. Yokoi,T., Shinoda,K., Ohno,I., Kato,K., Miyawaki,T. and Taniguchi,N. (1991) A 3′ splice site consensus sequence mutation in the intron 3 of the alpha-galactosidase A gene in a patient with Fabry disease. *Jinrui Idengaku Zasshi*, **36**, 245–250.

113. Matsumura,T., Osaka,H., Sugiyama,N., Kawanishi,C., Maruyama,Y., Suzuki,K., Onishi,H., Yamada,Y., Morita,M., Aoki,M. *et al.* (1998) Novel acceptor splice site mutation in the invariant AG of intron 6 of alpha-galactosidase A gene, causing Fabry disease. Mutations in brief no. 146. Online. *Hum. Mutat.*, **11**, 483.

114. Huang,C.H., Blumenfeld,O.O., Reid,M.E., Chen,Y., Daniels,G.L. and Smart,E. (1997) Alternative splicing of a novel glycophorin allele GPHe(GL) generates two protein isoforms in the human erythrocyte membrane. *Blood*, **90**, 391–397.

115. Steingrimsdottir,H., Rowley,G., Dorado,G., Cole,J. and Lehmann,A.R. (1992) Mutations which alter splicing in the human hypoxanthine-guanine phosphoribosyltransferase gene. *Nucleic Acids Res.*, **20**, 1201–1208.

116. Lin,Y.W., Perkins,J.J., Zhang,Z. and Aplan,P.D. (2004) Distinct mechanisms lead to *HPRT* gene mutations in leukemic cells. *Genes Chrom. Cancer*, **39**, 311–323.

117. Gibbs,R.A., Nguyen,P.N., McBride,L.J., Koepf,S.M. and Caskey,C.T. (1989) Identification of mutations leading to the Lesch-Nyhan syndrome by automated direct DNA sequencing of *in vitro* amplified cDNA. *Proc. Natl Acad. Sci. USA*, **86**, 1919–1923.

118. Gibbs,R.A., Nguyen,P.N., Edwards,A., Civitello,A.B. and Caskey,C.T. (1990) Multiplex DNA deletion detection and exon sequencing of the hypoxanthine phosphoribosyltransferase gene in Lesch-Nyhan families. *Genomics*, **7**, 235–244.

119. Takahashi,Y., Kadowaki,H., Ando,A., Quin,J.D., MacCuish,A.C., Yazaki,Y., Akanuma,Y. and Kadowaki,T. (1998) Two aberrant splicings caused by mutations in the insulin receptor gene in cultured lymphocytes from a patient with Rabson-Mendenhall's syndrome. *J. Clin. Invest.*, **101**, 588–594.

120. Newman,P.J., Seligsohn,U., Lyman,S. and Coller,B.S. (1991) The molecular genetic basis of Glanzmann thrombasthenia in the Iraqi-Jewish and Arab populations in Israel. *Proc. Natl Acad. Sci. USA*, **88**, 3160–3164.

121. Jonkman,M.F., Heeres,K., Pas,H.H., van Luyn,M.J., Elema,J.D., Corden,L.D., Smith,F.J., McLean,W.H., Ramaekers,F.C., Burton,M. *et al.* (1996) Effects of keratin 14 ablation on the clinical and cellular phenotype in a kindred with recessive epidermolysis bullosa simplex. *J. Invest. Dermatol.*, **107**, 764–769.

122. Naom,I., D'Alessandro,M., Sewry,C.A., Philpot,J., Manzur,A.Y., Dubowitz,V. and Muntoni,F. (1998) Laminin alpha 2-chain gene mutations in two siblings presenting with limb-girdle muscular dystrophy. *Neuromuscl. Disord.*, **8**, 495–501.

123. Castiglia,D., Posteraro,P., Spirito,F., Pinola,M., Angelo,C., Puddu,P., Meneguzzi,G. and Zambruno,G. (2001) Novel mutations in the *LAMC2* gene in non-Herlitz junctional epidermolysis bullosa: effects on laminin-5 assembly, secretion, and deposition. *J. Invest. Dermatol.*, **117**, 731–739.

124. Maruyama,T., Miyake,Y., Yamamura,T., Tajima,S., Funahashi,T., Matsuzawa,Y. and Yamamoto,A. (1998) A novel point mutation in a splice acceptor site of intron 1 of the human low density lipoprotein receptor gene which causes severe hypercholesterolemia: an unexpected absence of exon skipping. Mutations in brief no. 139. Online. *Hum. Mutat.*, **11**, 480–481.

125. Yu,L., Heere-Ress,E., Boucher,B., Defesche,J.C., Kastelein,J., Lavoie,M.A. and Genest,J.Jr (1999) Familial hypercholesterolemia. Acceptor splice site (G→C) mutation in intron 7 of the *LDLR* gene: alternate RNA editing causes exon 8 skipping or a premature stop codon in exon 8. *LDLR* (Honduras-1) [LDL-R1061(-1) G→C]. *Atherosclerosis*, **146**, 125–131.

126. Rodningen,O.K., Tonstad,S., Saugstad,O.D., Ose,L. and Leren,T.P. (1999) Mutant transcripts of the LDL receptor gene: mRNA structure and quantity. *Hum. Mutat.*, **13**, 186–196.

127. Webb,J.C., Patel,D.D., Shoulders,C.C., Knight,B.L. and Soutar,A.K. (1996) Genetic variation at a splicing branch point in intron 9 of the low density lipoprotein (LDL)-receptor gene: a rare mutation that disrupts mRNA splicing in a patient with familial hypercholesterolaemia and a common polymorphism. *Hum. Mol. Genet.*, **5**, 1325–1331.

128. Machinis,K., Pantel,J., Netchine,I., Leger,J., Camand,O.J., Sobrier,M.L., Dastot-Le Moal,F., Duquesnoy,P., Abitbol,M., Czernichow,P. *et al.* (2001) Syndromic short stature in patients with a germline mutation in the LIM homeobox *LHX4*. *Am. J. Hum. Genet.*, **69**, 961–968.

129. Bonnevie-Nielsen,V., Leigh Field,L., Lu,S., Zheng,D.J., Li,M., Martensen,P.M., Nielsen,T.B., Beck-Nielsen,H., Lau,Y.L. and Pociot,F. (2005) Variation in antiviral 2′,5′-oligoadenylate synthetase (2′5′AS) enzyme activity is controlled by a single-nucleotide polymorphism at a splice-acceptor site in the *OAS1* Gene. *Am. J. Hum. Genet.*, **76**, 623–633.

130. Carstens,R.P., Fenton,W.A. and Rosenberg,L.R. (1991) Identification of RNA splicing errors resulting in human ornithine transcarbamylase deficiency. *Am. J. Hum. Genet.*, **48**, 1105–1114.

131. Piriev,N.I., Shih,J.M. and Farber,D.B. (1998) Defective RNA splicing resulting from a mutation in the cyclic guanosine monophosphate-phosphodiesterase beta-subunit gene. *Invest. Ophthalmol. Vis. Sci.*, **39**, 463–470.

132. Tsujino,S., Servidei,S., Tonin,P., Shanske,S., Azan,G. and DiMauro,S. (1994) Identification of three novel mutations in non-Ashkenazi Italian patients with muscle phosphofructokinase deficiency. *Am. J. Hum. Genet.*, **54**, 812–819.

133. Kanno,H., Fujii,H., Wei,D.C., Chan,L.C., Hirono,A., Tsukimoto,I. and Miwa,S. (1997) Frame shift mutation, exon skipping, and a two-codon deletion caused by splice site mutations account for pyruvate kinase deficiency. *Blood*, **89**, 4213–4218.

134. Oppliger,T., Thony,B., Kluge,C., Matasovic,A., Heizmann,C.W., Ponzone,A., Spada,M. and Blau,N. (1997) Identification of mutations causing 6-pyruvoyl-tetrahydropterin synthase deficiency in four Italian families. *Hum. Mutat.*, **10**, 25–35.

135. Strautnieks,S.S., Thompson,R.J., Gardiner,R.M. and Chung,E. (1996) A novel splice-site mutation in the gamma subunit of the epithelial sodium channel gene in three pseudohypoaldosteronism type 1 families. *Nature Genet.*, **13**, 248–250.

136. Morita,M., Ho,M., Hosler,B.A., McKenna-Yasek,D. and Brown,R.H.Jr (2002) A novel mutation in the spastin gene in a family with spastic paraplegia. *Neurosci. Lett.*, **325**, 57–61.

137. Chavanas,S., Bodemer,C., Rochat,A., Hamel-Teillac,D., Ali,M., Irvine,A.D., Bonafe,J.L., Wilkinson,J., Taieb,A., Barrandon,Y. *et al.* (2000) Mutations in *SPINK5*, encoding a serine protease inhibitor, cause Netherton syndrome. *Nature Genet.*, **25**, 141–142.

138. Villa,A., Notarangelo,L.D., Di Santo,J.P., Macchi,P.P., Strina,D., Frattini,A., Lucchini,F., Patrosso,C.M., Giliani,S., Mantuano,E. *et al.* (1994) Organization of the human CD40L gene: implications for molecular defects in X chromosome-linked hyper-IgM syndrome and prenatal diagnosis. *Proc. Natl Acad. Sci. USA*, **91**, 2110–2114.

139. Varley,J.M., Attwooll,C., White,G., McGown,G., Thorncroft,M., Kelsey,A.M., Greaves,M., Boyle,J. and Birch,J.M. (2001) Characterization of germline *TP53* splicing mutations and their genetic and functional analysis. *Oncogene*, **20**, 2647–2654.

140. Varley,J.M., McGown,G., Thorncroft,M., White,G.R., Tricker,K.J., Kelsey,A.M., Birch,J.M. and Evans,D.G. (1998) A novel *TP53* splicing mutation in a Li-Fraumeni syndrome family: a patient with Wilms' tumour is not a mutation carrier. *Br. J. Cancer*, **78**, 1081–1083.

141. Bendig,I., Mohr,N., Kramer,F. and Weber,B.H. (2004) Identification of novel *TP53* mutations in familial and sporadic cancer cases of German and Swiss origin. *Cancer Genet. Cytogenet.*, **154**, 22–26.

142. Takahashi,T., D'Amico,D., Chiba,I., Buchhagen,D.L. and Minna,J.D. (1990) Identification of intronic point mutations as an alternative mechanism for p53 inactivation in lung cancer. *J. Clin. Invest.*, **86**, 363–369.

143. Gantla,S., Bakker,C.T., Deocharan,B., Thummala,N.R., Zweiner,J., Sinaasappel,M., Roy Chowdhury,J., Bosma,P.J. and Roy Chowdhury,N. (1998) Splice-site mutations: a novel genetic mechanism of Crigler-Najjar syndrome type 1. *Am. J. Hum. Genet.*, **62**, 585–592.

144. Sappal,B.S., Ghosh,S.S., Shneider,B., Kadakol,A., Chowdhury,J.R. and Chowdhury,N.R. (2002) A novel intronic mutation results in the use of a cryptic splice acceptor site within the coding region of *UGT1A1*, causing Crigler-Najjar syndrome type 1. *Mol. Genet. Metab.*, **75**, 134–142.

145. Satokata,I., Tanaka,K., Miura,N., Miyamoto,I., Satoh,Y., Kondo,S. and Okada,Y. (1990) Characterization of a splicing mutation in group A xeroderma pigmentosum. *Proc. Natl Acad. Sci. USA*, **87**, 9908–9912.

146. Shah,A.B., Chernov,I., Zhang,H.T., Ross,B.M., Das,K., Lutsenko,S., Parano,E., Pavone,L., Evgrafov,O., Ivanova-Smolenskaya,I.A. *et al.* (1997) Identification and analysis of mutations in the Wilson disease gene (*ATP7B*): population frequencies, genotype-phenotype correlation, and functional analyses. *Am. J. Hum. Genet.*, **61**, 317–328.

147. Li,S.S., Tseng,H.M., Yang,T.P., Liu,C.H., Teng,S.J., Huang,H.W., Chen,L.M., Kao,H.W., Chen,J.H., Tseng,J.N. *et al.* (1999) Molecular characterization of germline mutations in the *BRCA1* and *BRCA2* genes from breast cancer families in Taiwan. *Hum. Genet.*, **104**, 201–204.

148. Hoffman,J.D., Hallam,S.E., Venne,V.L., Lyon,E. and Ward,K. (1998) Implications of a novel cryptic splice site in the *BRCA1* gene. *Am. J. Med. Genet.*, **80**, 140–144.

149. Beck,S., Penque,D., Garcia,S., Gomes,A., Farinha,C., Mata,L., Gulbenkian,S., Gil-Ferreira,K., Duarte,A., Pacheco,P. *et al.* (1999) Cystic fibrosis patients with the 3272-26A>G mutation have mild disease, leaky alternative mRNA splicing, and *CFTR* protein at the cell membrane. *Hum. Mutat.*, **14**, 133–144.

150. Fanen,P., Ghanem,N., Vidaud,M., Besmond,C., Martin,J., Costes,B., Plassa,F. and Goossens,M. (1992) Molecular characterization of cystic fibrosis: 16 novel mutations identified by analysis of the whole cystic fibrosis conductance transmembrane regulator (*CFTR*) coding regions and splice site junctions. *Genomics*, **13**, 770–776.

151. Boot,R.G., Renkema,G.H., Verhoek,M., Strijland,A., Bliek,J., de Meulemeester,T.M., Mannens,M.M. and Aerts,J.M. (1998) The human chitotriosidase gene. Nature of inherited enzyme deficiency. *J. Biol. Chem.*, **273**, 25680–25685.

152. Rosipal,R., Lamoril,J., Puy,H., Da Silva,V., Gouya,L., De Rooij,F.W., Te Velde,K., Nordmann,Y., Martasek,P. and Deybach,J.C. (1999) Systematic analysis of coproporphyrinogen oxidase gene defects in hereditary coproporphyria and mutation update. *Hum. Mutat.*, **13**, 44–53.

153. Stasia,M.J., Bordigoni,P., Martel,C. and Morel,F. (2002) A novel and unusual case of chronic granulomatous disease in a child with a homozygous 36-bp deletion in the *CYBA* gene (A22(0)) leading to the activation of a cryptic splice site in intron 4. *Hum. Genet.*, **110**, 444–450.

154. Higashi,Y., Tanae,A., Inoue,H., Hiromasa,T. and Fujii-Kuriyama,Y. (1988) Aberrant splicing and missense mutations cause steroid 21-hydroxylase [P-450(C21)] deficiency in humans: possible gene conversion products. *Proc. Natl Acad. Sci. USA*, **85**, 7486–7490.

155. Fisher,C.W., Lau,K.S., Fisher,C.R., Wynn,R.M., Cox,R.P. and Chuang,D.T. (1991) A 17-bp insertion and a Phe215Cys missense mutation in the dihydrolipoyl transacylase (E2) mRNA from a thiamine-responsive maple syrup urine disease patient WG-34. *Biochem. Biophys. Res. Commun.*, **174**, 804–809.

156. Roest,P.A., Bout,M., van der Tuijn,A.C., Ginjaar,I.B., Bakker,E., Hogervorst,F.B., van Ommen,G.J. and den Dunnen,J.T. (1996) Splicing mutations in *DMD/BMD* detected by RT–PCR/PTT: detection of a 19AA insertion in the cysteine rich domain of dystrophin compatible with BMD. *J. Med. Genet.*, **33**, 935–939.

157. Weeda,G., van Ham,R.C., Vermeulen,W., Bootsma,D., van der Eb,A.J. and Hoeijmakers,J.H. (1990) A presumed DNA helicase encoded by *ERCC-3* is involved in the human repair disorders xeroderma pigmentosum and Cockayne's syndrome. *Cell*, **62**, 777–791.

158. Morgan,N.V., Tipping,A.J., Joenje,H. and Mathew,C.G. (1999) High frequency of large intragenic deletions in the Fanconi anemia group A gene. *Am. J. Hum. Genet.*, **65**, 1330–1341.

159. Weber,Y., Steinberger,D., Deuschl,G., Benecke,R. and Muller,U. (1997) Two previously unrecognized splicing mutations of *GCH1* in Dopa-responsive dystonia: exon skipping and one base insertion. *Neurogenetics*, **1**, 125–127.

160. Metherall,J.E., Collins,F.S., Pan,J., Weissman,S.M. and Forget,B.G. (1986) β⁰-thalassemia caused by a base substitution that creates an alternative splice acceptor site in an intron. *EMBO J.*, **5**, 2551–2557.

161. Busslinger,M., Moschonas,N. and Flavell,R.A. (1981) β⁺ thalassemia: aberrant splicing results from a single point mutation in an intron. *Cell*, **27**, 289–298.

162. Fukumaki,Y., Ghosh,P.K., Benz,E.J.Jr, Reddy,V.B., Lebowitz,P., Forget,B.G. and Weissman,S.M. (1982) Abnormally spliced messenger

RNA in erythroid cells from patients with β⁺ thalassemia and monkey cells expressing a cloned β⁺-thalassemic gene. *Cell*, **28**, 585–593.

163. Spritz,R.A., Jagadeeswaran,P., Choudary,P.V., Biro,P.A., Elder,J.T., deRiel,J.K., Manley,J.L., Gefter,M.L., Forget,B.G. and Weissman,S.M. (1981) Base substitution in an intervening sequence of a β⁺-thalassemic human globin gene. *Proc. Natl Acad. Sci. USA*, **78**, 2455–2459.

164. Dlott,B., d'Azzo,A., Quon,D.V. and Neufeld,E.F. (1990) Two mutations produce intron insertion in mRNA and elongated β-subunit of human β-hexosaminidase. *J. Biol. Chem.*, **265**, 17921–17927.

165. O'Neill,J.P., Rogan,P.K., Cariello,N. and Nicklas,J.A. (1998) Mutations that alter RNA splicing of the human HPRT gene: a review of the spectrum. *Mutat. Res.*, **411**, 179–214.

166. Rossi,A.M., Tates,A.D., van Zeeland,A.A. and Vrieling,H. (1992) Molecular analysis of mutations affecting hprt mRNA splicing in human T-lymphocytes *in vivo*. *Environ. Mol. Mutagen.*, **19**, 7–13.

167. Nelson,C., Rabb,H. and Arnaout,M.A. (1992) Genetic cause of leukocyte adhesion molecule deficiency. Abnormal splicing and a missense mutation in a conserved region of CD18 impair cell surface expression of β2 integrins. *J. Biol. Chem.*, **267**, 3351–3357.

168. Rosenthal,A., Jouet,M. and Kenwrick,S. (1992) Aberrant splicing of neural cell adhesion molecule *L1* mRNA in a family with X-linked hydrocephalus. *Nature Genet.*, **2**, 107–112.

169. Jouet,M., Moncla,A., Paterson,J., McKeown,C., Fryer,A., Carpenter,N., Holmberg,E., Wadelius,C. and Kenwrick,S. (1995) New domains of neural cell-adhesion molecule *L1* implicated in X-linked hydrocephalus and MASA syndrome. *Am. J. Hum. Genet.*, **56**, 1304–1314.

170. Laccone,F., Huppke,P., Hanefeld,F. and Meins,M. (2001) Mutation spectrum in patients with Rett syndrome in the German population: Evidence of hot spot regions. *Hum. Mutat.*, **17**, 183–190.

171. FitzPatrick,D.R., Hill,A., Tolmie,J.L., Thorburn,D.R. and Christodoulou,J. (1999) The molecular basis of malonyl-CoA decarboxylase deficiency. *Am. J. Hum. Genet.*, **65**, 318–326.

172. Tanner,S.M., Laporte,J., Guiraud-Chaumeil,C. and Liechti-Gallati,S. (1998) Confirmation of prenatal diagnosis results of X-linked recessive myotubular myopathy by mutational screening, and description of three new mutations in the *MTM1* gene. *Hum. Mutat.*, **11**, 62–68.

173. Jaaskelainen,P., Kuusisto,J., Miettinen,R., Karkkainen,P., Karkkainen,S., Heikkinen,S., Peltola,P., Pihlajamaki,J., Vauhkonen,I. and Laakso,M. (2002) Mutations in the cardiac myosin-binding protein C gene are the predominant cause of familial hypertrophic cardiomyopathy in eastern Finland. *J. Mol. Med.*, **80**, 412–422.

174. Spritz,R.A., Fukai,K., Holmes,S.A. and Luande,J. (1995) Frequent intragenic deletion of the P gene in Tanzanian patients with type II oculocutaneous albinism (*OCA2*). *Am. J. Hum. Genet.*, **56**, 1320–1323.

175. Dworniczak,B., Aulehla-Scholz,C., Kalaydjieva,L., Bartholome,K., Grudda,K. and Horst,J. (1991) Aberrant splicing of phenylalanine hydroxylase mRNA: the major cause for phenylketonuria in parts of southern Europe. *Genomics*, **11**, 242–246.

176. Alonso,J., Frayle,H., Menendez,I., Lopez,A., Garcia-Miguel,P., Abelairas,J., Sarret,E., Vendrell,M.T., Navajas,A., Artigas,M. *et al.* (2004) Identification of 26 new constitutional *RB1* gene mutations in Spanish, Colombian, and Cuban retinoblastoma patients. *Hum. Mutat.*, **25**, 99.

177. Sippel,K.C., DeStefano,J.D., Berson,E.L. and Dryja,T.P. (1998) Evaluation of the human arrestin gene in patients with retinitis pigmentosa and stationary night blindness. *Invest. Ophthalmol. Vis. Sci.*, **39**, 665–670.

178. Blanck,C., Kohlhase,J., Engels,S., Burfeind,P., Engel,W., Bottani,A., Patel,M.S., Kroes,H.Y. and Cobben,J.M. (2000) Three novel *SALL1* mutations extend the mutational spectrum in Townes-Brocks syndrome. *J. Med. Genet.*, **37**, 303–307.

179. Jochmans,K., Lissens,W., Yin,T., Michiels,J.J., van der Luit,L., Peerlinck,K., De Waele,M. and Liebaers,I. (1994) Molecular basis for type 1 antithrombin deficiency: identification of two novel point mutations and evidence for a *de novo* splice site mutation. *Blood*, **84**, 3742–3748.

180. Sapp,P.C., Rosen,D.R., Hosler,B.A., Esteban,J., McKenna-Yasek,D., O'Regan,J.P., Horvitz,H.R. and Brown,R.H.Jr (1995) Identification of three novel mutations in the gene for Cu/Zn superoxide dismutase in patients with familial amyotrophic lateral sclerosis. *Neuromuscul. Disord.*, **5**, 353–357.

181. Bulman,M.P., Harries,L.W., Hansen,T., Shepherd,M., Kelly,W.F., Hattersley,A.T. and Ellard,S. (2002) Abnormal splicing of hepatocyte nuclear factor 1 alpha in maturity-onset diabetes of the young. *Diabetologia*, **45**, 1463–1467.

182. Matsumoto,T., Imamura,O., Yamabe,Y., Kuromitsu,J., Tokutake,Y., Shimamoto,A., Suzuki,N., Satoh,M., Kitao,S., Ichikawa,K. *et al.* (1997) Mutation and haplotype analyses of the Werner's syndrome gene based on its genomic structure: genetic epidemiology in the Japanese population. *Hum. Genet.*, **100**, 123–130.

183. Chan,A.C., Kadlecek,T.A., Elder,M.E., Filipovich,A.H., Kuo,W.L., Iwashima,M., Parslow,T.G. and Weiss,A. (1994) ZAP-70 deficiency in an autosomal recessive form of severe combined immunodeficiency. *Science*, **264**, 1599–1601.