

# Identification and Characterization of Aberrant *GAA* Pre-mRNA Splicing in Pompe Disease Using a Generic Approach

Atze J. Bergsma,<sup>1,2,3</sup> Marian Kroos,<sup>1,3</sup> Marianne Hoogeveen-Westerveld,<sup>1,3</sup> Dicky Halley,<sup>4</sup> Ans T. van der Ploeg,<sup>2,3</sup> and W. W. M. Pim Pijnappel<sup>1,2,3\*</sup>

<sup>1</sup>Molecular Stem Cell Biology, Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, The Netherlands;

<sup>2</sup>Department of Pediatrics, Erasmus MC University Medical Center, Rotterdam, The Netherlands; <sup>3</sup>Center for Lysosomal and Metabolic Diseases, Erasmus MC University Medical Center, Rotterdam, The Netherlands; <sup>4</sup>Molecular Diagnostics, Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, The Netherlands

Communicated by Garry R. Cutting

Received 8 July 2014; accepted revised manuscript 11 September 2014.

Published online 22 September 2014 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22705

**ABSTRACT:** Identification of pathogenic variants in monogenic diseases is an important aspect of diagnosis, genetic counseling, and prediction of disease severity. Pathogenic mechanisms involved include changes in gene expression, RNA processing, and protein translation. Variants affecting pre-mRNA splicing are difficult to predict due to the complex mechanism of splicing regulation. A generic approach to systematically detect and characterize effects of sequence variants on splicing would improve current diagnostic practice. Here, it is shown that such approach is feasible by combining flanking exon RT-PCR, sequence analysis of PCR products, and exon-internal quantitative RT-PCR for all coding exons. Application of this approach to one novel and six previously published variants in the acid-alpha glucosidase (*GAA*) gene causing Pompe disease enabled detection of a total of 11 novel splicing events. Aberrant splicing included cryptic splice-site usage, intron retention, and exon skipping. Importantly, the extent of leaky wild-type splicing correlated with disease onset and severity. These results indicate that this approach enables sensitive detection and in-depth characterization of variants affecting splicing, many of which are still unrecognized or poorly understood. The approach is generic and should be adaptable for application to other monogenic diseases to aid in improved diagnostics. *Hum Mutat* 00:1–12, 2014. © 2014 Wiley Periodicals, Inc.

**KEY WORDS:** *GAA*; splicing; monogenic disease; generic assay; Pompe disease

## Introduction

Identification of variants involved in human-inherited disease is important for diagnosis, prediction of disease severity, and

genetic counseling. Around 7,000 rare genetic diseases are known, and in approximately 50% of cases the gene involved has been identified [Boycott et al., 2013]. At present, over 148,000 variants have been deposited and include missense/nonsense variants (55%), small deletions (15%), splicing variants (9%), gross deletions (7%), and other variants at lower frequency (<http://www.hgmd.cf.ac.uk/ac/index.php>). Detection of variants may be missed in diagnostic settings that involve sequencing of the exons only. This would exclude detection of variants in promoters, UTRs, or introns, which may affect gene expression, RNA stability, or pre-mRNA splicing. Improved methods for DNA sequencing result in the identification of an increasing number of sequence variants of which the potential pathogenic nature must be determined. With the exception of nonsense variants, it is difficult to predict the effect of other types of variants. Coding variants can be easily studied by introducing the variant in the cDNA from a gene of interest and testing the effect on protein activity in a transient transfection assay. It would be valuable to apply a routine functional approach that would detect aberrant RNA expression or processing without the need for prior DNA sequence analysis and targeted analysis of that variant.

The difficulty to predict the effects of variants on pre-mRNA splicing is related to the complex mechanism of splicing regulation. Many cis-acting elements that are located either close to or more distant from the splice sites may be involved. These include the polypyrimidine tract, branchpoint, and loosely defined regulatory elements present in either exons (exonic splicing enhancers and exon splicing silencers) or introns (intronic splicing enhancers and intronic splicing silencers) (reviewed in [Havens et al., 2013]). A number of splicing prediction programs exist [Reese et al., 1997; Pertea et al., 2001; Yeo and Burge, 2004; Desmet et al., 2009], but they produce different predictions for the same variant, obscuring data interpretation. Furthermore, when weakening of a splice site is likely from *in silico* predictions, the effect on alternative splicing is difficult to predict, and may include exon skipping/inclusion, intron retention, utilization of a cryptic splice site, or generation of a novel splice site. The outcome is important for the pathogenic nature of the variant. For example, perfect skipping of an exon while the reading frame is unchanged may generate a truncated protein with significant residual function, whereas a change of the reading frame results in a premature termination codon leading to mRNA degradation via the nonsense-mediated decay (NMD) pathway [Palacios, 2013].

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

\*Correspondence to: W.W.M. Pim Pijnappel, Department of Clinical Genetics, Erasmus MC University Medical Center, Faculty building, Room Ee-916a, P.O. Box 2040, Rotterdam 3000 CA, The Netherlands. E-mail: w.pijnappel@erasmusmc.nl

Contract grant sponsor: Sophia Children's Hospital Foundation (SSWO) (grant S-687).

In this study, a generic approach has been developed for the identification and characterization of variants affecting splicing and mRNA expression in a fast, cheap, and accurate manner. It relies on a combination of RT-PCR, qPCR, and sequence analysis using standard reactions amenable to validation in a diagnostic setting. The approach has been applied to Pompe disease (Glycogen Storage Disease II [GSDII]; MIM #232300), an autosomal-recessive monogenic disease caused by variants in the acid-alpha glucosidase gene (*GAA*; MIM #606800; reviewed in van der Ploeg and Reuser 2008). Although >460 *GAA* variants have been described (<http://www.pompecenter.nl>), and at least 38 variants have been predicted to affect splicing, only a few splicing variants have been fully characterized [Huie et al., 1994b; Boerkoel et al., 1995; Zampieri et al., 2011; Herzog et al., 2012; Dardis et al., 2014]. Pompe disease is caused by failure to degrade lysosomal glycogen, resulting in glycogen accumulation that is particularly harmful for cardiac and skeletal muscle cells. Severe variants that completely abrogate *GAA* enzyme activity cause a classic infantile disease course with hypertrophic cardiomyopathy, general skeletal muscle weakness, and respiratory failure and result in death within 1.5 years of life. Milder variants leave partial *GAA* enzyme activity resulting in a milder phenotype with onset varying from childhood to adult [Kroos et al., 2007; Laforet et al., 2013]. In general, a higher residual enzyme activity in primary fibroblasts is associated with later onset of Pompe disease [Umaphathysivam et al., 2005]. Enzyme replacement therapy (ERT) has been developed for Pompe disease, in which recombinant human *GAA* protein is administered intravenously [Van den Hout et al., 2000; Kishnani et al., 2006]. This treatment can rescue the lives of patients with classic infantile Pompe disease [Kishnani et al., 2007] and delay disease progression of patients with later onset [Regnery et al., 2012; Gungor et al., 2013; Toscano and Schoser, 2013], but the effects are heterogeneous. One important prognostic factor for the success of ERT is the stage of disease at the time of diagnosis combined with the severity of the *GAA* genotype. The detection and analysis of the pathogenic nature of sequence variants can therefore be relevant for the nature and timing of therapeutic intervention.

Application of the approach to three previously characterized Pompe patients confirmed published information. Testing of the effects of previously partially or uncharacterized variants present in five patients resulted in the detection, characterization, and quantification of novel-splicing products that could be linked to disease phenotype. The approach is generic and should be applicable to other monogenic diseases as well. The information obtained is useful for confirmation of diagnosis, prediction of disease severity, and genetic counseling.

## Materials and Methods

Analysis was performed as part of diagnostic practice with informed consent of Pompe patients and a healthy control. NM\_000152.3 was used as reference sequence for *GAA* mRNA. NP\_000143.2 was used as a reference for *GAA* protein. c.1 represents the first nucleotide of the coding region of the *GAA* mRNA. The coordinates for the genomic *GAA* sequence used are chr17:80,101,556–80,119,880 (hg38). Nomenclature was according to HGVS standards [den Dunnen and Antonarakis, 2000]. All variants are listed in our Pompe Mutation Database (<http://www.pompecenter.nl>). Cell culture, DNA, and RNA analysis were performed using standard conditions. Splicing predictions were performed using Alamut Visual v.2.4.2 (interactive software, see Supp. Table S1 for details). The activity of *GAA* in fibroblasts was measured with

4-methylumbelliferone (4-MU)- $\alpha$ -D-glucoopyranoside as substrate as described previously [Kroos et al., 2007]. A more detailed description of *Materials and Methods* can be found in the Supporting Information (Supp. Materials and Methods).

## Results

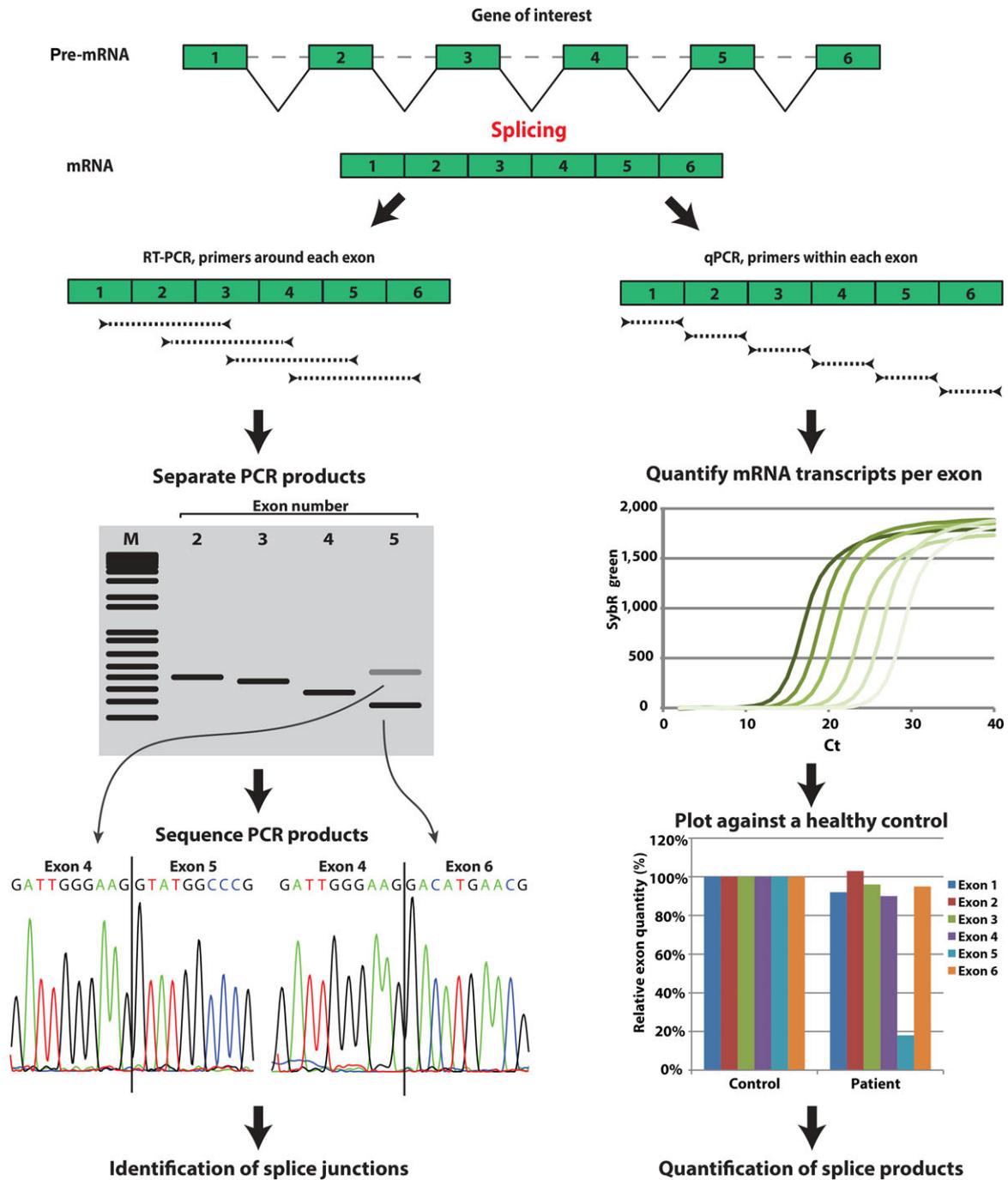
### A Generic Approach to Detect Aberrant Splicing

A number of variants that may affect pre-mRNA splicing can be predicted using several algorithms, but (1) this is not the case for all variants; (2) the molecular consequences of alternative splicing are very difficult to predict; and (3) predictions must be verified experimentally. The standard experimental approach is to use total RNA from primary cells such as fibroblasts and to perform RT-PCR of individual exons that are suspected to be affected, based on the location of the variant. Aberrant PCR products obtained are sequenced to define new splice-site boundaries. We reasoned that currently available methods in principle enable a generic approach to experimentally identify and verify bona fide splicing variants. The approach consists of two parts. First (Fig. 1, left), a generic RT-PCR is performed of the mRNA of interest using standard primers that flank each individual canonical exon (flanking exon PCR). The products are separated by agarose gel electrophoresis. Changes in product size are indicative of alternative/aberrant splicing. Splicing junctions can be precisely determined using sequencing of products isolated from gel or by direct sequencing of the PCR reaction. It should be noted that the flanking exon PCR is at most semiquantitative and yields primarily qualitative information on the identity of the aberrantly spliced products. Second (Fig. 1, right), a standard qPCR is performed to quantify each individual coding exon (exon-internal qPCR). Primers that anneal within each exon are used. Results are normalized for  $\beta$ -actin mRNA and for expression in a healthy control. Results quantify exon skipping/inclusion, and may also indicate whether a splicing variant allows leaky wild-type splicing.

### Development and Validation of the Approach

#### Healthy control

As previously determined, *GAA* activities in healthy individuals can range between 40 and 160 nmol 4-MU/hr/mg protein [van der Ploeg and Reuser, 2008]. To develop the approach, cells from a healthy control were tested first. This individual showed *GAA* enzyme activity of 122.4 nmol 4-MU/hr/mg protein, representing an average activity within this range (Table 1). To detect splicing junctions and exon sizes, flanking exon PCR analysis was performed on cDNA prepared from primary fibroblasts using primers that annealed to flanking exons (Fig. 2A; Supp. Table S2). Gel electrophoresis and ethidium bromide staining showed the correct molecular weight products in all cases. This indicated canonical splicing for all exons in these cells. Some additional products were observed at minor amounts, notably, just above exon 6 and 7. Sequence analysis indicated that these represent products in which intron 6 was retained. The products were observed in this healthy control as well as in most Pompe patients studied and may indicate noisy aberrant splicing, which is a known phenomenon [Pickrell et al., 2010]. Individual coding exons 2–20 (the translation start codon is localized in exon 2) were quantified using exon-internal qPCR (Fig. 2E). Values were normalized for  $\beta$ -actin expression (as



**Figure 1.** Workflow for the generic analysis of splice-site variants. Qualitative analysis of changes in splice-site usage is performed by PCR using primers annealing to the flanking exons (flanking exon PCR), followed by sequencing (left part). Quantitative analysis of aberrant splicing products is performed using primers annealing within each exon (exon-internal qPCR; right part).

measured by qPCR analysis), and were then ready to use for normalization of patient samples.

### Patients 1, 2, and 3

The approach was validated using variants with known effects on RNA expression and splicing. A detailed description is provided in the legends to Supp. Figures S1–S4 and a brief description is given below. Patient 1 harbored the c.-32–13T>G (IVS1) splicing variant

on one allele, and the c.1636+5G>T variant, which induces NMD, on the other allele. The three major splicing variants normal (N), splicing variant (SV)2, and SV3 caused by the IVS1 allele were identified by flanking exon PCR, and leaky wild-type splicing of 10%–20% was quantified using exon-internal qPCR, in agreement with previous reports [Huie et al., 1994b; Boerkoel et al., 1995; Dardis et al., 2014] (Supp. Fig. S1A–C and F). In addition, an isoform-specific qPCR method was developed that allowed quantification of the splicing variants separately. This confirmed NMD of SV2 and SV3 (Supp. Fig. S1D and E). All five splicing prediction programs used here

**Table 1. Laboratory Diagnosis of Pompe Patients Included in This Study**

	Variant allele 1	Variant allele 2	GAA activity in primary fibroblasts with 4-MU as substrate (nmol 4-MU/hr/mg protein)	Age at diagnosis	Onset infantile/juvenile/adult
Control	-	-	122.4	-	None
Patient 1	c.-32-13T>G (IVS1)	c.1636+5G>T	14.1	59 years	Adult
Patient 2	c.525delT	c.525delT	1.3	0.5 months	Infantile
Patient 3	c.1548G>A	c.2481+102_2646+31del (del ex18)	0.1	3.5 months	Infantile
Patient 4	c.-32-3C>G	c.1551+1G>A	6.9	8.5 years	Juvenile
Patient 5	c.1075G>A	c.1075G>A	0.6	8.5 months	Infantile
Patient 6	c.1552-3C>G	c.1552-3C>G	12.6	16 years	Juvenile
Patient 7	c.1437G>A	c.1437G>A	3.0	37 years	Adult
Patient 8	c.1256A>T	c.1551+1G>T	5.4	1.3 years	Juvenile

Reference sequence used for cDNA annotation is NM\_000152.3.

(SpliceSiteFinder-like [SSF], MaxEntScan [MES], NNSplice [NNS], GeneSplicer [GS], and Human Splicing Finder [HSF]) failed to detect an effect of the IVS1 variant on splicing (Supp. Fig. S2A).

Patient 2 was homozygous for the c.525delT variant, which undergates NMD [Hermans et al., 1994]. Flanking exon PCR detected low levels of products for all exons, which were not derived from genomic DNA contamination (Supp. Fig. S3A and B). This enabled identification of the c.525delT variant by sequence analysis (Supp. Fig. S3A and D). Exon-internal qPCR analysis showed expression of 3%–9% for all exons relative to the healthy control (Supp. Fig. S3C).

Patient 3 carried a well-known deletion removing the entire exon 18 plus its flanking sequences (del ex18, or c.2481+102\_2646+31del) on one allele, and c.1548G>A [Hermans et al., 2004], a nonsense variant that induces NMD, on the other allele. This case is interesting because the splice sites of exon 18 are removed. Flanking exon PCR confirmed that the del ex18 variant resulted in the precise skipping of exon 18 with a normal splice junction between exon 17 and exon 19, as reported previously [Huie et al., 1994a] (Supp. Fig. S4A, B, and D). Exon-internal qPCR analysis showed expression (relative to healthy control) of 3% of the c.1548G>A allele and 40%–50% of the del ex18 allele, the product of which was in-frame (Supp. Fig. S4C). GAA enzyme activities were consistent with leaky wild-type splicing and adult disease onset in patient 1, and the absence of leaky wild-type splicing and classic infantile Pompe disease in patients 2 and 3 (Table 1). Splicing products are summarized in Table 2. Taken together, these results validate that the approach can be used for the detection and identification of normal and aberrant splicing products, for their quantification, and for prediction of disease severity. The sensitivity of the approach can allow sequence identification of a variant inducing NMD.

### Characterization of Novel Splicing Variants

Next, a number of patients were analyzed that contained partially characterized or uncharacterized variants.

#### Patient 4

Patient 4 contained a novel variant at c.-32-3C>G located in intron 1 close to the splice acceptor site of exon 2 (Fig. 2C). This variant is suspected to affect splicing of exon 2 based on its similarity to the published c.-32-3C>A variant [Pittis et al., 2008]. In this study, a perfect skip of exon 2 was reported. Splicing prediction programs indicated that the c.-32-3C>G variant weakens the

splice acceptor site of exon 2 for some but not all programs (Supp. Fig. S2C). The second allele contained a previously reported [Orlikowski et al., 2011] but uncharacterized variant at c.1551+1G>A which is located in intron 10 close to the splice donor site of exon 10 (Fig. 2D). Based on the similarity to the published c.1551+1G>C variant [Huie et al., 1994b; Stroppiano et al., 2001], the c.1551+1G>A variant is suspected to affect exon 10 splicing. Splicing prediction programs indicated loss of the splice donor site of exon 10 (Supp. Fig. S2C).

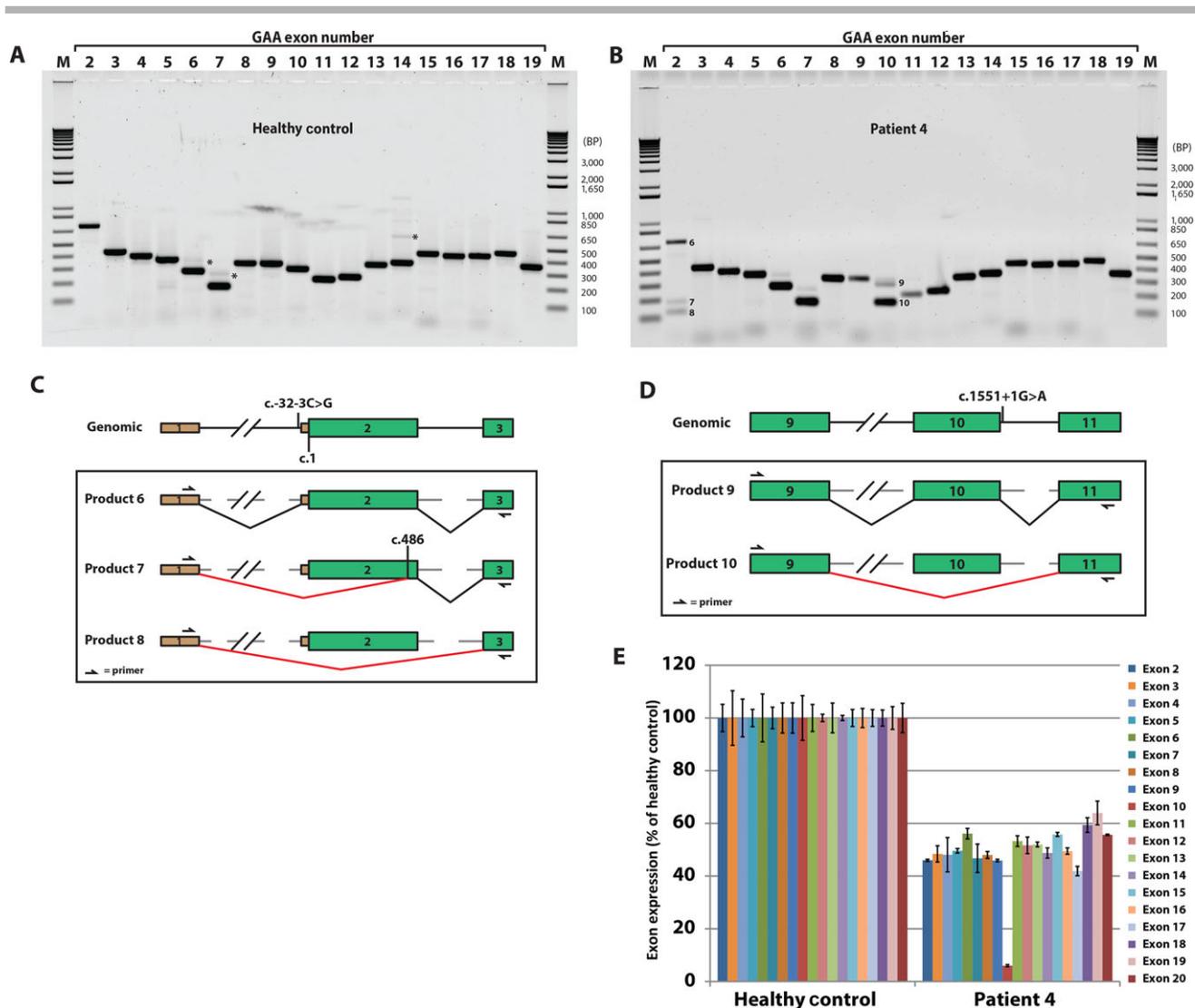
The results of the flanking exon PCR analysis indicated aberrant splicing of two exons: exon 2 and exon 10 (Fig. 2B). Amplification of exon 2 resulted in three major products, numbered 6–8, and sequence analysis indicated that these products included wild-type splicing, partial skipping of exon 2 via the cryptic splice acceptor site at c.486 in exon 2, and perfect skipping of exon 2, respectively (Fig. 2C; Supp. Fig. S5A). This indicates that two independent variants in intron 1, namely, c.-32-13T>G, which is located in the polypyrimidine tract, and c.-32-3C>G, located near the splice acceptor site, have the same qualitative outcome with respect to exon 2 splicing. Splicing prediction programs were insufficient to accurately predict this outcome. Flanking exon PCR amplification of exon 10 resulted in two major products, 9 and 10 (Fig. 2B). Sequence analysis showed that product 9 contained wild-type junctions between exons 9, 10, and 11, and that product 10 represented precise skipping of exon 10 mRNA (Fig. 2D; Supp. Fig. S5B) in which the reading frame remains intact.

To determine the extent of splicing defects, exon-internal qPCR was performed. Exon 10 was expressed at ~6%, whereas all other exons were expressed at ~50% of healthy control levels (Fig. 2E). This is consistent with the idea that the majority of mRNA is derived from the c.1551+1G>A allele in which exon 10 is skipped. The shorter product has an unchanged reading frame and is expected to be stable. In contrast, the c.-32-3C>G allele results in (partial) exon 2 skipping, which is known to result in mRNA degradation analogous to the IVS1 variant. The c.-32-3C>G allele has only a minor contribution to the exon-internal qPCR results. Its contribution can be judged from exon 10 expression, which can result from leaky wild-type splicing of the c.-32-3C>G variant. However, an alternative source for exon 10 expression is leaky wild-type expression of the c.1551+1G>A allele. The very low level of exon 10 expression indicates that both the c.-32-3C>G and the c.1551+1G>A have low or absent levels of leaky wild-type expression. This indicates that the c.-32-3C>G variant may be more severe compared with the IVS1 variant, as the IVS1 variant allows a higher level of wild-type splicing of 10%–20% (Supp. Fig. S1C). In contrast to the situation with, for example, the c.525delT allele of patient 2, in which residual mRNA

**Table 2. Summary of Splicing Events Resulting from the Variants Studied**

Patient	Variant (cDNA HGVS nomenclature)	Location	Reference on patient/codon change	Effect on RNA processing	RNA HGVS nomenclature	Reading frame	Protein HGVS nomenclature	Reference on splicing
1	c.-32-13T>G (IVS1)	Intron 1	Huie et al. (1994b)	Leaky wt splicing Partial skipping exon 2 Perfect skipping exon 2 Intron 11 inclusion	r. = r.-32_486del r.-32_546del r.1636_1637ins1636+1_1636+957; 1636+5g>u	In frame Out of frame Out of frame Out of frame	p. = p.? p.? p.V547Hfs*146	Boerkoel et al. (1995); Huie et al. (1994b); Dardis et al. (2014) Kroos et al. (2006)
2	c.525delT	Exon 2	Hermans et al. (1994)	Premature stop codon	r.525delu	Out of frame	p.E176Rfs*45	Hermans et al. (1994)
3	c.1548G>A	Exon10	Hermans et al. (2004)	Premature stop codon	r.1548g>a	New stop codon	p.W516*	Hermans et al. (2004)
	c.2481+102_2646+31del (del ex18)	Intron 17-intron 18	Huie et al. (1994a)	Deletion of full exon 18	r.2482_2646del	In frame	p.G828_N882del	Huie et al. (1994a)
4	c.-32-3C>G	Intron 1	This study	Leaky wt splicing Partial skipping exon 2 Perfect skipping exon 2	r. = r.-32_486del r.-32_546del	In frame Out of frame Out of frame	p. = p.? p.?	This study This study This study
5	c.1551+1G>A	Intron 10	Orlikowski et al. (2011)	Perfect skipping exon 10	r.1438_1551del	In frame	p.V480_L517del	This study
	c.1075G>A	Exon 6	Muller-Felber et al. (2007)	Deletion of 4 nt of exon 6	r.1072_1075del	Out of frame	p.V358Dfs*33	This study
6	c.1552-3C>G	Intron 10	Kroos et al. (2006)	Leaky wt splicing Full intron 10 inclusion	r. = r.1551_1552ins1551+1_1551+100; 1552-3c>g	In frame Out of frame	p. = p.D518Vfs*7	Kroos et al. (2006) This study
7	c.1437G>A	Exon 9	Kroos et al. (2008)	Partial inclusion intron 10	r.1438_1551delins1437+71_1437+100;1552-3c>g r.1437g>a	In frame	p.V480_L517delins10	Kroos et al. (2006)
8	c.1256A>T	Exon 8	Kroos et al. (2012a)	Leaky wt splicing Perfect skipping exon 9	r.1327_1437del r.1256a>u	In frame In frame	p. = p.D443_K479del p.D419V	Kroos et al. (2008) This study This study
8	c.1551+1G>T	Intron 10	Kroos et al. (2012a)	Partial skipping exon 8 Leaky wt splicing Perfect skipping of exon 10	r.1255_1326del r. = r.1438_1551del	In frame In frame In frame	p.D419_V442del p. = p.V480_L517del	This study This study This study

Reference sequences used for cDNA and protein annotation are NM\_000152.3 and NP\_000143.2, respectively. Intronic sequences are derived chr11:780,101,556-80,119,880 (GRCh38).



**Figure 2.** Splicing analysis of a healthy control and Pompe patient 4. **A:** Flanking exon PCR analysis of a healthy control. Exon numbers are indicated above the lanes. PCR products were separated by electrophoresis on an agarose gel. Asterisks indicate alternative splicing events detected in both healthy individuals and in Pompe patients. **B:** As (A), but for Pompe patient 4. Numbers besides the bands refer to the products analyzed in further detail (see below). **C:** Cartoons of the genomic variant (upper cartoon) and the splicing variants 6–8 (lower cartoons, boxed) detected for patient 4. The translation start site is indicated as c.1. Exons are indicated as boxes. Noncoding exons are in brown; coding exons in green. Introns are depicted as lines. A broken line is used to indicate that the intron is longer than suggested in this drawing. An alternative splice site at c.486 is indicated. **D:** As (C), but now for splicing variants 9 and 10. **E:** Exon-internal qPCR analysis.  $\beta$ -actin was used for normalization. Values obtained from the healthy control were set to 100%. Error bars indicate SD ( $n = 3$ ).

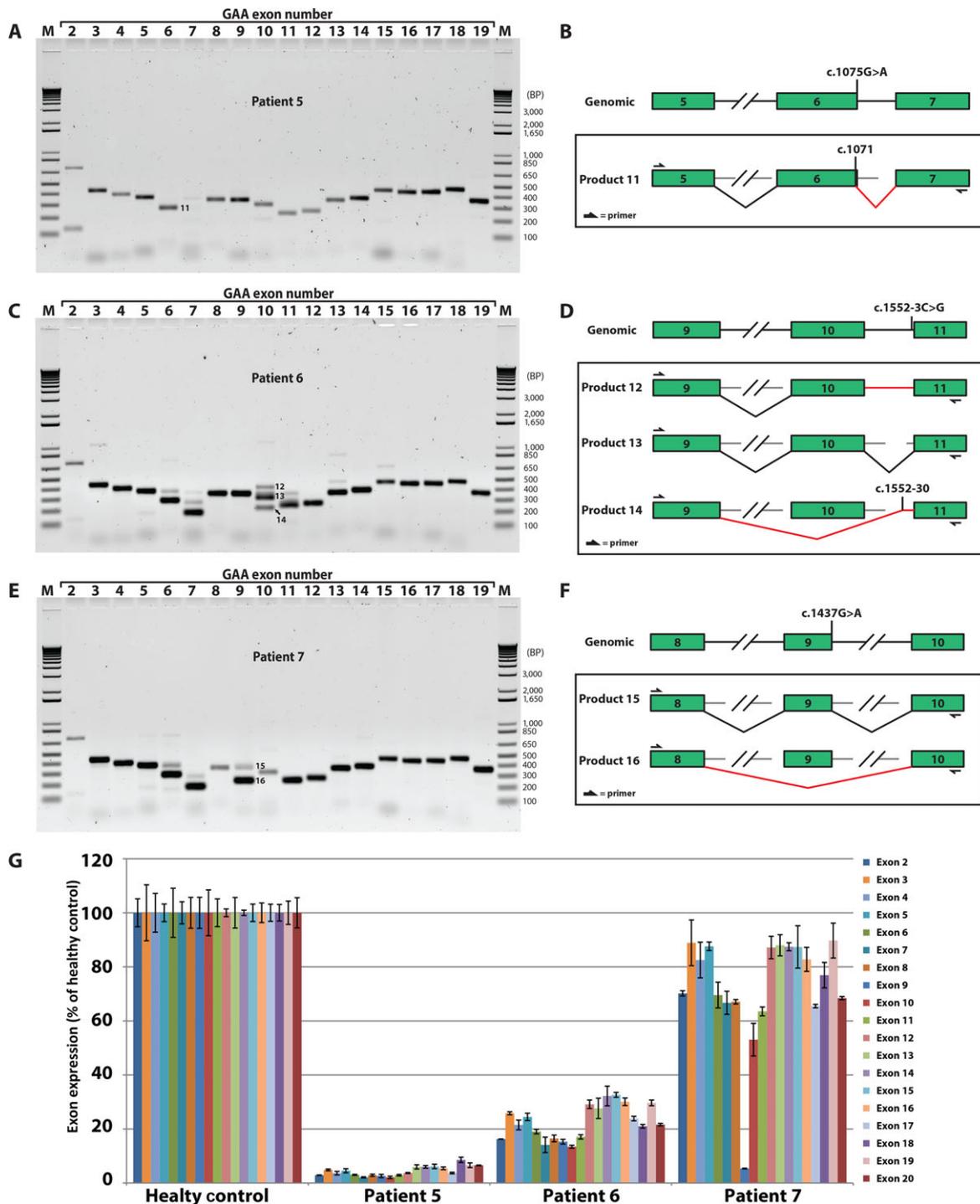
still carried a deletion and concomitant reading frame shift, the low levels of residual mRNA expression in patient 4 reflect wild-type mRNA. In agreement, the clinical course of Pompe disease indicated a juvenile onset for this patient, consistent with a low level of wild-type *GAA* expression and *GAA* enzyme activity levels that were lower compared with adult onset patients (Table 1).

### Patient 5

Patient 5 was homozygous for c.1075G>A, which is a p.G359R missense variant located at the last base pair of exon 6 (Fig. 3B) [Muller-Felber et al., 2007]. This variant has been classified as presumably nonpathogenic with possible effects on splicing [Kroos et al., 2012a]. It is located near the splice donor site of exon 6,

and splicing prediction analysis indicated weakening of this site and strengthening of a cryptic splice donor site 4 nucleotides upstream (Supp. Fig. S2D).

Flanking exon PCR analysis showed absence of a product for exon 7, low levels of the other exons, and a low level of a low MW product for exon 2 (Fig. 3A). Based on the predictions and on the location of this variant in exon 6, we suspected that splicing junctions around exon 6 and 7 may be altered. In agreement, sequencing of the exon 6 PCR product (product 11) showed that the cryptic splice donor site in exon 6 located four nucleotides upstream at c.1071 was used instead (Fig. 3B; Supp. Fig. S6B). This explains the absence of a product for exon 7, as the forward primer for exon 7 amplification has four mismatches due to the changed splice donor site. Remarkably, the flanking exon PCR assay failed to detect leaky wild-type splicing for this variant. This would have resulted in the



**Figure 3.** Splicing analysis of Pompe patients carrying homozygous variants. **A:** Flanking exon PCR analysis of patient 5. **B:** Cartoons of the genomic variants and the splicing variant detected for patient 5. **C:** Flanking exon PCR analysis of patient 6. **D:** Cartoons of the genomic variant and the splicing variants detected for patient 6. **E:** Flanking exon PCR analysis of patient 7. **F:** Cartoon of the genomic variant and the splicing variant detected for patient 7. **G:** Exon-internal qPCR analysis of patients 5-7. Error bars indicate SD ( $n = 3$ ).

presence of a wild-type band for exon 7 amplification, which was not observed. To further investigate splicing of exon 7, an alternative forward primer (qPCR GAA Exon 6 forward; see Supp. Table S3) located in exon 6 was used. The expected product was now obtained, and showed splicing from c.1071 in exon 6 to the canonical splice acceptor site of exon 7 (Supp. Fig. S6A), as was observed for sequence analysis of product 11 (data not shown). The reading frame

of the resulting mRNA has been changed leading to a premature termination codon (Table 2). The low MW product obtained with exon 2 amplification has not been pursued further. It may be caused by a yet unidentified intronic variant. Alternatively, wild-type GAA mRNA is known to have leaky exon 2 skipping, the product of which may be preferentially amplified because of mRNA degradation due to the c.1071 variant.

Quantification of *GAA* mRNA expression using the exon-internal qPCR assay showed that all *GAA* exons were expressed at very low levels, well below levels observed for the IVS1 variant but just above the levels observed for the c.525delT variant (Fig. 3G). This confirmed the notion that leaky wild-type splicing levels in this patient are very low or absent, whereas the majority of the mRNA is unstable. In agreement, very low *GAA* activity in fibroblasts was measured and the patient was diagnosed with the classic infantile form of Pompe disease (Table 1).

### Patient 6

Patient 6 carried a homozygous c.1552-3C>G variant. This variant is located in intron 10 close to exon 11 (Fig. 3D). Flanking exon PCR analysis showed aberrant splicing of exon 10 with three major products (12-14; Fig. 3C). Sequence analysis indicated that in product 14, exon 10 was completely skipped while a splice acceptor site near exon 11 at c.1552-30 was utilized (Fig. 3D; Supp. Fig. S6C). This mRNA leaves the reading frame intact (Table 2). Product 13 was identified as wild-type spliced mRNA. Product 12 consisted of mRNA in which the complete intron 10 was retained. The reading frame is disrupted in this splicing product. While products 13 and 14 have been detected previously [Kroos et al., 2006], product 12 is novel. Interestingly, splicing prediction programs were ambivalent on predicting the extent of utilization of the canonical or the cryptic splice acceptor sites of exon 11 (Supp. Fig. S2F). A priori, weakening of the splice acceptor site of exon 11 would not be expected to result in the skipping of exon 10. Instead, two products seemed more likely: one in which the splice donor site of exon 10 splices to the cryptic acceptor at c.1552-30, resulting in extension of exon 11 with a part of intron 10 and further normal splicing. The other expected product would be a perfect skipping of exon 11. The completely different outcome illustrates that experimental validation is required to analyze the molecular consequences of potential splicing variants.

Quantification of splicing defects was performed with the exon-internal qPCR assay. This showed expression of all exons at ~20% of healthy control levels (Fig. 3G). No extra reduction of exon 10 expression was observed, suggesting that the majority of mRNA included exon 10, favoring products 12 and 13 over 14. The presence of leaky wild-type splicing (product 13) is consistent with residual *GAA* enzyme activity and the milder phenotype of this patient with late juvenile onset Pompe disease (Table 1). In conclusion, c.1552-3C>G results in several splicing defects around exon 10 and intron 10, and it allows leaky wild-type splicing compatible with juvenile onset disease.

### Patient 7

Patient 7 was homozygous for c.1437G>A, a silent variant located at the splice donor site of exon 9 (Fig. 3F). Flanking exon PCR analysis showed two products instead of one for exon 9 amplification, and low yields for exon 8 and exon 10 amplification (Fig. 3E). Sequence analysis indicated that product 15 represented wild-type spliced exon 9, whereas in product 16, exon 9 was perfectly skipped, resulting in a shorter transcript in which the reading frame was unchanged (Fig. 3F; Supp. Fig. S6D). As expected from its location, the c.1437G>A variant was predicted in silico to weaken to splice donor site of exon 9 (Supp. Fig. S2E). Products of exon 8 and exon 10 amplification had correct sizes but lower yield because exon 9 had reduced availability to serve as template for annealing of the reverse PCR primer (for exon 8) or the forward PCR primer (for exon 10).

Quantification using exon-internal qPCR showed near-normal (70%–80% of control) expression levels for all exons except for exon 9, which showed expression of only 5% of healthy control. This suggests expression of 5% residual wild-type mRNA that should allow expression of low levels of wild-type *GAA* enzyme. In agreement, residual *GAA* enzyme activity was measured to be 3 nmol/hr/mg, which is consistent with nonclassic infantile onset (Table 1). It is remarkable that this patient was diagnosed at the age of 37 years, as other patients with higher enzyme activity showed juvenile disease onset. This may indicate the existence of Pompe disease-modifying factors, as suggested previously [Herzog et al., 2012; Kroos et al., 2012b].

In summary, the c.1437G>A variant results in precise skipping of exon 9 leaving the reading frame intact, and allows a low level of leaky wild-type *GAA* splicing.

## Characterization of a Complex Case: Patient 8

### Genotype

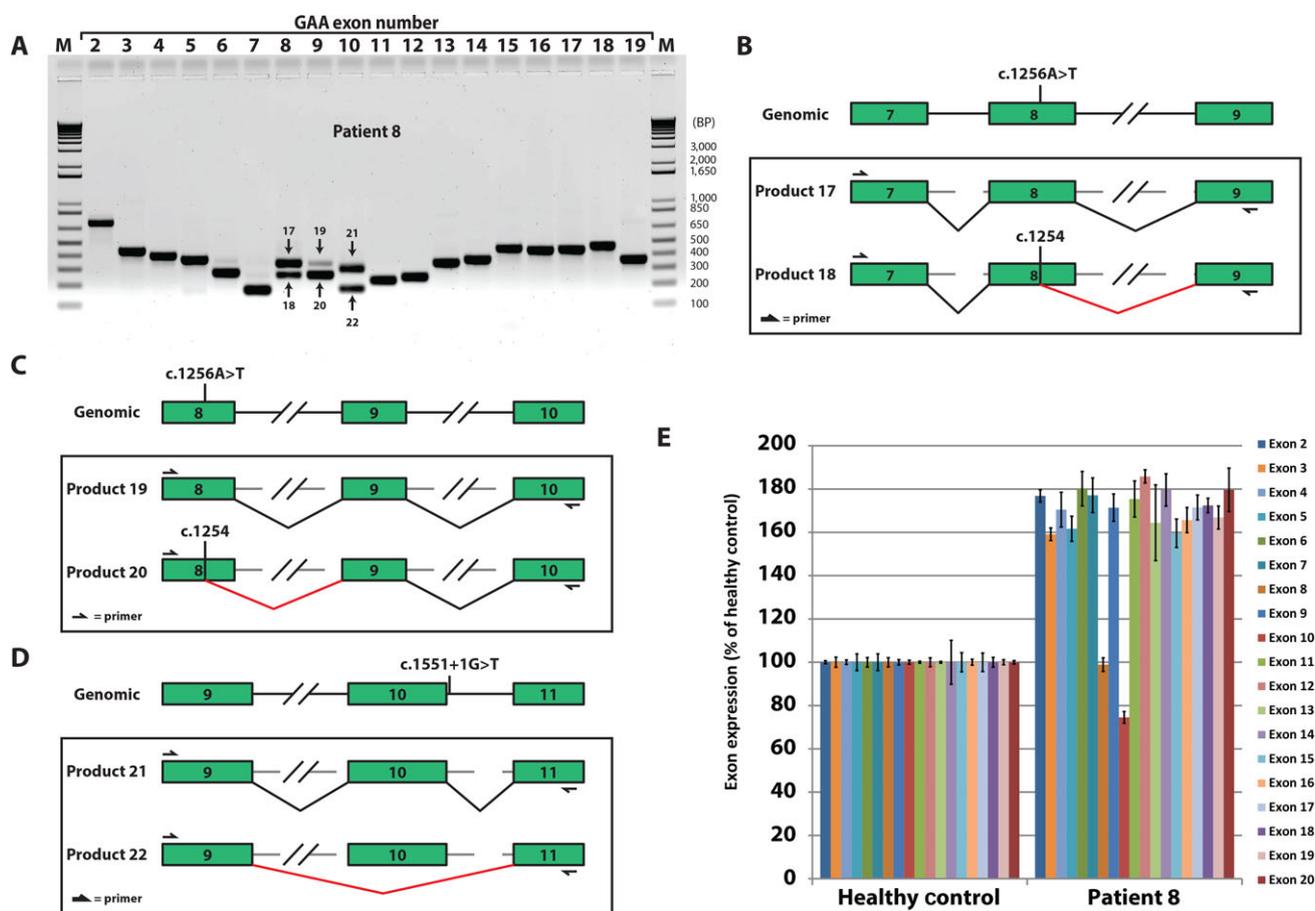
Patient 8 contained the missense variant c.1256A>T on allele 1. It is located in the middle of exon 8, results in p.D419V, and has been classified as mildly pathogenic (Fig. 4B and C) [Kroos et al., 2012a]. The second allele contained a c.1551+1G>T variant, which is located in intron 10 close to the splice donor site of exon 10 [Kroos et al., 2012a]. It resembles the c.1551+1G>A variant described above for patient 4 (Fig. 4D).

### Analysis of splicing products

Flanking exon PCR analysis indicated multiple PCR products from amplification of exons 8–10 (Fig. 4A). All these products were analyzed by sequencing (Supp. Fig. S7). This indicated the presence of wild-type exon 8 splicing (product 17) and utilization of a novel splice donor site in exon 8 at c.1254, which is located 2 nt upstream of the c.1256A>T variant (product 18 in Fig. 4B and product 20 in Fig. 4C). The c.1256A>T variant generated a consensus GT dinucleotide splice donor site. This donor spliced from between nucleotides c.1254G and c.1255G to the canonical splicing acceptor site of exon 9 and the resulting reading frame was unchanged (Table 2). Splicing prediction programs indeed showed that c.1254 turned into a splice donor site due to the c.1256A>T variant (Supp. Fig. S2G). The canonical splice donor site of exon 8 remained unchanged, and it was unclear which of the two sites would be preferred from in silico predictions using Alamut. In fact, prediction of relative splice-site strengths using ASSEDA [Mucaki et al., 2013] showed that the strength of the novel cryptic splice donor site was negligible (Ri of 2.4) relative to the canonical splice donor site of exon 8 (Ri of 5.2) (data not shown), which was the opposite of the results found in vivo. Product 21 represented wild-type splicing of exon 10, whereas product 22 was the result of perfect exon 10 skipping in which the reading frame remained intact (Fig. 4D; Supp. Fig. S7). Loss of the exon 10 splice donor site by the c.1551+1G>T variant was consistent with splicing predictions (Supp. Fig. S2G), but the precise outcome remained unclear from in silico analysis.

### Evidence for low levels of leaky wild-type splicing

Along with the exon-internal qPCR analysis described below, the flanking exon PCR assay provides information on the severity of



**Figure 4.** Analysis of complex splicing changes in Pompe patient 8. **A:** Flanking exon PCR analysis. **B:** Cartoons of the genomic variant and the splicing variants from allele 1, detected from flanking exon PCR analysis of exon 8. **C:** Cartoons of the genomic variant and the splicing variants from allele 1, detected from flanking exon PCR analysis of exon 9. **D:** Cartoons of the genomic variant and the splicing variants from allele 2, detected from flanking exon PCR analysis of exon 10. **E:** Exon-internal qPCR analysis. Error bars indicate SD ( $n = 3$ ).

the variants via the relative intensities of the products. These can be explained based on the identification of the splicing products (Fig. 4B–D) and on the locations of the primers used for amplification (Supp. Fig. S8).

### Exon 7

Detection of exon 7 is performed with a forward primer that anneals to the 3' end of exon 6 and a reverse primer to the 5' end of exon 8 (Supp. Fig. S8). The 5' end of exon 8 is retained in all cases whereas the 3' part is spliced out in the c.1256A>T allele. Flanking exon PCR detection of exon 7 should therefore not be affected in this patient and this was indeed the case (Fig. 4A).

### Exon 8

Flanking exon PCR primers used for detection of exon 8 anneal to exon 7 and 9 (Supp. Fig. S8). Both exons are not affected in this patient predicting that all splicing alterations of exon 8 itself should be detected in a semiquantitative manner. Indeed, a strong wild-type product (number 17) was detected, dominated by allele 2, and a slightly weaker smaller product 18 was detected due to the novel cryptic splice donor site at c.1254 in allele 1. Assuming that maximally 50% of product 17 can be derived from allele 2, its

stronger abundance compared to product 18 therefore suggests that allele 1 has leaky wild-type splicing.

### Exon 9

PCR primers for detection of exon 9 by flanking exon PCR anneal to the 5' part of exon 8, which is the part that is not skipped in allele 1, and to exon 10, which is completely skipped in allele 2 (Supp. Fig. S8). This complicates detection of exon 9 from these two alleles: a product from allele 1 would be shorter than normal due to the partial skipping of exon 8. A product from allele 2 is not possible due to the precise skipping of exon 10, because this exon is required for primer annealing. The predominant product obtained was the shorter product number 20 that was derived from allele 1. However, a small amount of wild-type product number 19 was also observed. This indicates that at least one of the two alleles allows leaky wild-type splicing.

### Exon 10

Flanking exon PCR analysis of exon 10 is performed with primers annealing in exon 9 and exon 11, both of which are unaffected. The result therefore reflects the splicing alterations of exon 10 in a semiquantitative manner. Product 21 representing wild-type

splicing was the most abundant, whereas product 22 in which exon 10 was perfectly skipped was slightly less abundant. Because exon 10 splicing of allele 1 is unaffected and can account for 50% of wild-type product, this result suggests that allele 2 also has leaky wild-type splicing similar to allele 1.

### Quantification using exon-internal qPCR analysis

Quantification of mRNA expression of each exon revealed that all exons except exons 8 and 10 showed approximately twofold higher abundance compared with the healthy control. Exons 8 and 10 were expressed at twofold lower levels with respect to the other exons but still at 80%–120% of the levels of the healthy control. This indicates high mRNA expression in this patient. RNA from allele 1 (c.1256A>T) shows partial skipping of exon 8 resulting in failure of detection of a qPCR product. The detection of residual exon 8 is therefore derived from allele 2 (c.1551+1G>T), expected to contribute 50%, and the remaining expression is likely derived from leaky wild-type splicing from allele 1. The same rationale applies to detection of exon 10. In this case, expression was close to 50% relative to other exons, suggesting that the c.1551+1G>T variant allowed much lower levels of wild-type splicing. It should be noted that it is unclear why this patient shows twofold higher GAA expression relative to the healthy control, and whether this increase applies to both alleles to similar extents. This patient has a juvenile disease onset consistent with low levels of residual wild-type expression of GAA mRNA and of GAA enzymatic activity (Table 1).

In summary, patient 8 contained two splicing variants. c.1256A>T is a missense variant in exon 8 that causes p.D419V and in addition generates a novel splice donor site at c.1254, resulting in partial skipping of exon 8 and in leaky wild-type splicing. However, the leaky “wild type” product still contains the p.D419V variant that has been reported to lower the activity of the corresponding protein [Kroos et al., 2012a]. Therefore, c.1256A>T has a dual effect: it affects both splicing and enzyme activity of the remaining wild-type protein product. c.1551+1G>T is located in intron 10 and causes perfect skipping of exon 10 and in leaky wild-type splicing. The juvenile onset of Pompe disease suggests that both variants are moderately to severely pathogenic. This is consistent with the GAA enzyme activity levels, which are lower compared with adult onset patients.

## Discussion

A generic approach is described for the analyses of pathogenic variants at the level of RNA expression and splicing. It includes flanking exon PCR combined with sequencing to identify aberrant splicing products, and exon-internal qPCR to quantify exons at the RNA level. Although individual techniques are known and straightforward, these have so far not been combined and applied in a standardized manner to analyze the pathogenic nature of human gene variants. The approach has been developed for Pompe disease but can be adjusted for application to other monogenic disorders. Application of the approach to five partially or uncharacterized Pompe patients identified novel major splicing products in all cases, and enabled the quantification of leaky wild-type splicing, which correlated with disease onset and progression. Splicing prediction programs were largely unable to make accurate predictions. This approach is cheap and simple and should be suitable for implementation in the routine diagnosis of monogenic diseases including Pompe disease.

The approach provides three levels of information. First, aberrant splicing products can be detected in an unbiased and functional

manner. This expands the options for detecting splice site variants as current human variant analysis is based on prior knowledge and often involves sequence analysis of exons. Second, all splicing products can be sequenced. Sequencing of all products is recommended in a diagnostic setting to detect small changes in splice-site utilization that cannot be detected by agarose gel electrophoresis. This is illustrated by the detection of a splice-site switch of four nucleotides in patient five harboring the c.1075G>A variant. Another reason to sequence the products is to determine whether the reading frame is left intact or not, which predicts whether mRNA degradation via the NMD pathway may occur. Third, mRNA levels and the presence of each exon in the mRNA can be quantified by exon-internal qPCR. This provides quantitative information on mRNA expression and stability, and on the level of leaky wild-type splicing. Besides splicing variants, variants present in regulatory regions such as promoters and the UTRs should in principle also be detected using the exon-internal qPCR analysis.

Some limitations of this approach include the following. The first and the last exons cannot be analyzed by flanking exon PCR. A GC-rich sequence, which occurs frequently in the first exon, may prevent exon-internal qPCR analysis. This is the case for GAA exon 1. Solutions include: (1) other analyses of the approach that provide information on the missing exon. Variants in exon 1 that affect splicing may affect exon 2, which can be detected by flanking exon PCR. Effects on mRNA expression or stability should be detected using the exon-internal qPCR analysis of the remaining exons. (2) A variant-specific PCR strategy may be developed as shown in Supp. Figure S1D and E. Another limitation is that the approach is based on the analysis of RNA, which is not always available. On the other hand, at any moment a skin biopsy may be obtained, and the establishment of a fibroblast culture from such biopsy is a routine technique. The effect of a splicing variant may depend on the cell type and/or whether certain genes are expressed in a cell type-specific manner [Barbosa-Morais et al., 2012]. In the case of Pompe disease, expression of GAA is ubiquitous. However, it remains possible that splicing and/or mRNA stability differs between primary fibroblasts and the most affected cells, which include cardiac and skeletal muscle cells in the case of Pompe disease. Finally, certain aberrant splicing events may be missed by this approach. When one allele is fully expressed, it may dominate in the flanking exon PCR assay over expression of the second allele undergoing RNA degradation. Alternatively, a large aberrant splicing product may be formed that is outcompeted by smaller PCR products in the flanking exon PCR assay. False positive results can be minimized as shown by the –RT flanking exon PCR results, which failed to detect background PCR products (Supp. Fig. S3B). Taken together, whereas the false positive rate of the approach is likely very small, false negative results should be taken into consideration.

Alternative assays include exon arrays, deep RNA sequencing, or expression of minigenes containing part of the gene of interest. Exon arrays contain oligonucleotide probes to detect exons or exon junctions to which the labeled sample is hybridized [Wang and Cooper, 2007; Castle et al., 2008; Kwan et al., 2008]. This method is useful for initial screening but it cannot provide the same level of information obtained with the PCR-based methods described here. Sequence variants including SNPs that do not necessarily affect mRNA expression/processing may affect hybridization to the array. A quantitative analysis can therefore be difficult. This information is useful for estimating the extent of leaky wild-type splicing, which can have a predictive factor for disease severity. Deep RNA sequencing provides precise information on RNA splicing and processing [Wang et al., 2008; Lalonde et al., 2011; Lappalainen et al., 2013]. For a good sequence coverage of all exons, in-depth sequencing must be

performed, which increases the costs and can be a challenge for lowly expressed genes. Minigenes can be used when tissues of interest are difficult to obtain [Sharma et al., 2014]. Although this method can be very useful for the analysis of splicing variants, it is also time-consuming. For these reasons, the implementation of exon arrays, deep RNA sequencing, or minigenes in a standard diagnostic setting have not been widespread so far.

Previously characterized variants including the del ex18 and the IVS1 variant have been confirmed by the approach. The IVS1 variant yields three major products [Huie et al., 1994b; Boerkoel et al., 1995; Dardis et al., 2014], all of which were accurately detected. Alternative splicing forms are known to exist as well but these are expressed at low levels. It is interesting to note that the IVS1 variant has a rather poor genotype–phenotype relationship [Wokke et al., 1995]. This variant allows leaky wild-type splicing preventing the most severe classic infantile form of Pompe disease, but the onset of the disease can vary from early childhood to late adulthood. It is at present unknown what the underlying mechanism is for these differences. One could speculate that the level of leaky wild-type splicing in vivo in skeletal muscle cells varies between these patients, caused by variations in the expression or activity of splicing regulatory proteins. This will be interesting for future research.

Significant amount of novel information has been obtained using the approach (Table 2). (1) The c.525delT variant is a very severe variant resulting in undetectable GAA protein levels. It has therefore been classified as a CRIM (cross-reactive immunologic material) negative variant: because the patient does not express any GAA protein, a severe immunological response to intravenous treatment with recombinant enzyme can occur [Kishnani et al., 2010]. Surprisingly, all full size exons could be detected by flanking exon PCR (Supp. Fig. S3A), and exon-internal qPCR also detected low levels of mRNA expression for all exons. This indicates that the approach has high sensitivity and detected the low levels of mRNA that escaped NMD. The premature termination codon in the c.525delT alleles however prevents production of significant levels of (partial) protein from the remaining mRNA [Hermans et al., 1994]. (2) Three exonic variants have been found to induce pathogenic splicing events: c.1075G>A, c.1437G>A, and c.1256A>T. This highlights the need for including exonic variants in the approach described here to detect potential changes in mRNA expression/processing. (3) The c.-32–3C>G variant presents an interesting case. It results in the same splicing products as the IVS1 variant, with the exception that it does not allow detectable levels of leaky wild-type splicing. The mechanisms by which these two variants affect splicing are presumably different. The IVS1 variant is located in the polypyrimidine tract, and interferes with the in vitro binding of the general splicing factor U2AF65 [Lim et al., 2011; Dardis et al., 2014]. The c.-32–3C>G variant is located close to the splice acceptor site of exon 2, and is therefore expected to interfere with the binding of the U2snRNP complex to this site. These different mechanisms still don't explain why the IVS1 variant allows leaky wild-type splicing while the c.-32–3C>G variant does not. One might infer from this finding that variants that are located close to a canonical splice acceptor or donor site are more likely to inhibit leaky-wild type splicing compared to variants in more distant regulatory regions. Closer analysis of the data from this study suggests that this is not the case. Leaky wild type splicing has been detected in both types of variants, including IVS1 (c.-32–13T>G), c.1552–3C>G, c.1437G>A (splice junction site), c.1256A>T (in the middle of an exon), and c.1551+1G>T. The absence of detectable leaky wild-type splicing was the case for c.-32–3C>G, c.1551+1G>A, and c.1075G>A (splice junction site). Overall, these results show that it is not possible yet

to predict how splicing is affected and to what extent leaky wild type splicing is allowed, highlighting the need for functional validation.

Finally, these and other results show that splicing is a promiscuous process in which noisy splicing [Pickrell et al., 2010] and leaky wild-type splicing are often present. The implication is that splicing should be amenable to modulation. For example, small molecules targeting splicing factors [Fan et al., 2011; Webb et al., 2013] may present a worthwhile therapeutic strategy to treat genetic diseases caused by splicing variants.

## Acknowledgments

We would like to thank Dr. A Reuser for discussion and critical reading of the manuscript; Drs. F. Verheijen, H. Michaelakakis, M. van Hove, I. Baric, F. Eyskens, G. Gray, and P. Turnpenny for providing patient fibroblast cell cultures.

Ans T. van der Ploeg and W.W.M. Pim Pijnappel have received honoraria for educational presentations from Genzyme Corp., Cambridge, MA.

## References

- Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, Kim T, Misquitta-Ali CM, et al. 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338:1587–1593.
- Boerkoel CF, Exelbert R, Nicastrì C, Nichols RC, Miller FW, Plotz PH, Raben N. 1995. Leaky splicing mutation in the acid maltase gene is associated with delayed onset of glycogenosis type II. *Am J Hum Genet* 56:887–897.
- Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. 2013. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* 14:681–691.
- Castle JC, Zhang C, Shah JK, Kulkarni AV, Kalsotra A, Cooper TA, Johnson JM. 2008. Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat Genet* 40:1416–1425.
- Dardis A, Zanin I, Zampieri S, Stuanì C, Pianta A, Romanello M, Baralle FE, Bembì B, Buratti E. 2014. Functional characterization of the common c.-32–13T>G mutation of GAA gene: identification of potential therapeutic agents. *Nucleic Acids Res* 42:1291–1302.
- den Dunnen JT, Antonarakis SE. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 15:7–12.
- Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 37:e67.
- Fan L, Lagisetti C, Edwards CC, Webb TR, Potter PM. 2011. Sudemycins, novel small molecule analogues of FR901464, induce alternative gene splicing. *ACS Chem Biol* 6:582–589.
- Gungor D, Kruijshaar ME, Plug I, D'Agostino RB, Hagemans MLC, van Doorn PA, Reuser AJJ, van der Ploeg AT. 2013. Impact of enzyme replacement therapy on survival in adults with Pompe disease: results from a prospective international observational study. *Orphanet J Rare Dis* 8:49.
- Havens MA, Duelli DM, Hastings ML. 2013. Targeting RNA splicing for disease therapy. *Wiley Interdiscip Rev RNA* 4:247–266.
- Hermans MM, De Graaff E, Kroos MA, Mohkamsing S, Eussen BJ, Jooisse M, Willemsen R, Kleijer WJ, Oostra BA, Reuser AJ. 1994. The effect of a single base pair deletion (delta T525) and a C1634T missense mutation (pro545leu) on the expression of lysosomal alpha-glucosidase in patients with glycogen storage disease type II. *Hum Mol Genet* 3:2213–2218.
- Hermans MM, van Leenen D, Kroos MA, Beesley CE, Van Der Ploeg AT, Sakuraba H, Wevers R, Kleijer W, Michelakakis H, Kirk EP, Fletcher J, Bosshard N, et al. 2004. Twenty-two novel mutations in the lysosomal alpha-glucosidase gene (GAA) underscore the genotype–phenotype correlation in glycogen storage disease type II. *Hum Mutat* 23:47–56.
- Herzog A, Hartung R, Reuser AJ, Hermans P, Runz H, Karabul N, Gokce S, Pohlenz J, Kampmann C, Lampe C, Beck M, Mengel E. 2012. A cross-sectional single-centre study on the spectrum of Pompe disease, German patients: molecular analysis of the GAA gene, manifestation and genotype–phenotype correlations. *Orphanet J Rare Dis* 7:35.
- Huie ML, Chen AS, Brooks SS, Grix A, Hirschhorn R. 1994a. A de novo 13 nt deletion, a newly identified C647W missense mutation and a deletion of exon 18 in infantile onset glycogen storage disease type II (GSDII). *Hum Mol Genet* 3:1081–1087.
- Huie ML, Chen AS, Tsujino S, Shanske S, DiMauro S, Engel AG, Hirschhorn R. 1994b. 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* 3:2231–2236.
- Kishnani PS, Corzo D, Nicolino M, Byrne B, Mandel H, Hwu WL, Leslie N, Levine J, Spencer C, McDonald M, Li J, Dumontier J, et al. 2007. Recombinant human acid alpha-glucosidase—major clinical benefits in infantile-onset Pompe disease. *Neurology* 68:99–109.
- Kishnani PS, Goldenberg PC, DeArmy SL, Heller J, Benjamin D, Young S, Bali D, Smith SA, Li JS, Mandel H, Koeberl D, Rosenberg A, et al. 2010. Cross-reactive immunologic material status affects treatment outcomes in Pompe disease infants. *Mol Genet Metab* 99:26–33.
- Kishnani PS1, Nicolino M, Voit T, Rogers RC, Tsai AC, Waterson J, Herman GE, Amalfitano A, Thurberg BL, Richards S, Davison M, Corzo D, Chen YT. 2006. Chinese hamster ovary cell-derived recombinant human acid alpha-glucosidase in infantile-onset Pompe disease. *J Pediatr* 149(1):89–97.
- Kroos M, Hooegeven-Westerveld M, Michelakakis H, Pomponio R, Van der Ploeg A, Halley D, Reuser A, Consortium GAAD. 2012a. Update of the Pompe disease mutation database with 60 novel GAA sequence variants and additional studies on the functional effect of 34 previously reported variants. *Hum Mutat* 33:1161–1165.
- Kroos M, Hooegeven-Westerveld M, van der Ploeg A, Reuser AJ. 2012b. The genotype–phenotype correlation in Pompe disease. *Am J Med Genet C Semin Med Genet* 160C:59–68.
- Kroos M, Manta P, Mavridou I, Muntoni F, Halley D, Van der Helm R, Zaifeiriou D, Van der Ploeg A, Reuser A, Michelakakis H. 2006. Seven cases of Pompe disease from Greece. *J Inher Metab Dis* 29:556–563.
- Kroos MA, Pomponio RJ, Hagemans ML, Keulemans JL, Phipps M, DeRiso M, Palmer RE, Auems MG, Van der Beek NA, Van Diggelen OP, Halley DJ, Van der Ploeg AT, et al. 2007. Broad spectrum of Pompe disease in patients with the same c.-32–13T→G haplotype. *Neurology* 68:110–115.
- Kwan T, Benovoy D, Dias C, Gurd S, Provencher C, Beaulieu P, Hudson TJ, Sladek R, Majewski J. 2008. Genome-wide analysis of transcript isoform variation in humans. *Nat Genet* 40(2):225–231.
- Laforet P, Laloui K, Granger B, Hamroun D, Taouagh N, Hogrel JY, Orlikowski D, Bouhour F, Lacour A, Salort-Campana E, Penisson-Besnier I, Sacconi S, et al. 2013. The French Pompe registry. Baseline characteristics of a cohort of 126 patients with adult Pompe disease. *Rev Neurol (Paris)* 169:595–602.
- Lalonde E, Ha KC, Wang Z, Bemmo A, Kleinman CL, Kwan T, Pastinen T, Majewski J. 2011. RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. *Genome Res* 21:545–554.
- Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, Gonzalez-Porta M, Kurbatova N, Griebel T, Ferreira PG, Barann M, Wieland T, et al. 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501:506–511.
- Lim KH, Ferraris L, Filloux ME, Raphael BJ, Fairbrother WG. 2011. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc Natl Acad Sci USA* 108:11093–11098.
- Mucaki EJ, Shirley BC, Rogan PK. 2013. Prediction of mutant mRNA splice isoforms by information theory-based exon definition. *Hum Mutat* 34:557–565.
- Muller-Felber W, Horvath R, Gempel K, Podskarbi T, Shin Y, Pongratz D, Walter MC, Baethmann M, Schlotter-Weigel B, Lochmuller H, Schoser B. 2007. Late onset Pompe disease: clinical and neurophysiological spectrum of 38 patients including long-term follow-up in 18 patients. *Neuromuscul Disord* 17:698–706.
- Orlikowski D, Pellegrini N, Prigent H, Laforet P, Carlier R, Carlier P, Eymard B, Lofaso F, Annane D. 2011. Recombinant human acid alpha-glucosidase (rhGAA) in adult patients with severe respiratory failure due to Pompe disease. *Neuromuscul Disord* 21:477–482.
- Palacios IM. 2013. Nonsense-mediated mRNA decay: from mechanistic insights to impacts on human health. *Brief Funct Genom* 12:25–36.
- Perteau M, Lin X, Salzberg SL. 2001. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* 29:1185–1190.
- Pickrell JK, Pai AA, Gilad Y, Pritchard JK. 2010. Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet* 6:e1001236.
- Pittis MG, Donnarumma M, Montalvo AL, Dominissini S, Kroos M, Rosano C, Stroppiano M, Bianco MG, Donati MA, Parenti G, D'Amico A, Ciana G, et al. 2008. Molecular and functional characterization of eight novel GAA mutations in Italian infants with Pompe disease. *Hum Mutat* 29:E27–E36.
- Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. *J Comput Biol* 4:311–323.
- Regnery C, Kornblum C, Hanisch F, Vielhaber S, Strigl-Pill N, Grunert B, Muller-Felber W, Glocker FX, Spranger M, Deschauer M, Mengel E, Schoser B. 2012. 36 months observational clinical study of 38 adult Pompe disease patients under alglucosidase alfa enzyme replacement therapy. *J Inher Metab Dis* 35:837–845.
- Sharma N, Sosnay PR, Ramalho AS, Douville C, Franca A, Gottschalk LB, Park J, Lee M, Vecchio-Pagan B, Raraigh KS, Amaral MD, Karchin R, et al. 2014. Experimental assessment of splicing variants using expression minigenes and comparison with in silico predictions. *Hum Mutat* 35:1249–1259.
- Stroppiano M, Bonuccelli G, Corsolini F, Filocamo M. 2001. 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* 101:55–58.
- Toscano A, Schoser B. 2013. Enzyme replacement therapy in late-onset Pompe disease: a systematic literature review. *J Neurol* 260:951–959.
- Umapathysivam K, Hopwood JJ, Meikle PJ. 2005. Correlation of acid alpha-glucosidase and glycogen content in skin fibroblasts with age of onset in Pompe disease. *Clin Chim Acta* 361:191–198.
- Van den Hout H, Reuser AJ, Vulto AG, Loonen MC, Cromme-Dijkhuis A, Van der Ploeg AT. 2000. Recombinant human alpha-glucosidase from rabbit milk in Pompe patients. *Lancet* 356:397–398.
- van der Ploeg AT, Reuser AJ. 2008. Pompe's disease. *Lancet* 372:1342–1353.
- Wang GS, Cooper TA. 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 8:749–761.
- Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456:470–476.
- Webb TR, Joyner AS, Potter PM. 2013. The development and application of small molecule modulators of SF3b as therapeutic agents for cancer. *Drug Discov Today* 18:43–49.
- Wokke JH, Auems MG, van den Boogaard MJ, Ippel EF, van Diggelene O, Kroos MA, Boer M, Jennekens FG, Reuser AJ, Ploos van Amstel HK. 1995. Genotype–phenotype correlation in adult-onset acid maltase deficiency. *Ann Neurol* 38:450–454.
- Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 11:377–394.
- Zampieri S, Buratti E, Dominissini S, Montalvo AL, Pittis MG, Bembi B, Dardis A. 2011. Splicing mutations in glycogen-storage disease type II: evaluation of the full spectrum of mutations and their relation to patients' phenotypes. *Eur J Hum Genet* 19:422–431.