npg

ARTICLE

Splicing mutations in glycogen-storage disease type II: evaluation of the full spectrum of mutations and their relation to patients' phenotypes

Stefania Zampieri[1,4], Emanuele Buratti[2,4], Silvia Dominissini[3], Anna Lisa Montalvo[3], Maria Gabriela Pittis[3], Bruno Bembi[1] and Andrea Dardis[*,1]

Glycogen-storage disease type II is an autosomal recessive-inherited disorder due to the deficiency of acid α-glucosidase. A large number of mutations in the acid α-glucosidase gene have been described to date. Among them, ~15% are variations that may affect mRNA splicing process. In this study, we have for the first time comprehensively reviewed the available information on splicing mutations of the acid α-glucosidase gene and we have evaluated their possible impact on the splicing process using different *in silico* approaches. Out of the 39 different *GAA*-sequence variations described, an *in silico* analysis using seven different programs showed that 97% of them are predicted to have an impact on the splicing process. Moreover, this analysis showed a quite good correlation between the impact of the mutation on the splicing process and the clinical phenotype. In addition, we have performed the functional characterization of three novel sequence variants found in Italian patients and still uncharacterized. Using a minigene system, we have confirmed their pathogenic nature. In conclusion, this study has shown that *in silico* analysis represents a useful tool to select mutations that affect the splicing process of the acid α-glucosidase gene and provides an updated picture of all this kind of mutations reported till now.

INTRODUCTION

Glycogen-storage disease type II (GSDII; Pompe disease, acid maltase deficiency, MIM no. 232300) is an autosomal recessive-inherited disorder due to the deficiency of acid α-glucosidase (*GAA*; E. C.3.2.1.20) that results in impaired glycogen degradation, which accumulates within the lysosomes. The *GAA* gene (MIM no. 606800) has been localized to human chromosome 17q25.2–25.3. The enzyme is synthesized as an inactive precursor of 110 kD, which is transported to the prelysosomal and lysosomal compartment via the mannose-6-phosphate receptor where it is processed into a 95 kD intermediate and the fully active forms of 76 and 70 kD.[1–4]

Clinically, GSDII encompasses a continuous spectrum of phenotypes from a rapidly progressive infantile form to a slowly progressive late-onset (LO) form. Classic infantile GSDII manifests soon after birth and is characterized by absent or nearly absent enzyme activity, severe muscle weakness, cardiomyopathy and respiratory insufficiency, which typically lead to death within the first year of life.[3,5–7]

LO GSDII comprises all milder subtypes: partial enzyme deficiency manifests in children and adults as slowly progressive skeletal muscle weakness without cardiac involvement. Respiratory muscle weakness, particularly of the diaphragm, is the leading cause of death in the LO cases.[3,5,7–9]

A large number of sequence variations in the *GAA* gene have been described to date (http://www-fgg.eur.nl/ch1/pompe/en/?Molecular_ aspects:Mutations). Among them, ~15% are variations that may affect pre-messenger RNA (pre-mRNA)-splicing process.

Pre-mRNA defects seem to have a role in almost all known genetic disorders.[10,11] However, unless the mutation affects the highly conserved nucleotides at the exon 3'ss and 5'ss boundaries, it has often been very difficult to show a clear correlation between a suspected mutation and the disease. Recently, several methods have been developed to evaluate the clinical effect of mutations that may cause splicing defects.[12] As is intuitively obvious, direct analysis of the mature mRNA from the patient remains the most reliable method to determine whether or not a genetic variation affects splicing. However, cells/RNA from the patient might not be available or the transcript may be expressed only in highly selected tissues, making this approach not always possible. To overcome this problem, alternative systems such us minigene-based assays have been used.[13] All these approaches, however, require a substantial amount of time and skill, if they have to be applied to a large number of putative splicing mutations.

For this reason, several *in silico* approaches to assess the effects of sequence variants on splicing have been developed. In general, these splice-prediction programs (SPPs) evaluate the effect of putative splicing mutations on the strength of 5' and 3' splice-site sequences, or search for potential changes within the vast array of splicing regulatory elements (SREs) known to this date. Although the

predictions obtained are usually not enough to establish with sufficiently high accuracy the clinical impact of genetic variations on splicing, it has been proposed that SPPs could be used to perform a first selection of those variants that may have an effect on the pre-mRNA splicing before starting with time-consuming and labor-intensive mRNA analysis.[14]

In this study, we have comprehensively reviewed the available information on splicing mutations of the *GAA* gene and we have evaluated the possible impact of these genetic variations on pre-mRNA-splicing process using different *in silico* approaches. In addition, using a minigene system assay, we have performed the functional characterization of three sequence variants previously found in Italian patients affected with LO GSDII.

MATERIALS AND METHODS

Mutation nomenclature
All mutations are described according to the mutation nomenclature, considering nucleotide +1 as the A of the first ATG translation initiation codon (http://www.hgvs.org/mutnomen).[15,16] Nucleotide numbers are derived from cDNA *GAA* sequence (RefSeq cDNA Y00839.1).

Splice-site prediction of intronic variants previously described in the *GAA* gene
The sequence environment of all acceptor and donor sites was analyzed using Splice Site Prediction by Neural Network, NNSPLICE http://www.fruitfly.org/seq_tools/splice.html/.[17] Maximum entropy scores were obtained using the software based on the maximum entropy principle, MaxEntscan

Table 1 Sequences of oligonucleotides used for PCR amplification and SDM of *GAA* gene

Primers used for PCR amplification

6F	5′-CCGAAGCTTGATGTGGTCCTGCAGCCGAGC-3′
8R	5′-CCGGAATTCCACGATCATCATGTAGCGCCG-3′
10F	5′-CCGAAGCTTGATGGTATGGCCCGGGTCCACTGCCTTCCCCG-3′
12R	5′-CCGGAATTCCGGTCACTGTGGGAGGCGATGGCTTCGG-3′
17F	5′-CCGAAGCTTGATGGTGCCAGTAGAGGCCCTT-3′
19R	5′-CCGGAATTCCGGTCACTTGGTGTCGGGGCTGTAGG-3′

Primers used for SDM

1194mutF	5′-CCCACTTCCCCCTGGAGAGTTGGGGTGGTGG-3′
1194mutR	5′-CCACCACCCCAACTCTCCAGGGGGAAGTGGG-3′
1626mutF	5′-GAGCTGGAGAACCCACCGTACGTGCCTGGTCAGCTCGCCC-3′
1626mutR	5′-GGGCGAGCTGACCAGGCACGTACGGTGGGTTCTCCAGCTC-3′
2646mutF	5′-GTCATCTTCCTGGCCAGGAATGAGTCCTGGGGCTGCTCAG-3′
2646mutR	5′-CTGAGCAGCCCCAGGACTCATTCCTGGCCAGGAAGATGAC-3′

(http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html/).[18] The H-bond scores were calculated at http://www.uni-duesseldorf.de/rna/html/hbond_score.php.[19] Finally, the Sroogle scores were obtained using the software available at http://sroogle.tau.ac.il/.[20]

In addition, the potential effect of the nucleotide variants on SREs was also analyzed with ESEfinder (http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi),[21,22] RESCUE-ESE (http://genes.mit.edu/burgelab/rescue-ese/)[23] and PESX (http://cubweb.biology.columbia.edu/pesx/).[24,25]

Minigene constructs
To evaluate the expression of exons 7, 11 and 18 of the *GAA* gene, wild-type minigenes *GAA1194wt*, *GAA1626wt* and *GAA2646wt* were obtained by insertion in the pcDNA3 plasmid of PCR fragments containing the genomic *GAA* sequence from exons 6–8, 10–12 or 17–19, respectively. PCR amplification was performed using primers 6F–8R, 10F–12R and 17F–19R (listed in Table 1). The forward and reverse primers carried a *Hind*III and *EcoR*1 restriction site, respectively. Mutated minigenes *GAA1194m*, *GAA1626m* and *GAA2646m* carrying mutations c.1194+2T>A, c.1626C>G and c.2646_2646+1delTG, respectively, were prepared by site-directed mutagenesis (SDM) using the Quickchange Site-Directed Mutagenesis Kit (Stratagene, Cedar Creek, TX, USA) according to the manufacturer's instructions. Primers used for the SDM are listed in Table 1. Each clone was entirely sequenced to confirm that no other mutations were introduced.

Cell culture and transient transfection
COS-1, CHO and Hep3B cells were grown on monolayers in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, 2 mM L-glutamine and 50 mg/ml penicillin/streptomycin (Gibco, Paisley, UK). HeLa cells were cultured in RPMI 1640 supplemented with 10% fetal calf serum, 2 mM L-glutamine and 50 mg/ml penicillin/streptomycin (Gibco). Cells were transfected with Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA) using 4 μg of total plasmid DNA Endofree purified (Sigma, St Louis, MO, USA) following the manufacturer's instructions.

Minigene splicing assay
COS-1, CHO, Hep3b and HeLa cells were transfected with the wild-type and mutant minigene constructs. Total RNA was extracted after 48 h using TRIzol reagent (Gibco) and analyzed by RT-PCR. Reverse transcription was performed using the oligo (dT) primer; the PCR reaction was carried out with a forward vector-specific primer (5′-AGGGAGACCCAAGCTTGATG-3′) and the reverse primers 8R, 12R or 19R (Table 1). PCR products were resolved in a 1% agarose gel and sequenced.

RESULTS
The complete spectrum of *GAA* mutations that may affect pre-mRNA splicing is shown in Figure 1. So far, 39 different sequence variations that may alter the splicing process have been described (http://www.fgg.eur.nl/ch1/pompe/en/?Molecular_aspects:Mutations). All of them were analyzed

Figure 1 Schematic representation of *GAA* gene. Exons are represented by black squares and introns by lines. The position and the sequence variation of the GAA mutations that may affect pre-mRNA splicing are indicated by arrows.

npg

Table 2 *In silico* evaluation of 39 *GAA* sequence variants that may affect the pre-mRNA splicing

Modification	Position	MES	NNSplice	HBond	ESEfinder	Rescue-ESE	PESX	DG	Senep	Observed consequence	Ref
		MUT ss score (% of WT ss score)						*Sroogle*		*Observed consequence*	*Ref*
Donor 5'ss											
c.546G>A	Ex 2	NR	NR	5.8 (−38)	SRp40: 3.9 (−10)	None found	Loss 2 Pese	2.2 (+61)	57.60 (−17)	Normal spliced mRNA 6.3% of NC	29
c.546G>C	Ex 2	NR	NR	5.8 (−38)	SRp40: 3.5 (−19)	None found	Loss 2 Pese	−2.3 (+59)	56.32 (−19)		30
c.546G>T	Ex 2	NR	NR	5.8 (−38)	SF2/AF2 (Ig-BRCA): 2.3 (−23); SRp40: 3.7 (−14)	None found	None found	−2 (+64)	57.04 (−18)	Normal spliced mRNA reduced with respect to NC	31
c.546+1G>T	In 2	NR	NR	5.8 (−38)	SF2/AF2 (Ig-BRCA): 3.7 (+23)	None found	Loss 1 Pese	NR	NR		32
c.692+1G>C	In 3	NR	NR	NR	None found	None found	None found	NR	NR	r.0	33
c.692+5G>T	In 3	NR	NR	5.10 (−58)	SC35: 4.5(+32); SRp40: 4.7 (+88)	None found	None found	−0.3 (+92)	58.22 (−17)		34
c.955+2T>G	In 5	1.19 (−87)	NR	NR	SRp40: 2.2 (−19)	None found	1 Pese creation	NR	NR		34
c.1194+2T>A	In 7	1.91 (−81)	NR	NR	None found	None found	4 Pese creation	NR	NR		33
c.1194+2T>C	In 7	2.34 (−77)	NR	NR	None found	None found	3 Pese creation	−3.5 (+51)	69.19 (−20)	Retention 89 nt In 7	35
c.1326+1G>A	In 8	NR	NR	NR	None found	None found	None found	NR	NR	r.0	36
c.1437+2T>C	In 9	1.00 (−89)	NR	NR	SF2/AF2: 3.2 (+88)	None found	None found	−4.4 (+48)	67.39 (−20)	Skipping 111 nt Ex 9	37
c.1437G>A	Ex 9	2.67 (−70)	NR	9.3 (−34)	None found	None found	None found	−3.8 (+55)	72.59 (−14)	Normal spliced mRNA 1.2% of NC	34,35
c.1551+1G>C	In 10	NR	NR	NR	SC35: 4.0 (+14)	None found	1 Pese creation	NR	NR	Skipping 114 nt Ex 10	37–39
c.1636+1G>C	In 11	NR	NR (no change)	NR	SRp40: 2.0 (−31)	None found	1 Pese creation	NR	NR		34
c.1636+5G>C	In 11	NR	NR (no change)	NR	None found	None found	1 Pese creation	−1.0 (+68)	62.99 (−16)	Retention 957 nt In 11	40
c.1636+5G>T	In 11	NR	NR (no change) new cryptic ss creation,	8.30 (−27)	None found	None found	None found	−0.9 (+71)	66.48 (−12)	Normal spliced mRNA 13.7% of NC	34
c.1888+1G>A	In 13	1.42 (−85)	NR	NR	SF2/AF2: 3.8 (+65); SF2/AF2 (Ig-BRCA): 3.2 (+33)	+2 ESE	None found	NR	NR		34
c.2040G>A	Ex 14	NR	NR	8.90 (−26)	SC35: 3.7 (+12)	None found	Loss 2 Pese	−3.8 (+45)	64.18 (−16)	Retention 190 nt In 14	41
c.2189+1G>A	In 15	NR	NR	NR	None found	+1 ESE	None found	NR	NR		34
c.2331+2T>A	In 16	3.27 (−71)	NR	NR	None found	None found	Loss 1 Pese	NR	NR		34
c.2331+2T>C	In 16	3.70 (−68)	NR	NR	None found	None found	Loss 1 Pese	−3.8 (+49)	68.28 (−20)	Skipping 16 nt Ex 16	42
c.2331+4A>G	In 16	8.27 (−28)	0.28 (−75)	12.5 (−22)	SRp40: 1.8 (+40)	None found	Loss 1 Pese	−7.00 (no change)	72.42 (−15)	Normal spliced mRNA 3.6% of NC	43
c.2646_2646 +1delTG	Ex 18-In 18	NR	NR	NR	None found	None found	None found	NR	NR		33
c.2646+2T>A	In 18	NR	NR	NR	SRp40: 4.0 (+67); SRp55: 2.6 (+86)	+1 ESE	None found	NR	NR		44

Table 2 (Continued)

Modification	Position	MUT ss score (% of WT ss score)						Sroogle		Observed consequence	Ref
		MES	NNSplice	HBond	ESEfinder	Rescue-ESE	PESX	DG (PSSM)	Senep		
Acceptor 3′ ss											
c.-32-2A>G	In 1	5.75 (−58)	NR	NA	SF2/AF2: 2.3 (−70) SF2/AF2 (Ig-BRCA): 3.3 (−20) SRp55: 2.7 (+29)	None found	None found	66.02 (−26)			34
c.-32-3C>A	In 1	9.35 (−32)	0.94 (no change)	NA	SF2/AF2: 4.5 (+36) SRp55: 1.5 (−29)	None found	None found	85.46 (no change)		Skipping 578 nt Ex 2	40
c.1076-1G>C	In 6	NR	NR	NA	SF2/AF2: 3.2 (+33) SRp55: 1.5 (−50)	None found	Loss 1 Pese	56.11 (−30) Branch site NR[a]		Retention 79 nt In 6 and 89 nt In 7	30,33
c.1195-2A>G	In 7	NR	NR	NA	SF2/AF2: 4.0 (+21)	None found	None found	58.39 (−29)			45
c.1327-2A>G	In 8	NR	NR	NA	SF2/AF2: 1.3 (−70) SF2/AF2 (Ig-BRCA): 2.8 (−38) SRp40: 2.6 (−33)	None found	Loss 1 Pese	64.57 (−21)			46
c.1438-1G>T	In 9	NR	NR	NA	None found	None found	None found	60.63 (−28)			34
c.1438-1G>C	In 9	NR	NR	NA	SRp55: 4.3 (+34)	None found	None found	60.63 (−28)			5
c.1552-3C>G	In 10	NR	NR	NA	SRp40: 3.0 (−12)	+1 ESE	3 Pese creation	75.66 (−10)		-Normal spliced mRNA 8.3% of NC -SV1: retention 28 nt in 3′ In 10 (2.8% of NC)	43
c.1755-1G>A	In 12	NR	NR	NA	SF2/AF2: 3.0 (−14) SF2/AF2 (Ig-BRCA): 3.6 (−12) SC35: 2.6 (−19) SRp40: 3.1 (−14) SRp55: 1.7 (+113%)	None found	None found	56.86 (−29)			39
c.2041-1G>A	In 14	NR	NR new cryptic ss creation	NA	SC35: 3.3 (+11) SRp55: 2.9 (−38)	+1 ESE	1 Pese creation	NR		Skipping 6 nt at 5′ Ex 15	34
c.2041-2A>C	In 14	1.00 (−89)	NR new cryptic ss creation	NA	SRp55: 3.0 (−36%)	None found	None found	NR		Skipping 6 nt at 5′ Ex 15	29
Others								DG	Senep		
c.1626C>G	Ex 11	NR	NR (no change) new cryptic ss creation	11.38 (no change)	None found	None found	None found	NR	NR		33
c.-32-13T>G	In 1	11.96 (−13)	0.98 (no change)	NA	SRp55: 3.8 (+40)	None found	None found	87.03 (no change) PolyPyrimidin site[a] Algorithm by Kol et al[47] 0.88 (no change) Algorithm by Schwartz et al[48] 10.71 (−13%)		-Normal splice product -SV1 retention 36 nt 5′ In 1+ skipping 578 nt Ex 2 -SV2 skipping 578 nt Ex 2 -SV3 retention 60 nt 3′ Ex 2	38

Table 2 (Continued)

| Modification | Position | MUT ss score (% of WT ss score) | | | | Rescue-ESE | PESX | Sroogle | | Observed consequence | Ref |
		MES	NNSplice	HBond	ESEfinder			DG	Senep		
c.1076-22T>G	In 6	3.73 (no change)	0.38 (no change) New cryptic ss creation	NA	SF2/AF2: 4.9 (+69) SF2/AF2 (Ig-BRCA): 4.8 (+55) SRp40: 4.9 (+69)	+1 ESE	None found	79.56 (no change)		-SV1 skipping 120nt Ex 6. Retention 21nt In 6. -SV2 retention 393nt In 5. Skipping 45nt Ex 6. Utilization of a cryptic donor site in Ex 6. Retention 21nt In 6	49,50
c.1195-8G>T	In 7	9.13 (+16)	0.73 (+62)	NA	SRp55: 3.0 (−21)	None found	1 Pese creation	84.32 (no change)		Retention 21nt In 6	34

Abbreviations: DG, delta free energy; Ex, exon; Esefinder, exonic splicing enhancer finder; HBond, hydrogen bond; In, intron; MES, MaxEntScan (maximum entropy scan); Mut ss, mutant splice site; NA, non-applicable; NC, normal control; NNSplice, neural network splice site prediction; NR, not recognizable because below detection threshold; Pese, putative exonic splicing enhancer; PESX, putative exonic splicing enhancer/silencer; PSSM, position-specific scoring matrix; Rescue-ESE, relative enhancer and silencer classification by unanimous enrichment; r.0, mRNA transcribed from the mutant allele non-detected in patient's fibroblasts; Senep, senepathy; Sroogle, splicing regulation online graphical engine; SV, splice variant; WT ss, wild-type splice site.
Threshold for NNSplice was set at 0.1. For the other programs, default thresholds were used.
Mutations barely predicted to affect the splicing process are indicated by underline.
[a]For polypyrimidine tract and branch-site recognition, the algorithms developed by Kol et al[47] and Schwartz et al[48] were used.

in silico using three Splice Site Prediction (SSP) programs to evaluate the potential effect on 5′ss and 3′ss strengths (MaxEntScan, NNSplice and HBond). These programs were chosen on the basis that they are very common in the scientific literature that focuses on splicing mutations and use different approaches to measure splice-site strengths. For example, the MaxEntScan server scores 5′splice-site sequences taking into account known dependencies between adjacent and non-adjacent positions of the splice-site consensus, the NNSplice program is a machine-learning approach that recognizes sequence patterns, once it was trained with DNA sequences encompassing authentic splice sites and HBond analyses individual hydrogen-bonding patterns to the U1 snRNA 5′ end irrespective of nucleotide frequencies.

In addition, all mutations were analyzed with respect to the eventual presence of SR-binding sites or general enhancer/silencer SREs using the ESEfinder, RescueESE and PESX programs. Also in this case, these programs use different approaches to identify and score putative splicing regulatory sequences. The ESEfinder program is based on functional SELEX-binding analyses to identify, using a matrix-scoring approach, putative SR protein-binding sites and their eventual disruption following the introduction of nucleotide changes. On the other hand, both RescueESE and PESX servers have used statistical analyses to identify putative regulatory motifs (hexamers or octamers) whose frequency differs between exons/introns and between exons with weak or strong splice sites (RescueESE) or within internal non-coding exons vs adjacent pseudoexons or the 5′ UTR of intronless genes (PESX).

In general, for all these programs, the user is allowed to make his/her own decision with regards to decide whether a prediction is positive or negative. In our analyses, we have therefore decided to look predominantly to score variations rather than their basic numerical value. In general, we have considered that a mutant score should have at least a 10% difference from the wild-type one to be classified as a deleterious change. For all these programs, of course, the key question is represented by the degree of reliance that one can place in their predictions. In general, because the donor and acceptor elements tend to be reasonably conserved in human genes, the programs that evaluate these elements seem to be rather more successful than those that aim to target the much more loosely conserved SREs. However, all bioinformatics approaches tend to suffer from individual drawbacks that have been summarized recently by Spurdle et al.[26] In addition, for obvious reasons, none of these programs can easily take into account all the factors that influence splice-site selection and include transcriptional effects, influence of genomic context, relative abundance of splicing factors and, of course, the presence of still unidentified regulatory elements.[27] Therefore, it is common consensus that to evaluate a potential splicing mutation it is advisable to use as many programs as possible to perform the analysis.[28] For this reason, in addition to all these programs, we have also decided to test these mutations using a more recent in silico prediction program (Sroogle)[20] that employs an integrated approach to answer the question whether a nucleotide variation has a good chance of representing a splicing mutation as opposed to a benign polymorphism. The results of this analysis are reported in Table 2.

Out of the 39 mutations taken into consideration, 38 are clearly predicted to affect the splicing process by at least one of the programs used (but very often more), whereas only the c.-32-13T>G scores as negative with all programs except for MES and Sroogle where it barely clears the 10% cutoff score.

Among the programs tested, the most successful ones are represented by those that aim to predict the eventual disruption of the natural 5′ and 3′ splice sites. In fact, 37 mutations are clearly predicted to affect the 3′ or 5′ splice-site signal strengths. More interestingly, it is

Table 3 Correlation between the effect of *GAA* mutations on the splicing process and the clinical phenotype

Mutation	Clinical phenotype	Effect on RNA level (when available)	Reference
c.-32-3C>A	I	Exon skipping	40
c.692+1G>C	I	r.0	33
c.1194+2T>C	I	Intron retention	34
c.1195-2A>G	I		45
c.1326+1G>A	I	r.0	36
c.1327-2A>G	I		46
c.1437+2T>C	I	Exon skipping	37
c.1437G>A	I	Normal spliced mRNA 1.2% of NC	34,35
c.1551+1G>C	I	Exon skipping	37–39
c.1636+5G>C	I	Intron retention	40
c.1755-1G>A	I		39
c.2040G>A	I	Intron retention	41
c.2041-2A>C	I	Skipping 6 nt at 5′ exon 15	29
c.2331+2T>C	I	Skipping 16 nt exon16	42
c.2646+2T>A	I		44
c.-32-13T>G	LO	Normal spliced mRNA reduced with respect to NC	38
c.546G>A	LO	Normal spliced mRNA 6.3% of NC	29
c.546G>T	LO	Normal spliced mRNA reduced with respect to NC	31
c.1076-22T>G	LO	Exon skipping	49,50
c.1552-3C>G	LO	Normal spliced mRNA 8.3% of NC	43
c.1626C>G	LO		33
c.1636+5G>T	LO	Normal spliced mRNA 13.7% of NC	34
c.2331+4A>G	LO	Normal spliced mRNA 3.6% of NC	43
c.546G>C	LO (in association with c.-32-13T>G)		30
c.546+1G>T	LO (in association with c.-32-13T>G)		32
c.1076-1G>C	LO (in association with c.-32-13T>G)	Retention 79 nt intron 6 and 89 nt intron 7	30,33
c.1194+2T>A	LO (in association with c.-32-13T>G)		35
c.2646_2646+1delTG	LO (in association with c.-32-13T>G)		33

Abbreviations: I, infantile phenotype; LO, late-onset phenotype; NC, normal control; r.0, mutant transcript non-detected *in vivo*.

worthy to point out that with possibly just one exception (c.-32-3C>A), where MaxEntScan performs apparently better, in all other cases, the integrated Sroogle program performs in the same way as any of the individual programs. In some cases, such as c.1636+5G>C and c.1636+5G>T, it even performs better than the MaxEntScan and NNSplice programs.

We also looked at whether the programs were able to predict the degree of changes correctly. However, no correlation between the degree of score change and the consequence on the mRNA splicing observed *in vivo* was found.

Next, to analyze the possible correlation between the effect of *GAA* mutations on the splicing process and the clinical phenotype, we reviewed the clinical presentation of patients carrying these mutations. As shown in Table 3, 15 splicing mutations were reported in infantile patients and the impact on pre-mRNA processing has been reported for 11 of them.

Interestingly, 10 of these 11 mutations associated with infantile onset caused a severe defect on mRNA splicing, leading to either exon skipping, shift in the reading frame or the lack of the corresponding transcript *in vivo*.[29,33,34,36–42] In the light of these findings, therefore, it is not unexpected that patients carrying these mutations presented a severe form of GSDII. In fact, only in cells from a patient carrying the c.1437G>A mutation, a very low amount of normal mRNA was detected (1.2%),[34,35] suggesting that the amount of enzyme translated from this mRNA is probably not enough to prevent the development of a severe form of the disease.

In contrast to this picture, in patients affected with LO GSDII, eight mutations were reported in a homozygous state or in association with

severe mutations. In this case, the effect on pre-mRNA processing has been analyzed *in vivo* for seven of them. It is interesting to note that in the cells of patients presenting six of these mutations, variable amounts of normal spliced *GAA* mRNA have been detected (from 3.6 to 13.7% of *GAA* mRNA expressed in normal cells).[29,31,34,38,43] As a result, these data may explain the less severe phenotype reported in these patients.

Five of these LO GSDII-associated splicing mutations were found in patients with the common c.-32-13T>G mutation.[30,32,33,35] It is worth mentioning that the mutation c.-32-13T>G, the most common one among the Caucasian LO GSDII patients,[3] has been previously studied *in vitro* using a minigene system assay.[51] The results have shown that this mutation does not completely prevent normal splicing, as low levels of the correctly spliced mRNA were generated with the mutant construct. In fact, three splice variants (SV1, SV2 and SV3-Table2) were observed with both the wild-type and the mutant constructs, indicating that these forms represent normal alternative spliced products. Thus, the major effect of the mutation seems to affect mostly the overall splicing efficiency of the pre-mRNA transcript rather than in a qualitative way. As the presence of the c.-32-13T>G mutation in one allele is enough to determine the LO phenotype, independently of the mutation present in the second allele, it was not possible to establish a correlation between mutations present in association with the c.-32-13T>G allele and the clinical presentation. The only exception is represented by mutation c.1076-1G>C, which had been analyzed *in vivo*. In this case, RT-PCR analysis performed in cultured fibroblasts from a patient carrying the c.1076-1G>C mutation showed the inclusion of introns 6 (79 bp) and 7 (89 bp) into the
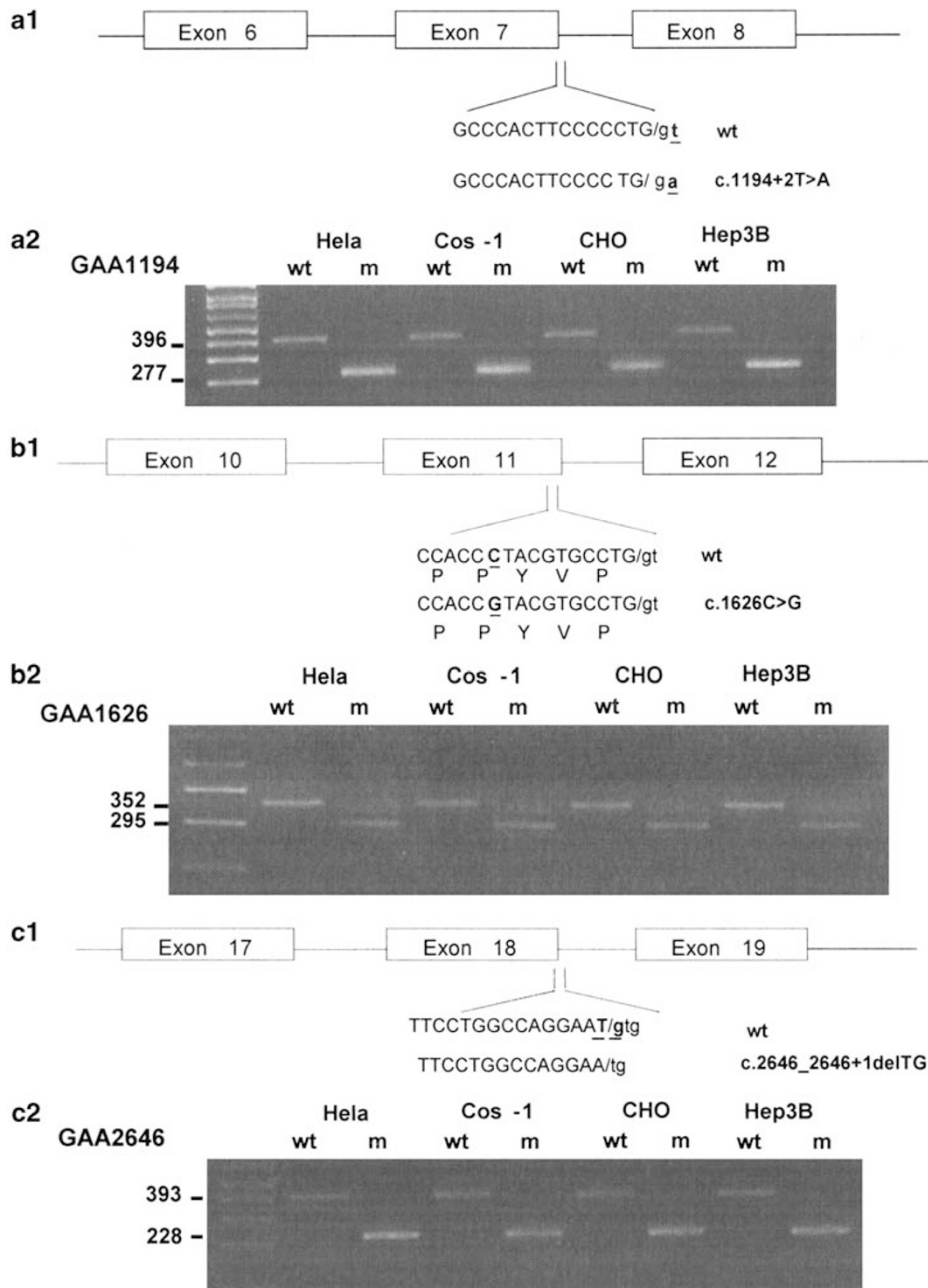
Figure 2 Schematic representation of the *GAA* regions affected by splicing mutations. Mutations c.1194+2T>A, c.1626C>G and c.2646_2646+1delTG (panels **a1**, **b1** and **c1**, respectively) are highlighted in bold. RT-PCR analysis of the *GAA* mRNA in cells transfected with wt and minigenes containing mutations c.1194+2T>A, c.1626C>G and c.2646_2646+1delTG (panels **a2**, **b2** and **c2**, respectively) using a forward vector-specific primer that amplified only the minigene product.

transcribed mRNA. If we consider the fact that the c.1076-1G>C mutation has also been found in homozygosis in infantile GSDII patients,[30] these data indicate that c.1076-1G>C could be classified as a severe mutation.

To complete this analysis of *GAA*-splicing mutations, it should also be noted that in a previous study, we have characterized the mutation profile of the *GAA* gene in 40 Italian patients with LO GSDII.[33] Overall, five mutations that might have affected the splicing process were found. However, as RNA and/or cells from patients carrying three of them (c.1626C>G, c.1194+2T>A and c.2646_2646+1delTG) were unavailable, their deleterious effect could not be confirmed. The sequence variation c.1626C>G was found in the homozygous state

and did not disrupt the reading frame and codon usage, whereas mutations c.1194+2T>A and c.2646_2646+1delTG affected the consensus 5′ splice donor sites of exons 7 and 18, respectively. Several SPPs (Table 2) clearly predicted that mutation c.1626C>G would create a novel donor site, which would have caused the exclusion of 11 bp of exon 11, whereas the mutated sequences c.1194+2T>A and c.2646_2646+1delTG would no longer be recognized as donor sites. Therefore, to test the predicted effects of these mutations on *GAA* pre-mRNA splicing, we have now performed a functional splicing assay.

As shown in Figure 2, cells transfected with the mutant constructs bearing these mutations produced aberrant transcripts in all cases. However, it is worth noting that in cells transfected with the

GAA1626m construct, a low amount of a transcript similar in size to the normal one was also present. As the wild-type acceptor and donor sites of intron 11 seem not to be abolished, this data suggest that even in the presence of this mutation, a low amount of wild-type transcript would still be produced.

Sequencing analysis of the aberrant PCR product showed that as predicted by the SPPs, the c.1194+2T>A and c.2646_2646+1delTG mutations led to the skipping of exons 7 and 18, respectively (Figure 3a and c). In the light of these results, both mutations should be considered to have a severe impact on pre-mRNA splicing. However, in both patients, these mutations were found in association with the common c.-32-13T>G mutation (Table 3) that, as mentioned above, would probably have been enough to determine the observed LO phenotype.

Finally, the analysis of the RT-PCR products found in cells transfected with the GAA1626m construct showed that, unexpectedly, the shorter transcript was not only lacking 11 nt of exon 11 as predicted by the SPPs, but also a portion of 46 nucleotides of exon 12 was missing (Figure 3b). In any case, the presence of low amounts of the wild-type transcript was confirmed (data not shown). If we consider that the c.1626C>G mutation was found at the homozygous state, this finding may also explain the LO form of the disease observed in this patient (Table 2).

DISCUSSION

The complexity of the splicing process has revealed that almost all types of nucleotide changes may potentially affect the pre-mRNA-splicing process. However, distinguishing between benign and disease-causing sequence substitutions may be challenging. In fact, it is not realistic to test all nucleotide modifications at RNA level. Therefore, several in silico approaches to assess the effects of sequence variants on splicing have been developed and it has been proposed that they could be used to perform a first selection of those variants that may have an effect on the pre-mRNA splicing.[28,52]

In this work, we have reviewed the genetic variations reported in the literature that might affect the GAA pre-mRNA-splicing process. The in silico analysis of 39 sequence variants showed that this approach may be useful to select those mutations that affect the splicing process. As previously mentioned, among these programs, the most successful ones are represented by those that aim to predict the eventual disruption of the natural 5′ and 3′ splice sites. This observation is consistent with previous analyses that show that even a moderate consensus, such as the one present in the 5′ and 3′ ss splice sites of higher eukaryotes, greatly helps this kind of computational studies. It is common knowledge, in fact, that the much more loosely conserved enhancer or silencer elements are much more difficult to be defined, and in this respect, it is worthy to point out the comment by Rogan PK[53] that highlights how only one in four programs, which aim to predict ESE motifs, was actually capable of correctly recognizing the creation of a novel ESE element in a pathological pseudoexon activation event.[54]

Considering that splicing signals are highly context specific, it is not surprising that they are hard to predict and that programs that uses an integrated approach perform better. In keeping with this hypothesis, the analysis performed here indeed has shown that the integrated Sroogle program performed at least as well as any of the individual programs. This conclusion is supported by the observation that another integrated program (Automated Splice Site Analysis, ASSA, available at https://splice.uwo.ca)[55] used to analyze splicing mutations in the RB1 gene provided several advantages with respect to individual programs.[28] Therefore, it is possible to conclude that integrated
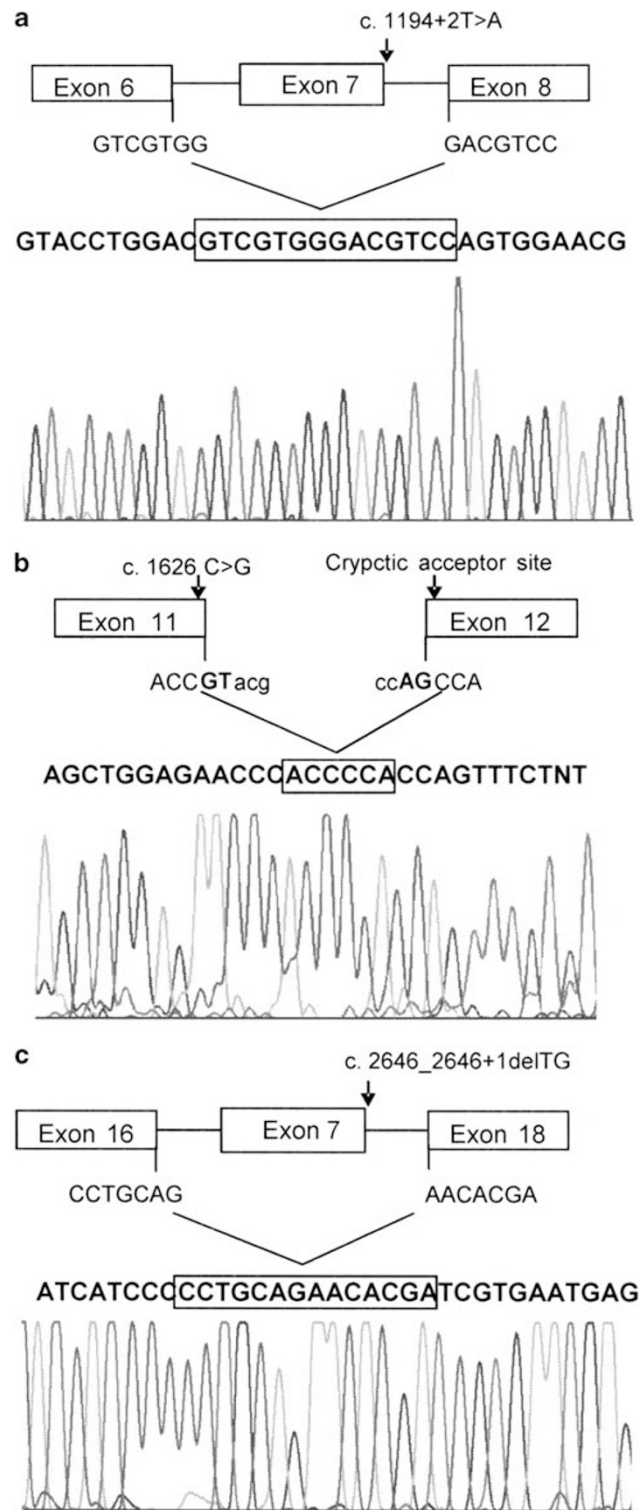

Figure 3 Sequencing analysis of the RT-PCR products obtained from amplification of GAA mRNA in transfected cells. (**a**) Cells transfected with c.1194+2T>Am construct, (**b**) cells transfected with c.1626C>Gm construct and (**c**) cells transfected with c.2646_2646+1delTGm constructs.

approaches, such as the one represented by Sroogle, may successfully substitute for the use of separate analyses programs to shorten the time to complete a preliminary screening of a large number of mutations.

There are also some cases, however, where *in silico* analysis did not perform so well. Not surprisingly, these cases include mutations that occur more distant to the splice site (ie, c.-32-13T>G and c.1195-8G>T). However, in the case of c.1195-8G>T mutation, the use of the ESEfinder and PESX programs might have substituted for the relative inability of the other softwares to detect the splicing-modifying potential of this mutations. In the case of the c.-32-13T>G substitution, even using a very favorable degree of score change (10%), it was barely predicted to affect the splicing process. There are several reasons that may account for this failure but the fact that this mutation still allows a certain amount of normal splicing (as discussed above) most probably places it in the limit of the detection threshold of these bioinformatics programs. Taken together, therefore, ~97% of the *GAA*-splicing mutations taken in consideration in this analysis could have been clearly predicted to have an impact on the splicing process by *in silico* analysis alone.

The remarkable heterogeneity of mutations found in GSDII patients makes it difficult to correlate the genotype with the phenotype. However, the results reviewed here show a good correlation between the impact of the mutation on the splicing process and the clinical phenotype (ie, the more severe the impact on the splicing process, the more likely the presence of an infantile form of the disease and *vice versa*).

Finally, using a minigene assay, we confirmed the pathogenic effect of the three sequence variants found in Italian patients affected with GSDII. All these three mutations were predicted to affect the splicing process by SPPs. However, in the case of mutation c.1626C>G, although several SPPs predicted that the mutation would create a novel donor site, which would cause the exclusion of 11 bp of exon 11, the functional analysis showed that this mutation not only creates a new donor site as predicted but also causes the activation of a cryptic splice site localized in exon 12, downstream of the normal acceptor site of exon 11.

In conclusion, this study has shown that *in silico* analysis represents a useful tool to select mutations that affect the splicing process of the acid α-glucosidase. This type of analysis is quite straightforward and reliable. However, it is worth to highlight the importance of functional studies for the correct evaluation of sequences' variations. In fact, a comprehensive analysis of the mechanism by which a sequence variant affect the mRNA splicing is crucial to analyze possible correlations between the mutation and the clinical phenotype, and to evaluate the feasibility of using emerging splicing-based therapeutic approaches.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Hoefsloot LH, Hoogeveen-Westerveld M, Reuser AJ, Oostra BA: Characterization of the human lysosomal alpha-glucosidase gene. *Biochem J* 1990a; **272**: 493–497.
2 Martiniuk F, Bodkin M, Tzall S, Hirschhorn R, Hirschhorn R: Isolation and partial characterization of the structural gene for human acid alpha glucosidase. *DNA Cell Biol* 1991; **10**: 283–292.
3 Hirschhorn R, Reuser AJJ: Glycogen storage disease type II: acid α-glucosidase (acid maltase) deficiency; in Scriver CR, Beaudet AL, Sly WS, Valle D (eds): *The Metabolic and Molecular Basis of Inherited Disease*. New York: McGraw-Hill, 2001, Vol 3, pp 3389–3420.
4 Moreland RJ, Jin X, Zhang XK *et al*: Lysosomal acid alpha-glucosidase consists of four different peptides processed from a single chain precursor. *J Biol Chem* 2005; **280**: 6780–6791.
5 Raben N, Plotz P, Byrne BJ: Acid alpha-glucosidase deficiency (glycogenosis type II, Pompe disease). *Curr Mol Med* 2002; **2**: 145–166.
6 Van den Hout HM, Hop W, van Diggelen OP *et al*: The natural course of infantile Pompe's disease: 20 original cases compared with 133 cases from the literature. *Pediatrics* 2003; **112**: 332–340.
7 Kishnani PS, Howell RR: Pompe disease in infants and children. *J Pediatr* 2004; **144** (5 Suppl): S35–S43.
8 Hagemans ML, Janssens AC, Winkel LP *et al*: Late-onset Pompe disease primarily affects quality of life in physical health domains. *Neurology* 2004; **63**: 1688–1692.
9 Hagemans ML, Winkel LP, Van Doorn PA *et al*: Clinical manifestation and natural course of late-onset Pompe's disease in 54 Dutch patients. *Brain* 2005; **128**(Part 3): 671–677.
10 Tazi J, Bakkour N, Stamm S: Alternative splicing and disease. *Biochim Biophys Acta* 2009; **1792**: 14–26.
11 Wang GS, Cooper TA: Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 2007; **8**: 749–761.
12 Baralle D, Lucassen A, Buratti E: Missed threads. The impact of pre-mRNA splicing defects on clinical practice. Review. *EMBO Rep* 2009; **10**: 810–816.
13 Vibe-Pedersen K, Kornblihtt AR, Baralle FE: Expression of a human alpha-globin/fibronectin gene hybrid generates two mRNAs by alternative splicing. *EMBO J* 1984; **3**: 2511–2516.
14 Vreeswijk MP, Kraan JN, van der Klift HM *et al*: Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs. *Hum Mutat* 2009; **30**: 107–114.
15 den Dunnen JT, Antonarakis SE: Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 2000; **15**: 7–12.
16 den Dunnen JT, Paalman MH: Standardizing mutation nomenclature: why bother? *Hum Mutat* 2003; **22**: 181–182.
17 Reese MG, Eeckman FH, Kulp D, Haussler D: Improved splice site detection in Genie. *J Comput Biol* 1997; **4**: 311–323.
18 Yeo G, Burge CB: Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004; **11**: 377–394.
19 Freund M, Asang C, Kammler S *et al*: A novel approach to describe a U1 snRNA binding site. *Nucleic Acids Res* 2003; **31**: 6963–6975.
20 Schwartz S, Hall E, Ast G: SROOGLE: webserver for integrative, user-friendly visualization of splicing signals. *Nucleic Acids Res* 2009; **37** (Web Server issue): W189–W192.
21 Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, Krainer AR: An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum Mol Genet* 2006; **15**: 2490–2508.
22 Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR: ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 2003; **3**: 3568–3571.
23 Fairbrother WG, Yeh RF, Sharp PA, Burge CB: Predictive identification of exonic splicing enhancers in human genes. *Science* 2002; **297**: 1007–1013.
24 Zhang XH, Chasin LA: Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev* 2004; **18**: 1241–1250.
25 Zhang XH, Kangsamaksin T, Chao MS, Banerjee JK, Chasin LA: Exon inclusion is dependent on predictable exonic splicing enhancers. *Mol Cell Biol* 2005; **25**: 7323–7332.
26 Spurdle AB, Couch FJ, Hogervorst FB, Radice P, Sinilnikova OM, IARC Unclassified Genetic Variants Working Group: Prediction and assessment of splicing alterations: implications for clinical testing. *Hum Mutat* 2008; **29**: 1304–1313.
27 Buratti E, Baralle M, Baralle FE: Defective splicing, disease and therapy: searching for master checkpoints in exon definition. *Nucleic Acids Res* 2006; **34**: 3494–3510.
28 Houdayer C, Dehainault C, Mattler C *et al*: Evaluation of in silico splice tools for decision-making in molecular diagnosis. *Hum Mutat* 2008; **29**: 975–982.
29 Hermans MM, van Leenen D, Kroos MA *et al*: Twenty-two novel mutations in the lysosomal alpha-glucosidase gene (GAA) underscore the genotype-phenotype correlation in glycogen storage disease type II. *Hum Mutat* 2004; **23**: 47–56.
30 Gort L, Coll MJ, Chabás A: Glycogen storage disease type II in Spanish patients: high frequency of c.1076-1G>C mutation. *Mol Genet Metab* 2007; **92**: 183–187.
31 Maimaiti M, Takahashi S, Okajima K *et al*: Silent exonic mutation in the acid-alpha-glycosidase gene that causes glycogen storage disease type II by affecting mRNA splicing. *J Hum Genet* 2009; **54**: 493–496.
32 Nascimbeni AC, Fanin M, Tasca E, Angelini C: Molecular pathology and enzyme processing in various phenotypes of acid maltase deficiency. *Neurology* 2008; **19**: 617–626.
33 Montalvo AL, Bembi B, Donnarumma M *et al*: Mutation profile of the GAA gene in 40 Italian patients with late onset glycogen storage disease type II. *Hum Mutat* 2006; **27**: 999–1006.
34 Kroos M, Pomponio RJ, van Vliet L *et al*: GAA Database Consortium. Update of the Pompe disease mutation database with 107 sequence variants and a format for severity rating. *Hum Mutat* 2008; **29**: E13–E26.
35 Wan L, Lee CC, Hsu CM *et al*: Identification of eight novel mutations of the acid alpha-glucosidase gene causing the infantile or juvenile form of glycogen storage disease type II. *J Neurol* 2008; **255**: 831–838.
36 Raben N, Lee E, Lee L, Hirschhorn R, Plotz PH: Novel mutations in African American patients with glycogen storage disease Type II. Mutations in brief no. 209. Online. *Hum Mutat* 1999; **13**: 83–84.

37 Stroppiano M, Bonuccelli G, Corsolini F, Filocamo M: Aberrant splicing at catalytic site as cause of infantile onset glycogen storage disease type II (GSDII): molecular identification of a novel IVS9 (+2GT→GC) in combination with rare IVS10 (+1GT→CT). *Am J Med Genet* 2001; **101**: 55–58.

38 Huie ML, Chen AS, Tsujino S *et al*: Aberrant splicing in adult onset glycogen storage disease type II (GSDII): molecular identification of an IVS1 (-13T→G) mutation in a majority of patients and a novel IVS10 (+1GT→CT) mutation. *Hum Mol Genet* 1994; **3**: 2231–2236.

39 Palmer RE, Amartino HM, Niizawa G, Blanco M, Pomponio RJ, Chamoles NA: Pompe disease (glycogen storage disease type II) in Argentineans: clinical manifestations and identification of 9 novel mutations. *Neuromuscul Disord* 2007; **17**: 16–22.

40 McCready ME, Carson NL, Chakraborty P *et al*: Development of a clinical assay for detection of GAA mutations and characterization of the GAA mutation spectrum in a Canadian cohort of individuals with glycogen storage disease, type II. *Mol Genet Metab* 2007; **92**: 325–335.

41 Hermans MM, van Leenen D, Kroos MA, Reuser AJ: Mutation detection in glycogen storage-disease type II by RT-PCR and automated sequencing. *Biochem Bioph Res Commun* 1997; **241**: 414–418.

42 Fernandez-Hojas R, Huie ML, Navarro C *et al*: Identification of six novel mutations in the acid alpha-glucosidase gene in three Spanish patients with infantile onset glycogen storage disease type II (Pompe disease). *Neuromuscul Disord* 2002; **12**: 159–166.

43 Pittis MG, Donnarumma M, Montalvo AL *et al*: Molecular and functional characterization of eight novel GAA mutations in Italian infants with Pompe disease. *Hum Mutat* 2008; **29**: E27–E36.

44 Oba-Shinjo SM, da Silva R, Andrade FG *et al*: Pompe disease in a Brazilian series: clinical and molecular analyses with identification of nine new mutations. *J Neurol* 2009; **256**: 1881–1890.

45 Hamdan MA, Almalik MH, Mirghani HM: Early administration of enzyme replacement therapy for Pompe disease: short-term follow-up results. *J Inherit Metab Dis* 2008; e-pub ahead of print 12 December 2008.

46 Kroos M, Manta P, Mavridou I *et al*: Seven cases of Pompe disease from Greece. *J Inherit Metab Dis* 2006; **29**: 556–563.

47 Kol G, Lev-Maor G, Ast G: Human-mouse comparative analysis reveals that branch-site plasticity contributes to splicing regulation. *Hum Mol Genet* 2005; **14**: 1559–1568.

48 Schwartz SH, Silva J, Burstein D, Pupko T, Eyras E, Ast G: Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res* 2008; **18**: 88–103.

49 Adams EM, Becker JA, Griffith L, Segal A, Plotz PH, Raben N: Glycogenosis type II: a juvenile-specific mutation with an unusual splicing pattern and a shared mutation in African Americans. *Hum Mutat* 1997; **10**: 128–134.

50 Vorgerd M, Burwinkel B, Reichmann H, Malin JP, Kilimann MW: Adult-onset glycogen storage disease type II: phenotypic and allelic heterogeneity in German patients. *Neurogenetics* 1998; **1**: 205–211.

51 Raben N, Nichols RC, Martiniuk F, Plotz PH: A model of mRNA splicing in adult lysosomal storage disease (glycogenosis type II). *Hum Mol Genet* 1996; **5**: 995–1000.

52 Baralle D, Baralle M: Splicing in action: assessing disease causing sequence changes. *J Med Genet* 2005; **42**: 737–748.

53 Rogan PK: Deeper understanding of unclassified intronic variants and ESEs. *Hum Mutat* 2010; **31**: V.

54 Homolova K, Zavadakova P, Doktor TK, Schroeder LD, Kozich V, Andresen BS: The deep intronic c.903+469T>C mutation in the MTRR gene creates an SF2/ASF binding exonic splicing enhancer, which leads to pseudoexon activation and causes the cblE type of homocystinuria. *Hum Mutat* 2010; **31**: 437–444.

55 Nalla VK, Rogan PK: Automated splicing mutation analysis by information theory. *Hum Mutat* 2005; **25**: 334–342.