ARTICLE

Noncanonical and canonical splice sites: a novel mutation at the rare noncanonical splice-donor cut site (IVS4+1A>G) of *SEDL* causes variable splicing isoforms in X-linked spondyloepiphyseal dysplasia tarda

Feng Xiong[1,8], Jianjun Gao*[,2,3,4,8], Jun Li[2], Yun Liu[4], Guoyin Feng[2], Wenli Fang[3], Hongfen Chang[3], Jiang Xie[1], Haitao Zheng[5], Tingyu Li*[,1] and Lin He*[,2,6,7]

[1]*Children's Hospital of Chongqing Medical University, Central District, Chongqing, PR China;* [2]*Bio-X Center, Shanghai Jiao Tong University, Shanghai, China;* [3]*Changning Greenland Hospital, Institute for Neuropsychiatric Science and Metabonomics, Shanghai Jiao Tong University, Shanghai, China;* [4]*Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences; Graduate School of Chinese Academy Sciences, Shanghai, China;* [5]*Department of Abdominal Surgery, The Affiliated Yantai Yuhuangding Hospital of Qingdao University Medical College, Yantai, Shangdong Province, China;* [6]*The Obstetrics and Gynecology Hospital, Fudan University, Shanghai, China;* [7]*Institutes of Biomedical Sciences, Fudan University, Shanghai, China*

X-linked spondyloepiphyseal dysplasia tarda can be caused by mutations in the *SEDL* gene. This study describes an interesting novel mutation (IVS4+1A>G) located exactly at the rare noncanonical AT–AC consensus splicing donor point of *SEDL*, which regained the canonical GT–AG consensus splicing junction in addition to several other rarer noncanonical splice patterns. The mutation activated several cryptic splice sites and generated the production of seven erroneous splicing isoforms, which we confirmed by sequencing of RT-PCR products and resequencing of cDNA clones. All the practical splice donors/acceptors were further assessed using FSPLICE 1.0 and SPL(M) Platforms to predict potential splice sites in genomic DNA. Subsequently, the expression levels of *SEDL* among the affected patients, carriers and controls were estimated using real-time quantitative PCR. Expression analyses showed that the expression levels of *SEDL* in both patients and carriers were decreased. Taken together, these results illustrated how disruption of the AT donor site in a rare AT–AC intron, leading to a canonical GT donor site, resulted in a multitude of aberrant transcripts, thus impairing exon definition. The unexpected splicing patterns resulting from the special mutation provide additional challenges and opportunities for understanding splicing mechanisms and specificity.

*Correspondence: Dr J Gao, Bio-X Center, Shanghai Jiao Tong University, PO Box 509, Hao Ran Building, 1954 Hua Shan Road, Shanghai 200030, China.
Tel: +86 21 62932779; Fax: +86 21 62932779; E-mail: gaoj3@niehs.nih.gov
or Dr T Li, Children's Hospital of Chongqing Medical University, 136 2nd Zhongshan Road, Central District, Chongqing 400014, China.
Tel: +86 23 63623604; Fax: +86 23 63623604; E-mail: tyli@vip.sina.com
or Dr L He, The Obstetrics and Gynecology Hospital, Fudan University, Shanghai 200011, China. Tel: +86 21 62822491; Fax: +86 21 62822491;
E-mail: helinhelin@gmail.com
[8]These authors contributed equally to this work.

Introduction

X-linked spondyloepiphyseal dysplasia tarda (SEDT, MIM313400) can be caused by mutations in the *SEDL* gene (MIM300202). *SEDL* is a highly conserved gene with orthologs identified in yeast, flies and vertebrates, and it is widely expressed in tissues, including fibroblasts, lymphoblasts and fetal cartilage.[1] It contains six exons[2] and spans a genomic region of approximately 23 kb on Xp22 in the human genome; it also escapes from X-inactivation.[3] As a small gene with alternative splicing involving exon 2,[4] ubiquitously expressed *SEDL* mRNA is about 2.8 kb in size. Exons 1 and 2 are noncoding, and the remaining 420 bp coding region is split among exons 3–6. The corresponding introns vary in size from 339 to 14 061 bp and their splice-site sequences for four of the five introns conform to the general canonical **GT–AG** splicing rule[5] using the U1–U2 snRNPs,[6] whereas the splicing sites with **AT–AC** ends between exons 4 and 5 implicate a rare noncanonical splicing mechanism[3] often utilizing the minor U11–U12 snRNPs.[6] In fact, there are different views regarding the precise splice sites of exon 4/ intron 4/ exon 5. The region was originally reported as an **AT–AT** splice site.[1,2] However, after a splice database search on the basis of the high-throughput genome sequencing project,[7] Mumm and colleagues argued that rather than an **AT–AT** site as proposed previously, the splice site is more likely to be a rare noncanonical **AT–AC** sequence.[3] As the 'catat' sequences in the boundaries of both exon 4/intron 4 and intron 4/exon 5 are the same, it had not been possible to identify the exact splicing site by simply comparing the genomic DNA with cDNA. Here, we report that an interesting novel mutation, exactly at the boundary of exon 4/intron 4, suggests a factual exon 4/5 definition.

In this study, involving members of a large Chinese family with X-linked spondyloepiphyseal dysplasia tarda, we provide substantial evidence that the mutation at the splicing donor site of intron 4 results in aberrant splicing of *SEDL* in the affected males and carrier females. This is a fairly unusual mutation occurring exactly at the **AT** splicing donor end, leading to a **GT** donor site and resulting in the activation of certain cryptic splice sites. A series of erroneous splices were generated due to the change from a rare noncanonical **AT–AC** splice joint to an ordinary canonical **GT–AG** pattern or an even rarer noncanonical **GT–AT** splice site.

Materials and methods
Patients and clinical diagnosis
After giving informed written consent, a total of 24 individuals descended directly from one antecedent were examined clinically and radiographically. The pedigree was consistent with an X-linked recessive pattern of inheritance for SEDT (see family pedigree, Supplementary

Figure 1). All the adult patients had suffered back and hip pain beginning in their 20s, and the pain had gradually deteriorated. Radiographs of the lumbar spine showed platyspondylia, a posterior hump, and narrowing of the disc space accompanied by end plate sclerosis. They were of short stature in the range of 115.5–133.5 cm, with upper body measurements in the 55.5–65 cm range. Arm spans exceeded height measurements and were in the 132–159 cm range.

PCR and DNA sequencing
DNA was obtained from the 24 family members, comprising 13 affected males, 5 obligate carrier females and 6 unaffected family members. Given the small number of *SEDL* gene exons and the relative ease of their PCR amplification, we made direct sequencing the primary choice for the detection of the *SEDL* mutation.[8] DNA extraction was performed using the salting-out method.[9] Exons 3, 4, 5 and 6 (containing the coding sequence) of *SEDL* and adjacent splice sites were amplified by PCR and sequenced in both directions, as described by Gedeon et al.[1] Sequencing reactions were performed with Big-Dye terminators and an ABI 3730 automated sequencer. All of the sequences were compared against the normal sequences of unrelated controls and the UCSC and NCBI databases.

cDNA cloning, resequencing and database searches
To further assess the consequence of the defect on *SEDL* gene transcription and RNA splicing, RT-PCR experiments were performed. Total RNA from blood lymphocytes was extracted using Trizol reagent (Invitrogen). RNA integrity was confirmed by direct visualization of 18S and 28S rRNA bands after agarose gel electrophoresis. RNA samples were incubated using 10 U of DNase I (Novagen) at 37°C for 20 min to remove residual DNA, followed by inactivation at 65°C for 10 min. RNA samples were further purified using a HiBind spin column (Omega) according to the manufacturer's instructions.[10] The purified RNA samples (0.5 μg) were then reverse-transcribed using the SuperScript first-strand synthesis system (Invitrogen) and oligo-dT18. Subsequently, PCR was performed using primers E1F2 (5′-CTTCCGCGGAAACTGACATTGC-3′) and 5Rx (5′-GTA TACACCATTGTGGTGACATC-3′), as described by Gedeon et al.[11] RT-PCR products were sequenced in both directions. Simultaneously, to more precisely characterize the effect of the mutation, the unpurified PCR products were directly cloned into the pGEM-T vector (Promega) and resequenced in both directions using SP6 and T7primers.

A database of canonical and noncanonical mammalian splice sites (SpliceDB) was used to search for cDNA and genomic DNA of *SEDL*-related splice variants.[7] The FSPLICE 1.0 and SPL platforms (http://www.softberry.com/) were used to predict potential splice sites in genomic DNA. SPLM was implemented to substitute for SPL only in case

of noncanonical splice sites. Blast (NCBI) was used to search databases for *SEDL*-related sequences.

Real-time quantitative PCR and expression estimates

To determine the expression levels of SEDL among the affected subjects, carriers and normal controls, extra primers spanning exons 3–4 (E3F (5′-CCAGCTGGGAAGG CAGAAT-3′) and E4R (5′-TCGAGAGCAGCATGAGCTATG-3′), amplicon of 71 bp) were designed. Real-time quantitative PCR was performed using the ABI 7900 system (Applied Biosystems). Reactions were performed in a 10 μl volume including diluted cDNA samples, primers and SYBR Green I Mastermix (Applied Biosystems). Diluted cDNA samples produced from 10 ng total RNA were added to each well. Real-time PCR data were collected using SDS software (version 2.1, Applied Biosystems). Expression of SEDL E3-4 was normalized in relation to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) levels (denoted as ΔC_t).[10] Primer pairs for GAPDH were GAPDHF: 5′-ATCACCATCTTCCAG GAGCGAG-3′ and GAPDHR: 5′-GTTGTCATGGAT GACCTTGGCC-3′. Both GAPDH and SEDL were tested three times in a reaction for each sample.

We were aware that seven *SEDL* pseudogenes have been detected in the human genome. Of these, *SEDLP1*, a transcribed retropseudogene (or retro-xaptonuon), is located on chromosome 19q13.4 and shows only six nucleotide differences in the ORF as compared with that of the X-linked *SEDL* gene. To avoid the interference of *SEDLP*, both primer combinations for RT-PCR and for real-time quantitative PCR were designed to amplify only the

X-linked *SEDL* gene and not the *SEDLP1* (chromosome 19) pseudogene sequence or any other pseudogenes.

Results

Genomic DNA sequencing revealed that a single nucleotide transition of A + 3431G (position given relative to the ATG start code) occurred at the exon 4/intron 4 boundary (Figure 1a). All affected male subjects showed a single nucleotide substitution of 'G' for 'A,' and all obligate carriers showed a heterozygous 'A/G' substitution, in contrast to a homozygous 'A/A' at the corresponding locus in the unaffected male family members (Figure 1b), unrelated controls and public databases.

To identify whether the mutation caused either missense coding or erroneous splicing, it was necessary to explore the precise boundaries of exon 4/intron 4/exon 5 using a direct biological approach, as opposed to previously used bioinformatic methods.[3] We sequenced RT-PCR products in both directions. Sequence chromatograms showed distinct peaks ahead of the mutation point and ambiguous peaks thereafter in both patients and obligate carriers (Figure 1c). No changes were present in control chromosomes (Figure 1d), which suggested deletion or insertion events around the mutation site. After unscrambling the sequences with ambiguous peaks, we found the mutation was an intronic defect rather than an exonic missense codon, which disrupted the splicing processing. These splicing errors at the donor site (IVS4 + 1A > G or A + 3431G) resulted in a dramatic increase in the percentage
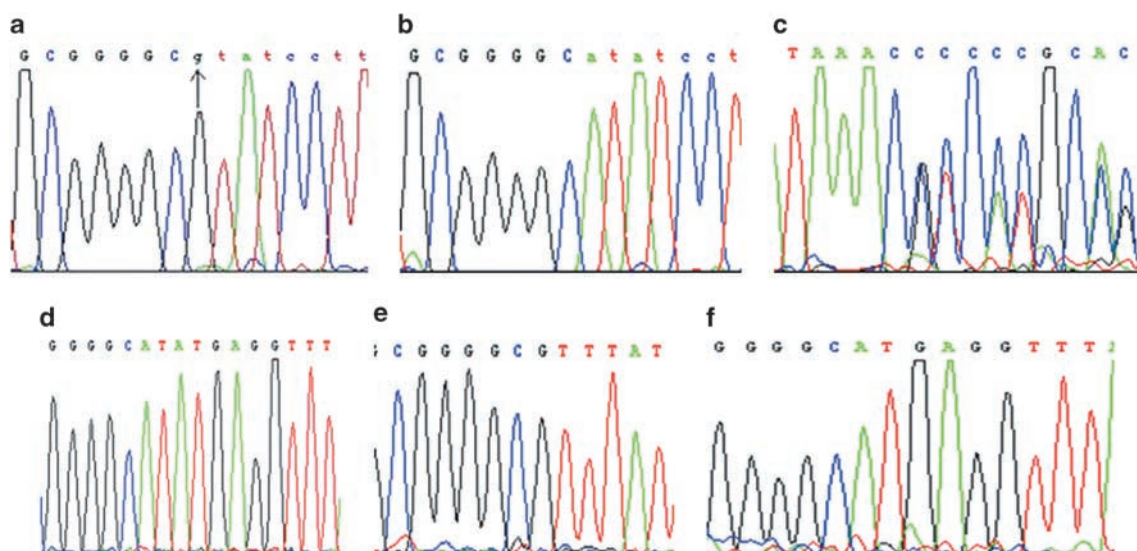

Figure 1 Sequence chromatograms showing mutation in *SEDL* of affected males and its consequences on splicing junctions. (**a**) gDNA sequence denotes the IVS4 + 1A > G mutation. The arrow indicates where the mutation occurred. Exonic sequence is shown in capital letters, intronic sequence in lower case; (**b**) gDNA sequence of normal control (**c**) cDNA sequence of RT-PCR product in the reverse-sequencing reaction of patients; (**d**) cDNA sequence of normal control (GGGGC−ATATGAGGTTT splicing junction); (**e**) del **ATATGAG** splice pattern in either patients or carriers; (**f**) del **AT** splice pattern in either patients or carriers.

of consensus strength from 29.0 to 49.1, as calculated by Shapiro and Senapathy.[12] Factually, alteration in RNA splice processing was generated with extra deletions of a seven-nucleotide 'ATATGAG' or 'AT' dinucleotide (Marked as 'del **ATATGAG**' and 'del **AT**,' respectively, Table 1, Figure 1e and f).

The results of sequencing cDNA clones confirmed that the two abnormal splice patterns described above presented in clones from patients and carriers' cDNA (Table 1 and Figure 1), whereas 33 clones from three normal controls showed only the ordinary sequence. The normal splicing junction, with the sequence 'atatcctt…ttatttcac' spliced out, had a noncanonical **AT–AC** splice pair rather than the **AT–AT** pair. However, the splicing errors caused by the mutation showed the sequence '…GCGGGGC' of exon 4 directly joined to 'GTTTATT…' of exon 5, without the 5'-flanking seven 'ATATGAG' nucleotides of exon 5 being spliced into the transcript (Figure 2a). It appeared to resume the canonical '**GT–AG**' splice rule, with the sequence '**gt**atcctt…ttatttcacATATG**AG**' being spliced out. Here also, another splice pattern seemed to implicate a noncanonical **GT–AT** splice pair, with '**gt**atcctt…ttattt cac**AT**' being spliced out. Both of the splice junctions caused by the single nucleotide mutation predicted a frameshift in the ORF. Thus, premature termination resulted in the truncation of the putative protein.

In addition, we also found that a series of variable splicing isoforms occurred infrequently in the patients' and carriers' cDNA clones. A different RNA splice junction with an extra deletion of a four-nucleotide '**ATAT**' was identified in the transcripts, which indicated the presence of another noncanonical **GT–AT** splice pair with '**gt**atcct t…ttatttcacAT**AT**' being spliced out (marked as del **ATAT**; Table 1, Figure 2a and Supplementary Figure 2a). Consequently, the frameshift resulted in a premature stop codon as well as a truncated SEDL protein.

Furthermore, we discovered several other variable splicing isoforms related to the removal of large portions of coding sequence. Generally, the splicing implied that the mutation created stronger cryptic splice pairs that may compete with the constitutively used splice site pairs, thus impairing exon definition, as described by Dietz.[13] Occasionally, however, some factual splice sites were not so 'strong,' which may be caused by the mutation created/ disrupted exonic splicing enhancer/silencer or intronic splicing enhancer/silencer sites.[6]

One isoform was generated by the splicing pattern that is schematically described as '**E1 + E3 + E4 + partial intron 4 + E5 + E6**' (Table 1, Figure 2b and Supplementary Figure 2b). It was quite likely that the mutation prevented the primary RNA splicing process, and that an excessive partial intron 4 sequence from IVS4 + 1 to + 23 (or position + 3, 431− + 3, 453) was spliced into the transcript, producing a new noncanonical '**GT–AT**' splice junction, with '**gt**aactaac…tttcac**AT**' spliced out. The ends

Table 1 Splicing patterns, frequencies and weight matrices of different transcript isoforms in patients and carriers

Splicing pattern	Schematic sequence	Patients	Carriers	SPL(M)/FSPLICE		
				Donor	Acceptor	
Wild type: GCGGGGC|+3431 atatccttacc… atttc**ac**+3874|ATATGAGGTTTATT		2/18	27/41	(10)/N.A.	N.A.	
del ATATGAG: GCGGGGC|+3431 **gt**atccttacc… atttcac**ATATGAG**+3881| GTTTATT		4/18	4/41	N. A.	N.A.	
del AT: GCGGGGC|+3431 **gt**atccttacc… atttcac**AT**+3876|ATGAGGTTT		7/18	1/41	N. A.	N.A.	
del ATAT: GCGGGGC|+3431**gt**atccttacc… atttcac**ATAT**+3878| GAGGAGGTTT		1/18	0/41	N. A.	N.A.	
E1+E3+E 4+partial intron 4+E5+E6: E1−+E3+E4… GCGGGGC**gt**atccttaccttcttagtaaag|+3454 **gt**aactaac…atttcac**AT** +3876 ATGAGGTTTATT… E5+E6		0/18	3/41	0.78/11.98	N.A.	
E1+E4+broken intron4+E5+E6: E1 +|−14561 **gt**aggga…tcatctag +3285|+E4+ gtatcct…ggtgaaaat|+3546 **gt**aagtttcc…catttctaag +3849| aaagattatgccttaattttcac**ATATGAG**GTTT +E5+E6		1/18	0/41	−14561: 0.83/11.00	+3285: 0.69/5.95 + 3849: 0.67/5.50	
E1+partial intron 2+E3, 4 +partial intron4 +E5, 6: E1+|−14561 **gt**agggaatg…tccttctag −535| **gt**aagtgac…ggcaag |−464 **gt**aattact…acttaattag −20|+exon 3,4 + gtatccttaccttagtaaag +3454 **gt**aactaac…atttcac**AT**+3876|,+E5, E6		1/18	0/41	+3546: 0.85/13.10 −535: 0.66/8.78	N.A.:	
E1+partial intron 2+E3, 4, 5, 6: E1…ACTGCAG|−14561 **gt**agggaatg…tccttctag −535| gtaagtgac…ggcaag |−464 **gt**aattact…acttaattag −20| E3, 4, 5, 6		1/18	0/41	−464: 0.83/12.26	−20: 0.67/5.50	

The exonic sequence in wild type is shown in capital letters, with the intronic sequence in lower case. Acceptor and donor ends are shown in bold and the corresponding sequence being spliced out is boxed.
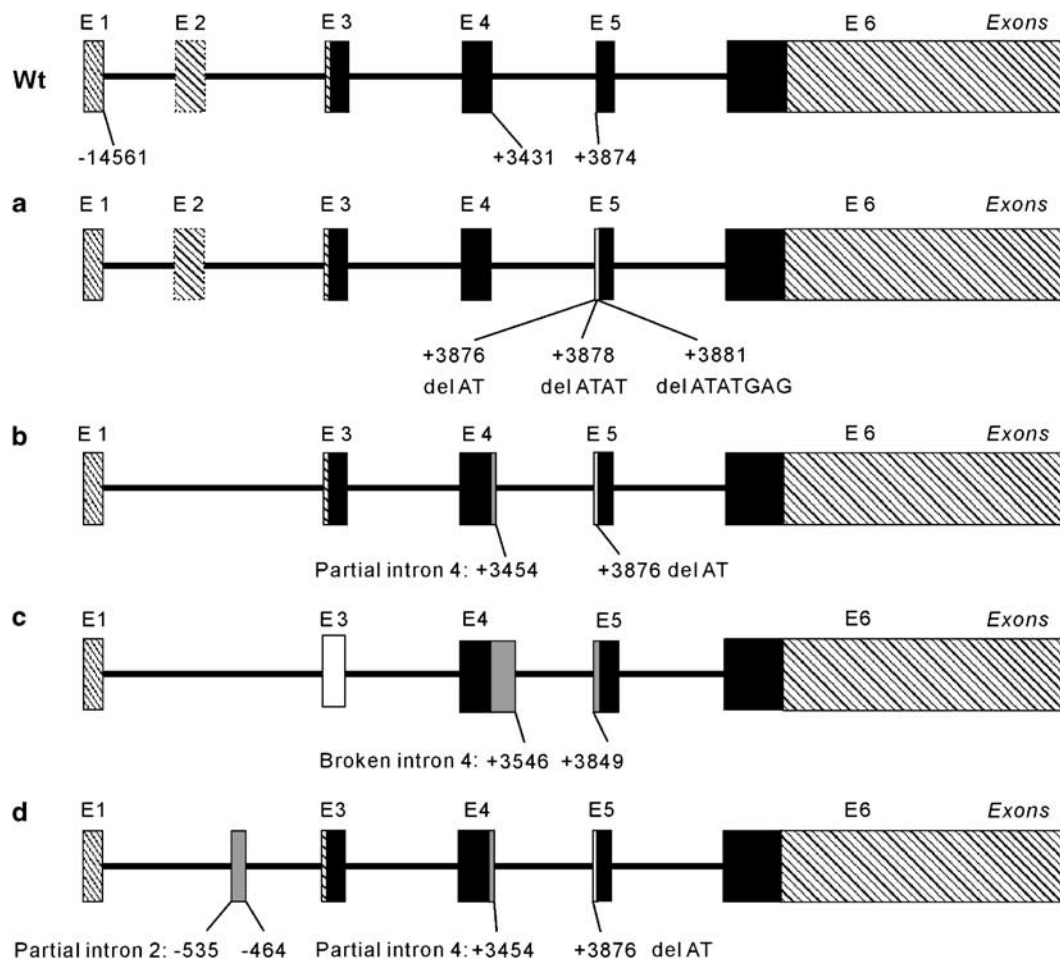
Figure 2 Schematic figures showing alternative splicing events caused by the IVS4 + 1A > G mutation. Wt: denotes the wild-type *SEDL* gene with alternative splicing Exon 2. The coding region is shaded black. Hatched regions represent the 5′- and 3′-untranslated regions (UTRs). White squares stand for the erroneous deleted parts of coding regions. Gray squares indicate the excessive sequences in the mature transcript. (**a**) Erroneous splice patterns of 'del ATATGAG', 'del AT' and 'del ATAT'; (**b**) Erroneous splice pattern of 'E1 + E3 + E4 + partial intron 4 + E5 + E6'; (**c**) 'E1 + E4 + broken Intron4 + E5 + E6'; (**d**) 'E1 + partial intron 2 + E3, 4 + partial intron4 + E5, 6.'

with '**AT**' indicated the first dinucleotide of exon 5. The altered exon 4 was referred to as exon 4a.[2,14] BLAST results predicted that the splicing pattern would cause eight amino-acid residue insertions into the 82nd to 83rd amino acids (reference: EAW98833) of sedlin.

Another complicated splice pattern was denoted as '**E1 + E4 + broken intron4 + E5 + E6**' (Figure 2c). 'E2 + intron2 + E3' was spliced out of the transcript, and an excessive sequence from intron 4 was detected in the transcript. It is more likely that two cryptic splice sites in accordance with the canonical '**GT–AG**' splicing mechanism were generated in intron 4. Both located at the downstream of the normal donor of exon 4, and upstream of the normal acceptor of exon 5 (Supplementary Figure 3).

A much more complicated splicing process was schematically presented as '**E1 + partial intron 2 + E3, 4 + partial intron4 + E5, 6**' (Figure 2d). The transcript was probably created by the activation of cryptic splice sites within both intron 2 and intron 4. A redundant fragment with 70 nucleotides (position −535 to −464, '**gt**aagtgac…ggca**ag**') from intron 2 was spliced into the transcript, which implied that two potential canonical '**GT–AG**' splice pairs were activated. Therefore, intron 2 broken into two fragments with 13 546 nucleotides (position −14561 to −535, **gt**agggaatg…tcctttct**ag**) and 445 nucleotides (position −464 to approximately −20 **gt**aattact…acttaatt**ag**) were spliced out (Supplementary Figure 4). The sequence of intron 4 was the same as the splicing pattern of '**E1 + E3 + E 4 + partial intron 4 + E5, 6**,' with the extra residues of 'gtatcct taccttcttagtaaag.'

Finally, we also found an erroneous transcript involving the partial intron 2 with 70 nucleotides as described in '**E1 + partial intron 2 + E3, 4 + partial intron4 + E5, 6**' but without partial intron4, denoted schematically as '**E1 + partial intron 2 + E3, 4, 5, 6**' (Table 1).
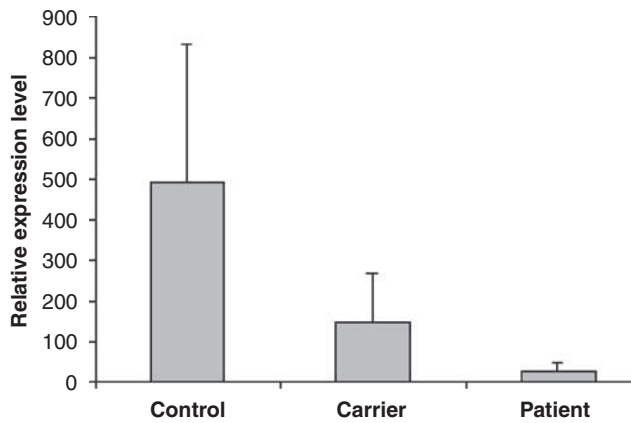
Figure 3 Relative expression levels of *SEDL* genes in controls, carriers and patients.

The predicted scores of SPL(M) and matrix weights of FSPLICE are also shown in Table 1. Our study suggests that some of the practical splice sites are not detectable by SPL, FSPLICE or even SPLM (marked as NA). It is most likely that those programs primarily predict **GT–AG** splice sites and are of limited value in detecting noncanonical splice sites.

Relative to the expression of the *GAPDH* housekeeping gene, expression values for each group were 25.30, 147.59 and 491.84, respectively (Figure 3). Analyses showed that expression levels in patients and carriers were both decreased. Patients' expression was approximately 1/20 of that of controls, whereas that of carriers was approximately from 1/6 to 1/3 of that of controls.

Discussion

SEDL gene mutations are spread along the entire length of the four *SEDL* exons[3–6] and their flanking introns. Examples include point mutations, splice alterations, an insertion, many deletions and several complex deletions/insertions. All characterized mutations within the *SEDL* gene led to loss of function of the SEDL protein.[8]

In our study, the novel IVS4 + 1A>G mutation caused seven erroneous transcript isoforms in total. Some patterns occurred more frequently than others and, in particular, the frequencies for the more complex splice patterns were very low. The possible reason was that the complex splice patterns implicated more erroneous splicing, whereas the simple ones simply involved only one erroneous splicing. The mutation was of particular interest, because it not only occurred exactly at the noncanonical splicing point, but also regained the canonical splice end with **GT**. In mammalian genomes, 99.24% of splice site pairs conform to the **GT–AG** rule, 0.69% to **GC–AG** and 0.05% to **AT–AC**, with only 0.02% consisting of other types of noncanonical splice sites.[5] The **AT–AC** introns are very

rare in the human genome, with only six introns showing this splicing pattern.[5] Intron 4 of *SEDL* would be the fifth to be identified. The specific mutation causes the splice site to be recognized as a canonical donor with a '**GT**' end, or to be unrecognized as a splice site. These splicing mechanisms are quite different from the one resulting from IVS4 + 4T>C.[14] The IVS4 + 4T>C mutation caused the splice site to be ignored, yielding exon 4a. On the contrary, when an '**AT**' donor changed into '**GT**' as reported in our study, it was identified as a canonical splice donor site in most clones and was less likely to be ignored. Comparing the different transcripts caused by IVS4 + 1A>G with IVS4 + 4T>C would provide greater insight into the splicing mechanism of the consensus sequence.

In genomes, conserved splice sequences are an essential component of the exon splicing process, and they provide a specific molecular signal for the RNA splicing machinery to identify the precise splice points. **GT–AG** introns generally use the U1-U2 snRNPs, whereas the **AT–AC** introns often use the U11-U12 snRNPs.[6] As a consequence, the new **GT** donor site is linked to an **AG** acceptor and not to the 'original' **AC** acceptor site. The reason why so many cryptic products were formed may be due to the disruption of U11-U12 splicing machinery and the activation of U1-U2 splicing machinery.

All of our transcripts resulting from variable splicing yielded altered SEDL proteins, most of which were truncated, leaving only about half of the NH_2-terminus. The putative Golgi-targeting domains within the COOH-terminus were also affected. This may be why the patients from this family are shorter (in the range of 115.5–133.5 cm) than those described elsewhere. However, more evidence is needed on this point.

To our knowledge, this is the only instance where a single nucleotide mutation activated so many cryptic splice sites, and it represents the only natural mutation known to cause a rare noncanonical splice site and retrieve a canonical site. It is also the only RNA splicing mechanism to expose a new pattern of '**GT–AT**.'

We were able to confirm the splicing effect of this mutation on the *SEDL* gene using a variety of software, including SPL, SPLM and FSPLICE. However, they were not as effective in identifying the cryptic splice acceptor sites, particularly those with the noncanonical ends. Some studies have shown that different components of the splicing machinery are involved in recognizing splice sites with infrequently occurring bases, particularly in the case of some **AT–AC** introns.[6] The related splice mechanisms in the noncanonical splice junctions should be quite different from those canonical patterns.[12] Our present knowledge about how the cell specifies splice sites is not sufficient for accurate and comprehensive computational identification of splice junctions in genomic sequences. Characterization of the splice sites with alternative splicing variants would expand our knowledge of splice junctions

and help us to improve the quality of gene structure prediction programs.[7]

SEDL expression levels of patients and carriers were decreased to 1/20 and 1/6, respectively, compared with those of normal controls. These data can help us to understand and quantify why patients affected by SEDT experience reduced endochondral bone growth at the epiphyses, particularly in the vertebral bodies, and why some adult female carriers experience problems with precocious osteoarthritis.

Taken together, identification of the IVS4 + 1A > G mutation in this SEDT group enables carrier detection and presymptomatic/prenatal diagnosis, but also reveals a significant mutation in the splice consensus. The disruption of the **AT** donor site in a rare **AT–AC** intron led to a **GT** donor site and resulted in a multitude of aberrant transcripts. This is an interesting mechanism and these splice patterns represent a fruitful area for further study on splicing mechanisms.

Acknowledgements

References

1 Gedeon AK, Colley A, Jamieson R *et al*: Identification of the gene (SEDL) causing X-linked spondyloepiphyseal dysplasia tarda. *Nat Genet* 1999; **22**: 400–404.

2 Gecz J, Hillman MA, Gedeon AK, Cox TC, Baker E, Mulley JC: Gene structure and expression study of the SEDL gene for spondyloepiphyseal dysplasia tarda. *Genomics* 2000; **69**: 242–251.

3 Mumm S, Zhang X, Vacca M, D'Esposito M, Whyte MP: The sedlin gene for spondyloepiphyseal dysplasia tarda escapes X-inactivation and contains a non-canonical splice site. *Gene* 2001; **273**: 285–293.

4 Mumm S, Zhang X, Gottesman GS, McAlister WH, Whyte MP: Preonset studies of spondyloepiphyseal dysplasia tarda caused by a novel 2-base pair deletion in SEDL encoding sedlin. *J Bone Miner Res* 2001; **16**: 2245–2250.

5 Burset M, Seledtsov IA, Solovyev VV: Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res* 2000; **28**: 4364–4375.

6 Wu Q, Krainer AR: AT-AC pre-mRNA splicing mechanisms and conservation of minor introns in voltage-gated ion channel genes. *Mol Cell Biol* 1999; **19**: 3225–3236.

7 Burset M, Seledtsov IA, Solovyev VV: SpliceDB: database of canonical and non-canonical mammalian splice sites. *Nucleic Acids Res* 2001; **29**: 255–259.

8 Savarirayan R, Thompson E, Gecz J: Spondyloepiphyseal dysplasia tarda (SEDL, MIM no. 313400). *Eur J Hum Genet* 2003; **11**: 639–642.

9 Miller SA, Dykes DD, Polesky HF: A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988; **16**: 1215.

10 Zhao X, Tang R, Gao B *et al*: Functional variants in the promoter region of chitinase 3-like 1 (CHI3L1) and susceptibility to schizophrenia. *Am J Hum Genet* 2007; **80**: 12–18.

11 Gedeon AK, Tiller GE, Le Merrer M *et al*: The molecular basis of X-linked spondyloepiphyseal dysplasia tarda. *Am J Hum Genet* 2001; **68**: 1386–1397.

12 Shapiro MB, Senapathy P: RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 1987; **15**: 7155–7174.

13 Dietz HC: Nonsense mutations and altered splice-site selection. *Am J Hum Genet* 1997; **60**: 729–730.

14 Shaw MA, Brunetti-Pierri N, Kadasi L *et al*: Identification of three novel SEDL mutations, including mutation in the rare, non-canonical splice site of exon 4. *Clin Genet* 2003; **64**: 235–242.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)