Pre-mRNA splicing in Pompe disease: Prospects for antisense therapy

Atze Jacobus Bergsma

The research described in this thesis was supported by the Sophia Children's Hospital Foundation [grant number S-687].

The studies presented in this thesis were performed at the departments of Pediatrics and Clinical Genetics of the Erasmus University Medical Center, Rotterdam, The Netherlands

ISBN: 978-94-6332-064-1

Author: Atze Bergsma
Cover design: Atze Bergsma
Layout: Pieter Bergsma and Atze Bergsma
Printed by: GVO drukkers & vormgevers B.V.

Pre-mRNA Splicing in Pompe Disease: Prospects for antisense therapy

Pre-mRNA splicing in de ziekte van Pompe: Vooruitzichten voor antisense therapie

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof.dr. H.A.P. Pols
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
dinsdag 20 september 2016 om 15.30 uur

door

Atze Jacobus Bergsma

Erasmus University Rotterdam

PROMOTIECOMMISSIE

Promotor	Prof.dr. A.T. van der Ploeg
Overige leden	Prof.dr.ir. S.M. van der Maarel
	Prof.dr. E.H.H.M. Rings
	Prof.dr. R.M.W. Hofstra
Copromotor	Dr. W.W.M. Pijnappel

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

2'O-mePS	2'O-methylphosphorotioate
adRP	autosomal dominant retinitis pigmentosa
AON	antisense oligonucleotide
ASSEDA	automated splice site and exon definition analysis
BMD	Becker muscular dystrophy
CNS	central nervous system
CypA	cyclophilin A
DMD	Duchenne muscular dystrophy
ERT	enzyme replacement therapy
ESE	exonic splice enhancer
ESS	exonic splice silencer
FDA	American food and drug association
GAA	acid α-glucosidase
GS	genesplicer
HEK cell	human embryonic kidney cell
hES cell	human embryonic stem cell
HGMD	human gene mutation database
HGVS	human genome variation society
hnRNP	Heterogeneous ribonucleoprotein particle
HSF	human splicing finder
iPS cell	induced pluripotent stem cell
ISE	intronic splice enhancer
ISS	intronic splice silencer
MEF cell	mouse embryonic feeder cell
MES	maxentscan
MHC	myosin heavy chain
MOE	2'O-methoxyethyl
NGS	next generation sequencing
NMD	nonsense-mediated decay
NNS	neural network splice
PAS	periodic acid Shiff
PMO	Morpholino
PNA	peptide nucleic acid
polyY	polypyrimidine tract
Pre-mRNA	messenger RNA precursor
PRPF	pre-mRNA processing factor
rhGAA	recombinant human GAA
RT-PCR	reverse transcriptase polymerase chain reaction
SF	splice factor
SMA	spinal muscular atrophy
SMaRT	spliceosome-mediated RNA trans-splicing
SMRTseq	single molecule real-time DNA sequencing
SNP	single nucleotide polymorphism

snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein
SR	serine/argenine rich
SROOGLE	splicing regulation online graphical engine
SSF	splicesitefinder-like
SV	splice variant
UTR	untranslated region

Chapter 1

General Introduction, Aims and Scope

INTRODUCTION

Genetic variation between humans is caused by the occurrence of sporadic variants in the DNA of individuals. Whenever these changes occur in the germ-line, they are inherited by the next generation. These processes are important for evolutionary changes and sustained variability within species. However, these changes can also lead to disease. To date, approximately 180,000 germline disease-related lesions have been described in the Human Gene Mutation Database, which occur in more than 30% of human protein-coding genes and are associated with human disease (http://www.hgmd.cf.ac.uk). With the advent of new next generation sequencing (NGS) techniques, many more variants and genes that are linked to disease are likely to be added to this database, broadening our understanding of genotype/phenotype relations. However, there is an enormous pace at which new potentially pathogenic variants are discovered, and functional consequences of these variants are difficult to investigate. Understanding the basic mechanism underlying disease caused by pathogenic variants is critical for development of therapies. This introduction highlights the current practice in genetic diagnostics in regard to functional investigation of variants, with a focus on variants that affect splicing. Furthermore, therapeutic strategies that modulate splicing are exemplified, highlighting the potential of antisense oligonucleotides in the treatment of Duchenne muscular dystrophy and spinal muscular atrophy. Current practice in genetic diagnostics and therapeutics of one particular monogenic disorder, Pompe disease, will be used as an example.

DIAGNOSIS OF MONOGENIC INHERITED DISORDERS

Inherited monogenic disorders are often caused by ectopic protein expression and/or protein activity of one or more proteins. The underlying cause of these changes can be traced back to the gene coding for that particular protein. The methods for identification of disease etiology have dramatically improved over the last century. A paradigm example is the discovery of the biological and genetic cause of Pompe disease. Pompe disease was first phenotypically described by the Dutch physician Johannes Cassianus Pompe in 1932 (1). Later, the cause was identified as a loss of activity and/or expression of the acid α–glucosidase (GAA) protein (2). The underlying genetic origin was unraveled after identification and cloning of the *GAA* gene, which encodes the GAA protein (3, 4). Thus far, more than 400 pathogenic variants have been described in the *GAA* gene (www.pompecenter.nl). Many other genes have been linked with various diseases through similar steps due to the direct functional link between the affected protein and the patient phenotype. In the absence of such link, other techniques are used for gene discovery, including linkage analysis and homozygosity mapping. With the advent of sequencing techniques based on Sanger sequencing (5), and later massive parallel sequencing, sequence variations underlying genetic disease can be readily identified for many disorders.

Sequence variations associated with disease can exert their pathogenic effect in various ways. Single nucleotide polymorphisms (SNPs), in which the nucleotide at one specific site is exchanged for another, account for more than half of the variants presented in the HGM database (http://www.hgmd.cf.ac.uk). Other variants are deletions and/or insertions of any size. The effect of a variant on protein expression and activity depends on the nature of the variant. For instance, missense variants, which cause amino acid changes, often have a less deleterious effect than nonsense variants, which generate premature stop codons. Furthermore, a variant can also effect splicing of messenger RNA precursor (pre-mRNA). In some cases, more than one effect can be exerted by a sequence variant, e.g. a missense variant can also effect splicing. This makes it difficult to predict what the effect of a variant on protein expression/activity is.

SPLICING AND HUMAN DISEASE

Of all variants that are annotated in the Human gene variation database, 16.356 entries are variants which are believed to alter canonical splicing (HGMD, February 2016) (Box 1). These account for 9.1% of all disease-associated variants in the database. Although splice switching variants seem to account for only a small portion of all known variants, caution should be taken. The data entered in HGMD is biased due to the methods used for finding pathogenic variants. In the early days of sequence analysis, functional analysis of reverse transcribed mRNAs using PCR techniques were often employed for finding variants that alter the length of mRNA transcripts. The first dataset of identified pathogenic splice variants presented 101 variants affecting splicing, accounting for more than 15% of all variants identified up to that point (6). However, functional genetic screening using RNA as a template was impractical to test for pathogenic variants. Furthermore, it relies on levels of expression in cell lines analyzed. Due to the easy standardization of genomic DNA isolation, Sanger sequencing of genomic DNA became the gold standard in genetic diagnostics. A drawback of this approach is that the functional effect on splicing of any type of variant is easily missed. Due to the high number of common SNPs (minor allele frequency of >1%) present in intronic regions, exonic sequences are prioritized for detection of missense and nonsense variants, leading to an overrepresentation of these variants in current mutation databases. Indeed, a recent survey shows that approximately 25% of missense and nonsense variants described in the HGM database may have an effect on splicing (7), suggesting that one third of all pathogenic SNPs in the database potentially affect splicing. Furthermore, a functional investigation of all knowns variants present in a single exon in *BRCA2* and *MLH1* yielded a high percentage of variants that affect splicing (42% and 77%, respectively) (8, 9). One mathematical model even suggests that up to 62% of all disease causing variants affect splicing (10). These data indicate that the effect of pathogenic variants on splicing of pre-mRNA may be considerably underestimated.

Variants that lead to abnormal splicing can exert their effect in trans or in cis.

Trans-acting variants affect proteins and/or Ribonucleoproteins (RNPs) that are part of the splice-regulatory mechanism. These factors regulate splicing of many genes and thus tend to have a more global effect. An example of a disease in which these trans-

BOX 1: PRE-MRNA SPLICING IN HUMANS VIA THE SPLICEOSOME

Current estimates state that the human genome contains 20,687 protein-coding genes (99). However, the number of proteins is estimated to be much greater. Processes responsible for this protein diversity include alternative promoter usage, alternative transcription termination, alternative splicing, and post translational modification. Transcripts of >90% of protein-coding genes produce multiple splice-isoforms. Splicing constitutes the removal of non-coding intronic sequences and ligation of coding exonic sequences to produce mature mRNA species, which encode for specific proteins. The differential inclusion of exons results in multiple mRNA isoforms and ultimately in protein diversity. This is a highly regulated process orchestrated by the spliceosome, a highly dynamic RNA/protein complex, and by additional splicing factors. There are two known spliceosome complexes, the major and the minor spliceosomes. The major spliceosome is responsible for the splicing of 95.5% of all introns and is the main topic of this box. The function of the minor spliceosome is reviewed in (100). The major spliceosome consists of the small nuclear ribonucleoproteins (snRNPs) U1, U2, U4, U5 and U6, which assemble in a specific order. Conserved splice elements are present in native RNA including the 5' and 3' splice sites, the branchpoint site, the polypyrimidine tract (PolyY), and exonic or intronic splice enhancers (ESEs, ISEs) or silencers (ESSs, ISSs), which can be recognized by the spliceosome and other splice regulatory factors (figure 1A). U1 recognizes the splice 5' splice site, while U2 recognizes the branchpoint and 3' splice site. These are then brought together by the action of the U4, U5 and U6 snRNPs (Figure 1B). Two transesterification reactions follow in which the intron lariat is cut and exons are ligated together (reviewed by (101)). The activity of the spliceosome is further regulated by transacting factors that (de dan en als van het Nederlands) bind to splice regulatory elements within the RNA molecule. These regulatory elements are recognized by a plethora of splicing factors, which can either promote or suppress exon recognition. The combination of positive and negative signals determines whether splicing can take place.

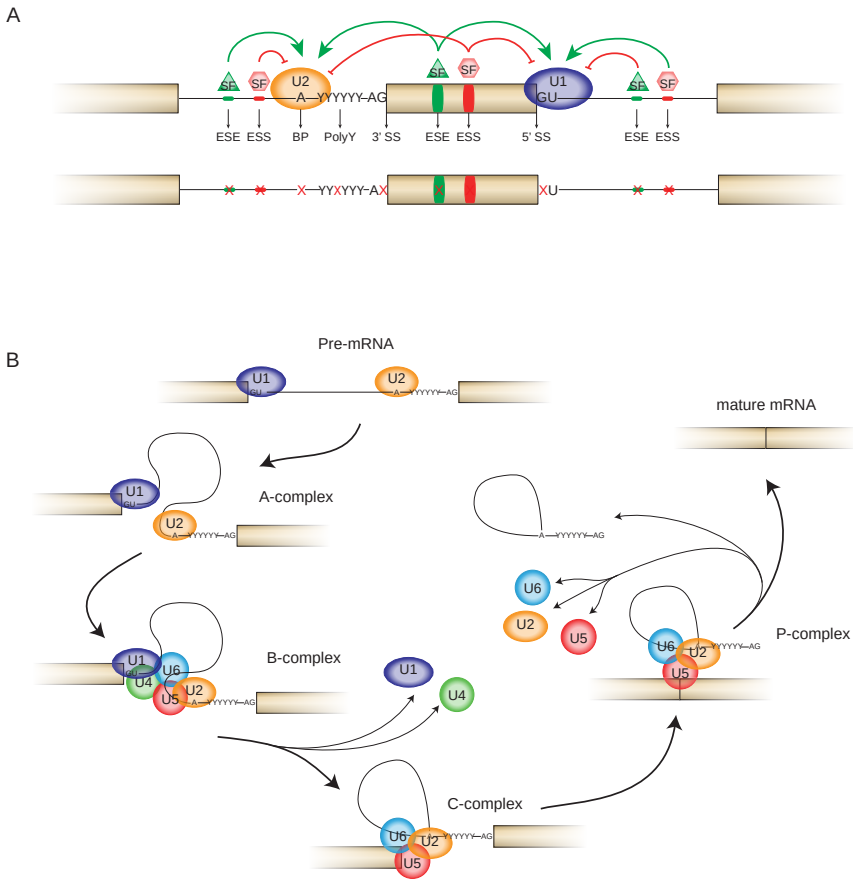


Figure 1. Regulation of splicing. (a) cartoon depicting elements within the pre-mRNA. *These elements can be well conserved sequences (3' and 5' splice site and the polyY tract), or more loosely defined sequences (exonic or intronic splice enhancers (ESEs, ISEs) or silencers (ESSs, ISSs)).These elements can be bound by splicing factors (SF), which in turn have a positive or negative influence on recognition of the exon they regulate. Binding sites of two major splicing components, U1 and U2 are indicated. Theoretically, any variant residing within the pre-mRNA sequence of a gene can cause changes in splicing outcome, as is shown in the lower cartoon with red crosses. Beige boxes are exonic regions, lines represent intronic regions. A represents the branch point site. (b) The subsequent steps of splicing out an intronic region from pre-mRNA leading to generation mature mRNA. The assembly and disassembly of the most important components of the major spliceosomal complex are indicated.*

effects are common is autosomal dominant retinitis pigmentosa (adRP), a condition in which degeneration of the retina leads to blindness in affected patients. Various variants in a number of pre-mRNA-processing factors (PRPFs) have been linked to the disorder (11). Although these PRPFs are expressed ubiquitously, expression in retinal tissue is significantly upregulated due to the relatively high level of constitutive mRNAs in this tissue, rendering these cells sensitive for variants in these proteins (12).

Cis-acting variants reside in pre-mRNA transcripts and alter splice regulatory elements, which shifts the canonical splicing equilibrium. These variants can lead to loss of recognition of existing elements by splicing factors, or a new splice regulatory element can

be generated. The most important and well conserved splice elements for correct splicing are the two intronic dinucleotides at the 5' and 3' borders of splice sites, the sequence of which is a GT or AG dinucleotide, respectively. More than 99% of exons have these consensus splice site sequences (13). Indeed, of all splicing variants in the HGM database 77% of variants located at the splice acceptor affect the AG dinucleotide and 64% affect the GT dinucleotide at the splice donor site (14). Pathogenic variants in one of these four bases are common. An example can be found in the *CFTR* gene. According to the Cystic Fibrosis Mutation Database 93 pathogenic variants lay within these two dinucleotide sequences (15). These variants lead to loss of exon recognition by spliceosomal factors and cause aberrant splicing via exon skipping, or lead to the usage of alternative splice sites (Figure 1B). The polypyrimidine tract is another well-defined splicing element, comprising a pyrimidine nucleotide rich stretch of approximately 15 to 20 bases, and is located 5 to 40 bases before the splice acceptor site. A well described polypyrimidine tract mutation has been identified in Pompe disease patients. The c.-32-13T>G variant in the *GAA* gene results in (partial) skipping of exon 2, resulting in a reduction of normal transcripts from this allele (16-19). Most variants that are known to influence splicing reside in these well-defined regions. However, variants that affect splicing do not necessarily have to be in close proximity to these elements. Variants that reside in less conserved splice elements like exonic or intronic splice enhancers and silencers (ESE, ISE, ESS, ISS) can be found at some distance from the splice sites of the exon that they regulate. An example can be found in variants identified to affect splicing of *NF1* exon 9, causing Neurofibromatosis type 1, in which several variants were found that alter splicing through mutation of exonic splice regulatory elements (20). Furthermore, variants can potentially generate a new splice site irrespective of its distance to the canonical splice site, which could affect splicing of the canonical exons (21, 22). Variants that affect less defined splicing elements are often misinterpreted due to the lack of diagnostic functional assays testing splicing. Deep intronic variants are missed completely, because diagnostic sequencing in monogenic disorders is limited to the sequencing of coding regions from genomic DNA of a gene. RNA-based assays would resolve these issues and seem to be a promising part of the toolkit in future diagnostics (23).

POTENTIAL THERAPEUTIC STRATEGIES FOR SPLICING INTERVENTION

Splicing plays an important role in transcription regulation and a great number of factors is involved in the splicing process. This makes splicing an attractive target for therapeutic intervention. The process of splicing can be manipulated in multiple ways with the use of therapeutic agents. This form of interference can be used to either up- or downregulate expression of targeted mRNA transcripts. The small molecule Spliceostatin A has been tested for its ability to globally inhibit splicing by targeting the splicing factor 3B (SF3B)

subunit of the *U2* snRNP, leading to cell cycle arrest in G1 and G2/M phase (24). These findings could have implications for the treatment of multiple types of cancer. A compound with more specific splice switching capacity was discovered in a drug screen to identify small compounds that can improve *SMN2* exon 7 inclusion for the potential treatment of spinal muscular atrophy (discussed in more detail below). The screen identified Aclarubicin as a compound improving exon 7 inclusion (25). These findings suggest the potential of using small molecules for targeting the splicing machinery. However, a drawback of the use small molecules is that they have a global effect on splicing, and thus the chance of off target effects is substantial. Therefore, techniques that target aberrant splicing events caused by specific variants are preferred. Another option for correcting aberrant splicing caused by the presence of splice altering variants is spliceosome-mediated RNA trans-splicing (SMaRT). This technique is based on the property of the spliceosome to enable the combination of transcripts from two independent RNA-molecules to form one mRNA transcript. SMaRT was proven to be effective after integration of introduced exogenous pre-mRNA products into the endogenous mRNA resulting in the generation of corrected full length *CFTR* mRNA transcript in a proof-of-concept study (26). Although SMaRT has been proven to be a feasible strategy for splice correction *in vitro* and in mouse models for multiple disorders, including Huntington's disease (27), inherited tauopathy (28), and p53-defective cancers (29), the process of trans-splicing rarely occurs in humans (30). Another specific and promising way to influence splicing can be accomplished with antisense oligonucleotides (AONs). These are oligonucleotides consisting of 15 to 30 RNA and/or DNA bases which can target and block splicing elements in pre-mRNA sequences. To prevent the AONs from being degraded by RNAses and DNAses, they have a modified backbone. The first AONs used carried a 2'O-methylphosphorotioate (2'O-mePS) backbone and were targeted against cryptic splice sites in mRNAs transcribed from mutated B-globin gene (31, 32). Currently, the most frequently used backbone chemistries are 2'O-mePS, 2'O-methoxyethyl (MOE), Morpholino (PMO) and peptide nucleic acids (PNA) (33). Recently, a promising new type of backbone chemistry, tricyclo-DNA, was also tested *in vivo* (34). Further modifications can be made to improve uptake, like cell penetrating peptides (reviewed in (35)) and nanoparticles (reviewed in (36)), and efficacy. For many disorders AON mediated therapy is currently being developed, including Hutchinson-Gilford progeria syndrome (37), type I Usher syndrome (38), Cystic Fibrosis (39) and cancer (reviewed in (40)). Particular progress has been made in the development of antisense therapeutics for Duchenne Muscular Dystrophy (DMD). DMD is a disease that results in severe muscle wasting, leading to loss of ambulation, cardiomyopathy and/or respiratory failure to which patients succumb before the age of 30 (41). The disease is caused by the impairment of expression of functional dystrophin protein, which is normally expressed in the muscle and acts as a bridge between the sarcolemma and the actin filaments driving muscle contraction. Dystrophin is transcribed from the *dystrophin* (*DMD*) gene, which is

located on the X chromosome and is one of the largest genes in the human genome with a genomic length of 2.5 megabases and 79 exons. The genetic cause of DMD can be found in partial deletions in the gene (65% of variants), point variants (25%) and insertions (6 to 10%) (42). Almost all pathogenic variants linked to DMD disease cause the generation of a premature termination codon due to either a reading frame shift or the presence of a nonsense variant. This leads to nonsense-mediated decay of the mRNAs produced, resulting in abrogated dystrophin expression. However, a milder phenotype of the disease, known as Becker Muscular Dystrophy (BMD) is caused by deletions that leave the reading frame intact (43). This causes shortening of the protein, but due to its modular structure it is still partially functional. This led researchers to believe that restoration of the reading frame could partially restore dystrophin function in DMD patients, which results in the milder BMD phenotype. To this end, antisense oligonucleotides were designed to block splicing of certain exons in the dystrophin gene in order to restore the reading frame (example shown in Figure 2A) (44-46). Furthermore, generation of 15 AONs targeting multiple *DMD* exons enabled the potential restoration of the reading frame for more >50% of DMD patients

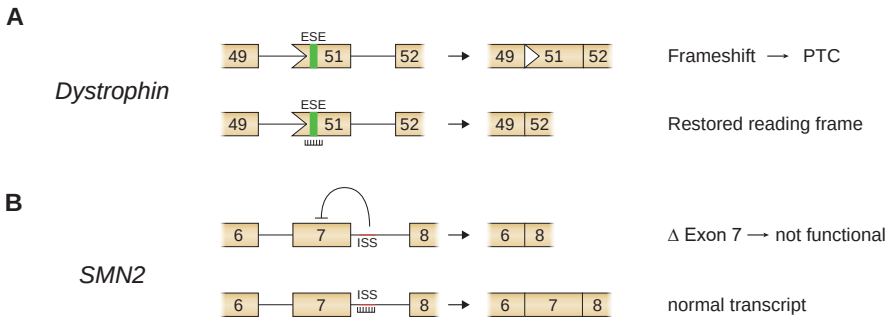


Figure 2. The use of antisense oligonucleotides in two well studied diseases. *(a) Schematic representation of a representative allele in a patient with Duchenne Muscular Dystrophy. This patient has a deletion encompassing exon 50 in the Dystrophin gene. This deletion variant leads to skipping of exon 50 in the Dystrophin mRNA transcript, leading to a frame shift (represented by the missing triangle). Restoration of the reading frame is achieved via AON mediated skipping of exon 51 by blockage of an exonic splice enhancer (ESE). (b) Spinal Muscular Atrophy is caused by variants in the SMN1 gene, leading to dysfunctional SMN protein. The pseudogene SMN2 contains a specific synonymous variant in exon 7 which leads to skipping of this exon in 90% of transcripts. Blockage of an intronic splice silencer (ISS) in intron 7 rescues this missplicing, leading to normal transcripts and functional SMN protein production.*

(47). The efficacy of AONs *in vivo* was confirmed with partial restoration of dystrophin expression in a mdx mouse model, both by local (48) and systemic administration (49, 50). Further investigation led to the start of multiple clinical trials, both with the 2'OmePS backbone (under the name of Drisapersen, by Prosensa) (51, 52) and the PMO backbone (under the name of Eteplirsen, Sarepta) (53, 54). Both AONs showed promising results in Phasel/II trials. However, the Phase III trial for Drisapersen failed to meet the primary endpoint. Although disappointing, much could be learned from these trials including better definition of primary outcomes and overall setup of such trials (55). In addition, these trails

were all started some time ago. Much progress has been made in recent years in the field of AON research, improving both efficacy and delivery of these antisense compounds.

Another disease in which AONs show a promising therapeutic approach is Spinal Muscular Atrophy (SMA). After cystic fibrosis, SMA is the most common autosomal recessive inherited disorder with a prevalence of a more than 1 in 10.000 (56). The disease is caused by loss of motorneurons in the anterior horn of the spinal cord, which in severe cases leads to death before the second year of life due to respiratory failure (57). Currently, no curative treatment is available for the disease and care is restricted to supportive and symptomatic attention. The genetic cause of SMA can be found in homozygous deletions or variants in the *SMN1* gene (58). Some residual protein expression often remains due to a large inverted duplication on Chromosome 5q, which encodes for one or more *SMN2* gene copies. Only a small amount of full length *SMN2* mRNA expression is produced from this gene due to a synonymous point mutation in *SMN2* exon 7, leading to the conversion of an exonic splice enhancer into an exonic splice silencer (59). In the background of weakly conserved canonical exon 7 3' and 5' splice sites, polypyrimidine tract and branch point sequences, this change from a ESE to an ESS results in extensive skipping of *SMN2* exon 7 (*SMN2Δ7*) in mRNA transcripts from this gene, although 10% of correct splicing remains (60, 61). Severity of the disease is scaled from type 0 as most severe, which is fatal *in utero* or short after birth, to type IV with only mild symptoms but normal life expectancy. Disease severity correlates with the amount of *SMN2* gene copies present (62). Furthermore, disease severity can be linked to the presence of variants that affect the splicing equilibrium of *SMN2* exon 7 (63). Since the discovery of the aberrant splicing present in transcripts from the *SMN2* gene, researchers have hypothesized that rescue of aberrant *SMN2* pre-mRNA splicing could lead to disease rescue. Currently, research focusses on the AON mediated correction of *SMN2* exon 7 exclusion. Several studies have investigated potential pre-mRNA targets for restoration of canonical *SMN2* splicing (64, 65). The optimal target is an intronic splicing silencer element present from base 10 to 24 in intron 7, termed ISS-N1 (Figure 2B) (66). Blockage of ISS-N1 site with 2'O-mePS AONs resulted in upregulation of *SMN2* canonical pre-mRNA splicing. The efficacy of the AON was further demonstrated *in vivo* in *SMN1* knockout mice carrying two copies of human *SMN2* (67) and in non-human primates (68), justifying further investigation of the AON in clinical trials. Phase I/II clincical trials are currently ongoing in which Ionis-SMNRx (or Nusinersen), a 2'-O-methoxyethyl modified AON developed by Ionis Pharmaceuticals/ Biogen which targets ISS-N1, is delivered intrathecally in SMA type I, II and III patients . First results indicate good safety and tolerability, which warrants further investigation (69).

POTENTIAL OF AON MEDIATED THERAPY IN POMPE DISEASE

As mentioned above, Pompe disease is an autosomal recessive inherited monogenic disorder caused by variants present in the *GAA* gene, which encodes for the GAA protein.

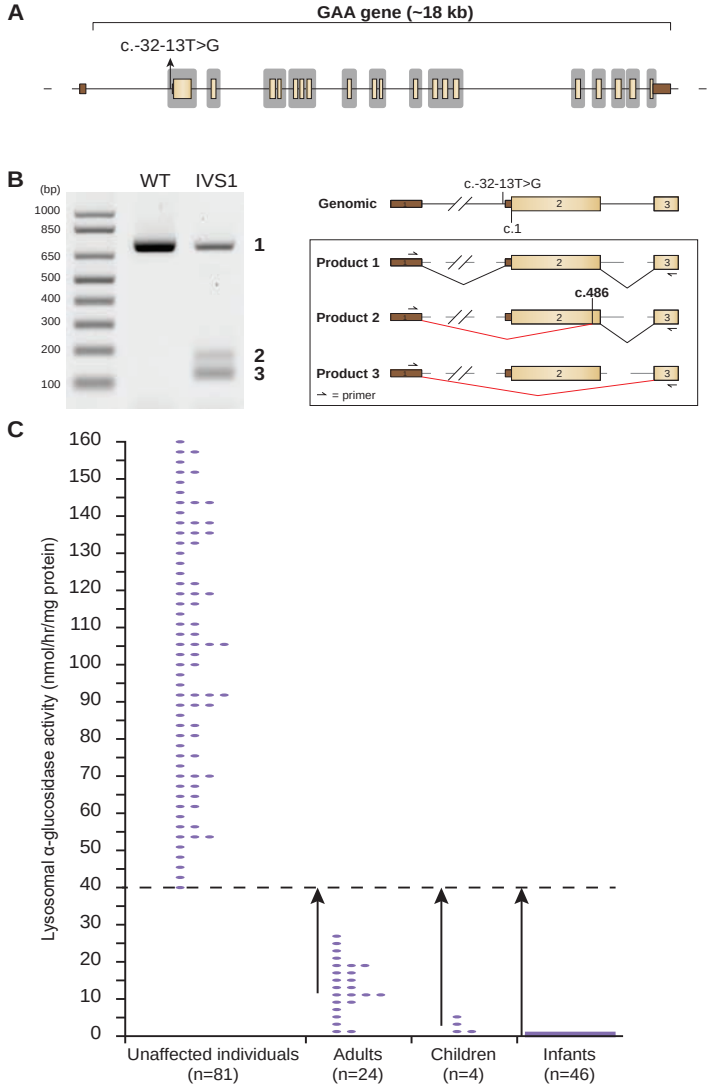


Figure 3. Leaky wild-type splicing due to the IVS1 variant in Pompe disease*. (a) Schematic representation of the GAA gene. Brown boxes are 3' and 5' UTRs, beige boxes are exonic regions, black lines are intronic regions. Grey boxes indicate the genomic regions screened in diagnostics to identify variants. The location of the IVS1 variant is indicated. (b) Representative RT-PCR gel from a patient carrying the IVS1 variant. The cartoon next to the gel depicts the specific spliced products indicated in the gel. (c) Graph depicting the measurement of GAA enzymatic activity in 155 individuals. Healthy controls do not have a lower activity than 40 nmol/hr/mg protein with an average of 123 nmol/hr/mg protein. Juvenile and adult Pompe patients do not have a higher GAA enzymatic activity then 27 nmol/hr/mg protein with an average of 15 nmol/hr/mg protein.*

The gene constitutes an 18 kb stretch on the long arm of chromosome 17 (Figure 3A). Currently, more than 400 (potentially) pathogenic variants have been described in the Pompe mutation database (http://www.pompecenter.nl) of which 54% are missense, 22% are deletions or duplications, 15% affect splicing, and 9% are nonsense. The presence of these

variants leads to partial or complete loss of GAA enzymatic activity leading to accumulation of glycogen in the lysosome, and particularly affects muscle tissue. It is a rare disease with an estimated incidence of 1 in 40.000 newborns (70, 71). A clinical spectrum of severity is observed in patients with Pompe disease (70, 72). The very severe early onset (or classic infantile) form, which accounts for approximately 10% of all patients, is characterized by symptoms of generalized muscle weakness, and hypertrophic cardiomyopathy and often results in death within the first year of life through cardiac or respiratory insufficiency, if left untreated (73, 74). The childhood/adult onset (or non-classic) form of Pompe disease has a wide spectrum of onset ranging from infancy to late in adulthood. Non-classic Pompe patients have symptoms ranging from hypotonia, limb-girdle weakness, and involvement of the diaphragm resulting in fatigue, wheelchair- and ventilator dependency, but they lack hypertrophic cardiomyopathy (75, 76). Pompe disease is generally diagnosed by measuring residual GAA enzymatic activity from dried blood spots (77), directly from muscle biopsies (78), in leukocytes (79), and cultured primary skin fibroblasts (80). For further confirmation, genomic DNA sequencing of the coding region of the *GAA* gene can be performed for identification of pathogenic variants (figure 3a, grey area's), after which the pathogenicity of the variants can be determined by cDNA expression analysis for missense variants or via *in silico* prediction (Box 2). Recently, newborn diagnostic screening for Pompe disease, which is based on the dried blood spot assay, has become standard practice in a number of countries, including Italy, Japan and Taiwan (81). Research is currently ongoing for the enrolment of Pompe disease in panels for prenatal screening (82). Another frequently used diagnostic assay is pathological staining of muscle biopsies for acid phosphatase and Periodic Acid Schiff (PAS) staining. Other markers have been investigated for their potential utilization in Pompe disease diagnostics, including identification of lysosomal glycogen accumulation in muscle biopsies using electron microscopy (83) and PAS staining of monocytes (84).

Soon after the discovery that GAA deficiency causes Pompe disease, the idea was raised to treat the disease by replacement of the enzyme with functional GAA protein (85, 86). However, these first trials did not result in clinical benefit. New insights revealed that this was caused by ineffective targeting of the enzyme in muscle tissue due to suboptimal glycosylation of GAA, which hampered receptor mediated uptake. The advent of new techniques that allow large-scale production of recombinant human GAA (rhGAA) led to production of the enzyme in Chinese hamster ovary cells, which in turn resulted in better glycosylation leading to improved uptake of rhGAA via the mannose-6-phosphate receptor (87). In 1999 the original hypothesis was proven by rescue of the phenotype in Pompe mice after administration of rhGAA isolated from transgenic rabbit milk (88). Further development led to the first clinical trial for enzyme replacement therapy (ERT) rhGAA in early onset Pompe patients, which showed improved survival for all four patients tested without experiencing adverse effects , warranting further study (89). Clear improvement

BOX 2: USE OF PREDICTION ALGORITHMS FOR PREDICTION OF THE EFFECT OF A VARIANT.

Functional assays for the detection of variants affecting pre-mRNA splicing are currently not widely used in diagnostics of monogenic disorders. This is mainly caused by the difficulty to implement RNA based assays in the diagnostic workflow and the limited availability of specific tissue samples. Another easier way to test the impact of a variant on splicing is by using *in silico* splice prediction algorithms. Algorithms for the prediction of branchpoint, 3' and 5' splice sites are often based on specific characteristics seen in these well characterized splice-elements, e.g. sequence conservation (102, 103), maximum entropy distribution (104) and machine learning techniques (105, 106). Various splice prediction algorithms for detection of these conserved splice elements have been extensively reviewed in (107). A number of algorithms are currently available for the prediction of splice enhancer and splice silencer elements. Prediction of ISEs, ISSs, ESEs and ESSs are much more difficult due to their low sequence conservation, and the limited availability of functional data on these sites. A much used tool for the prediction of enhancer-elements is the ESEfinder algorithm (108). The algorithm is based on results from functional assays in which sequence evolution was performed to find optimal motifs for promotion of exon inclusion by specific factors (109, 110). Another algorithm that is often used for the prediction of exonic splice enhancers is Rescue ESE, which is a method based on motif comparison (111). Methods for identification of potential splice silencer sequences are also available. Sironi's motif can be implemented for prediction of both intronic and exon splice silencer elements, and are purely based on motif recognition (112). The algorithm was designed based on the differences in motifs present in unspliced pseudo-exons in comparison to normally spliced exons to find sequences that have a negative effect on exon inclusion. This strategy was also used for the generation of an algorithm that can discriminate between silencer and enhancer elements (113). With the presence of a plethora of splice prediction algorithms, it is difficult for investigators to choose which one should be used. Different algorithms can have contradicting outcomes. For user convenience, several platforms have integrated multiple splice prediction tools to provide a better view on the effect of splicing. Two online platforms, Human Splicing Finder (HSF, (114)) and the Splicing RegulatiOn Online GraphicaL Engine (SROOGLE (115)), integrate multiple tools to

analyze sequences for the presence of splicing elements, albeit in different ways. HSF shows the outcome of specific algorithms separately, but within one screen as an overview. However, in SROOGLE outcomes of different algorithms are combined and give a significance score, which is then visualized along the input sequence. Various sequence analysis software packages are now also equipped with integrated splice prediction tools, which all have their own approach of how to visualize prediction outcomes, further facilitating integration of splice-prediction into the clinical diagnostics toolkit. Although splice prediction remains an indirect tool for evaluating the effect that sequence variations have on splicing, and verification of predicted effects need to be functionally validated, prediction will continue to be a valuable part in genetic diagnostics. Furthermore, continued optimization of prediction algorithms will ensure more accurate prediction capabilities, which will help in improving our understanding of splicing in general.

was seen in a phase III trial for ERT in early onset Pompe disease (90). Finally ERT with rhGAA (Myozyme®, Genzyme Corporation, a Sanofi company, Cambridge MA, USA) for the treatment of Pompe disease was approved by the FDA in 2006. Although outcomes vary for classic infantile Pompe patients, mostly due to adverse immune responses against rhGAA, a clear benefit is seen for this group of patients (91). A systematic review of clinical trials with Myozyme in juvenile and adult onset Pompe patients indicates that ERT is well tolerated and at least partially effective, with improvement or stabilization of ventilatory and/or ambulatory support in two-thirds of patients (92). However, caution needs to be taken as another report states that juvenile and adult patients improve in the first two years of treatment, but begin to decline after this period (93). In an observational clinical study over a period of 5 years, similar results were obtained with stabilization in most, but not all non-classic Pompe patients (94). These findings indicate that juvenile and adult Pompe patients show heterogeneous response to ERT, warranting further research to new types of treatment. Currently, a number of new treatment options have been under investigation, including improved versions of ERT (95), chaperone-mediated improvement of mutant GAA (96), gene therapy approaches (97), and use of AONs for lowering glycogen synthesis (98). In this thesis, we focused on antisense mediated therapy to correct a splicing defect caused by the most prevalent mutation in Pompe disease, the c.-32-13T>G (IVS1) variant. This variant has a high prevalence in the Caucasian population ranging from 40-70% of Pompe patients. It is located in the polypyrimidine tract 13 bases upstream of *GAA* exon 2 and causes aberrant splicing by weakening the recognition of the canonical splice acceptor site of exon 2 (Figure 3a) (16-19). This results in complete or partial skipping of

the exon in approximately 85-90% of *GAA* transcripts (Figure 3B). 10-15% of the *GAA* mRNA is normally spliced from the IVS1 allele, resulting in some residual GAA enzymatic activity. However, in combination with a deleterious mutation present on the second allele, the residual combined activity of approximately 10-15% is often insufficient for adequate glycogen clearance, leading to juvenile or adult onset Pompe disease (figure 3C). Restauration of canonical splicing could lead to an maximum increase of 50% of normal protein activity from the IVS1 allele, which would be sufficient for rescue from the disease, highlighting the potential of AON mediated therapy for this particular variant.

AIMS AND SCOPE OF THE THESIS

Aim

The number of gene lesions that affect pre-mRNA splicing are estimated to be considerably underrepresented in the human gene mutation database. This is a result of insufficient functional investigation of the effect that these lesions may have on the splicing process. The aim of this thesis is to investigate such effects by investigating splicing of the *GAA* gene in patients with Pompe. Moreover, a better understanding of the effect of aberrant pre-mRNA splicing in these patients could lead to the development of new AON based therapy by targeting these aberrant splicing events to restore wild type splicing.

Scope

Functional assays for the unbiased detection of aberrant pre-mRNA splicing in patients with Pompe disease are currently rarely used. In **Chapter 2**, we describe a PCR based assay that can be used for the unbiased detection of splicing defects in patient derived primary fibroblasts. The assay was able to identify aberrant splicing events, including splicing events that were not predicted using *in silico* splice prediction programs.

Further optimization of the splicing assay was achieved in **Chapter 3**, in which cycloheximide treatment of patient fibroblasts resulted in identification of aberrant splicing events that would normally not be detected due to their degradation by the nonsense-mediated decay pathway. Utilization of the splice splicing assay resulted in the detection of previously unidentified variants in two patients. Furthermore, the precise mechanism of pathogenicity of a *GAA* variant common in the Asian population was detected, leading to the successful targeting of an AON to partially rescue GAA enzymatic activity.

In **Chapter 4** we used the splicing assay for the identification of an unknown variant on one allele of a Pompe disease patient. We detected mis-splicing in this patient to be caused by a deep intronic variant that would not be detected in conventional diagnostic genomic DNA sequencing. We used an AON to correct aberrant splicing, leading to elevation of GAA enzymatic activity above the disease threshold. Furthermore, we tested AONs targeting cryptic splice sites that were identified in chapter 2 leading to improved GAA activity in two additional patients, underlining the potential of AON mediated therapy.

In **Chapter 5** we investigated the potential of AON mediated therapy for restoration of aberrant splicing caused by the common IVS1 variant. With a *U7* screening assay we were able to identify a specific AON target site, which resulted in upregulated wild-type splicing after blockage with an AON. These findings emphasize the possibility for treating IVS1 patients with AON therapy.

REFERENCES

1. Pompe JC. Over idiopathische hypertrophie van het hart. Vereenigingsverslag van het Genootschap ter bevordering van Natuur-, Genees- en Heelkunde te Amsterdam. Vergadering van de afdeling Geneeskunde, op woensdag 18 November. *Ned Tijdschr Geneeskd.* **1932**;76:304-11.
2. Hers HG. alpha-Glucosidase deficiency in generalized glycogenstorage disease (Pompe's disease). *Biochem J.* **1963**;86:11-6.
3. Hoefsloot LH, Hoogeveen-Westerveld M, Reuser AJ, Oostra BA. Characterization of the human lysosomal alpha-glucosidase gene. *Biochem J*. **1990**;272(2):493-7.
4. Hoefsloot LH, Willemsen R, Kroos MA, Hoogeveen-Westerveld M, Hermans MM, Van der Ploeg AT, et al. Expression and routeing of human lysosomal alpha-glucosidase in transiently transfected mammalian cells. *Biochem J.* **1990**;272(2):485-92.
5. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* **1977**;74(12):5463-7.
6. Krawczak M, Reiss J, Cooper DN. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet.* **1992**;90(1-2):41-54.
7. Sterne-Weiler T, Howard J, Mort M, Cooper DN, Sanford JR. Loss of exon identity is a common mechanism of human inherited disease. *Genome Res.* **2011**;21(10):1563-71.
8. Di Giacomo D, Gaildrat P, Abuli A, Abdat J, Frebourg T, Tosi M, et al. Functional analysis of a large set of BRCA2 exon 7 variants highlights the predictive value of hexamer scores in detecting alterations of exonic splicing regulatory elements. *Hum Mutat.* **2013**;34(11):1547-57.
9. Soukarieh O, Gaildrat P, Hamieh M, Drouet A, Baert-Desurmont S, Frebourg T, et al. Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. *PLoS Genet.* **2016**;12(1):e1005756.
10. Lopez-Bigas N, Audit B, Ouzounis C, Parra G, Guigo R. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.* **2005**;579(9):1900-3.
11. Liu MM, Zack DJ. Alternative splicing and retinal degeneration. *Clin Genet.* **2013**;84(2):142-9.
12. Tanackovic G, Ransijn A, Thibault P, Abou Elela S, Klinck R, Berson EL, et al. PRPF mutations are associated with generalized defects in spliceosome formation and pre-mRNA splicing in patients with retinitis pigmentosa. *Hum Mol Genet.* **2011**;20(11):2116-30.
13. Burset M, Seledtsov IA, Solovyev VV. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res.* **2000**;28(21):4364-75.
14. Krawczak M, Thomas NS, Hundrieser B, Mort M, Wittig M, Hampe J, et al. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum Mutat.* **2007**;28(2):150-8.
15. Tsui LC, Dorfman R. The cystic fibrosis gene: a molecular genetic perspective. *Cold Spring Harb Perspect Med.* **2013**;3(2):a009472.
16. Huie ML, Chen AS, Tsujino S, Shanske S, DiMauro S, Engel AG, 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(12):2231-6.
17. Boerkoel CF, Exelbert R, Nicastri C, Nichols RC, Miller FW, Plotz PH, et al. Leaky splicing mutation in the acid maltase gene is associated with delayed onset of glycogenosis type II. *Am J Hum Genet.* **1995**;56(4):887-97.
18. Dardis A, Zanin I, Zampieri S, Stuani C, Pianta A, Romanello M, et al. Functional characterization of the common c.-32-13T>G mutation of GAA gene: identification of potential therapeutic agents. *Nucleic Acids Res.* **2014**;42(2):1291-302.

19. Bergsma AJ, Kroos M, Hoogeveen-Westerveld M, Halley D, van der Ploeg AT, Pijnappel WW. Identification and characterization of aberrant GAA pre-mRNA splicing in pompe disease using a generic approach. *Hum Mutat.* **2015**;36(1):57-68.
20. Hernandez-Imaz E, Martin Y, de Conti L, Melean G, Valero A, Baralle M, et al. Functional Analysis of Mutations in Exon 9 of NF1 Reveals the Presence of Several Elements Regulating Splicing. *PLoS One.* **2015**;10(10):e0141735.
21. Anczukow O, Buisson M, Leone M, Coutanson C, Lasset C, Calender A, et al. BRCA2 deep intronic mutation causing activation of a cryptic exon: opening toward a new preventive therapeutic strategy. *Clin Cancer Res.* **2012**;18(18):4903-9.
22. Rodriguez-Pascau L, Coll MJ, Vilageliu L, Grinberg D. Antisense oligonucleotide treatment for a pseudoexon-generating mutation in the NPC1 gene causing Niemann-Pick type C disease. *Hum Mutat.* **2009**;30(11):E993-E1001.
23. Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet.* **2016**.
24. Kaida D, Motoyoshi H, Tashiro E, Nojima T, Hagiwara M, Ishigami K, et al. Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat Chem Biol.* **2007**;3(9):576-83.
25. Andreassi C, Jarecki J, Zhou J, Coovert DD, Monani UR, Chen X, et al. Aclarubicin treatment restores SMN levels to cells derived from type I spinal muscular atrophy patients. *Hum Mol Genet.* **2001**;10(24):2841-9.
26. Mansfield SG, Kole J, Puttaraju M, Yang CC, Garcia-Blanco MA, Cohn JA, et al. Repair of CFTR mRNA by spliceosome-mediated RNA trans-splicing. *Gene Ther.* **2000**;7(22):1885-95.
27. Rindt H, Yen PF, Thebeau CN, Peterson TS, Weisman GA, Lorson CL. Replacement of huntingtin exon 1 by trans-splicing. *Cell Mol Life Sci.* **2012**;69(24):4191-204.
28. Avale ME, Rodriguez-Martin T, Gallo JM. Trans-splicing correction of tau isoform imbalance in a mouse model of tau mis-splicing. *Hum Mol Genet.* **2013**;22(13):2603-11.
29. He X, Liu F, Yan J, Zhang Y, Yan J, Shang H, et al. Trans-splicing repair of mutant p53 suppresses the growth of hepatocellular carcinoma cells in vitro and in vivo. *Sci Rep.* 2015;5:8705.
30. Lei Q, Li C, Zuo Z, Huang C, Cheng H, Zhou R. Evolutionary Insights into RNA trans-Splicing in Vertebrates. *Genome Biol Evol.* **2016**;8(3):562-77.
31. Dominski Z, Kole R. Restoration of correct splicing in thalassemic pre-mRNA by antisense oligonucleotides. *Proc Natl Acad Sci U S A.* **1993**;90(18):8673-7.
32. Sierakowska H, Sambade MJ, Agrawal S, Kole R. Repair of thalassemic human beta-globin mRNA in mammalian cells by antisense oligonucleotides. *Proc Natl Acad Sci U S A.* **1996**;93(23):12840-4.
33. Geary RS, Norris D, Yu R, Bennett CF. Pharmacokinetics, biodistribution and cell uptake of antisense oligonucleotides. *Adv Drug Deliv Rev.* **2015**;87:46-51.
34. Goyenvalle A, Griffith G, Babbs A, El Andaloussi S, Ezzat K, Avril A, et al. Functional correction in mouse models of muscular dystrophy using exon-skipping tricyclo-DNA oligomers. *Nat Med.* **2015**;21(3):270-5.
35. Boisguerin P, Deshayes S, Gait MJ, O'Donovan L, Godfrey C, Betts CA, et al. Delivery of therapeutic oligonucleotides with cell penetrating peptides. *Adv Drug Deliv Rev.* **2015**;87:52-67.
36. Petrilli R, Eloy JO, Marchetti JM, Lopez RF, Lee RJ. Targeted lipid nanoparticles for antisense oligonucleotide delivery. *Curr Pharm Biotechnol.* **2014**;15(9):847-55.
37. Osorio FG, Navarro CL, Cadinanos J, Lopez-Mejia IC, Quiros PM, Bartoli C, et al. Splicing-directed therapy in a new mouse model of human accelerated aging. *Sci Transl Med.* **2011**;3(106):106ra7.
38. Lentz JJ, Jodelka FM, Hinrich AJ, McCaffrey KE, Farris HE, Spalitta MJ, et al. Rescue of hearing and vestibular function by antisense oligonucleotides in a mouse model of human deafness. *Nat Med.* **2013**;19(3):345-50.
39. Igreja S, Clarke LA, Botelho HM, Marques L, Amaral MD. Correction of a Cystic Fibrosis Splicing Mutation by Antisense Oligonucleotides. *Hum Mutat.* **2016**;37(2):209-15.
40. Farooqi AA, Rehman ZU, Muntane J. Antisense therapeutics in oncology: current status. *Onco Targets Ther.* **2014**;7:2035-42.
41. Townsend D, Yasuda S, Metzger J. Cardiomyopathy of Duchenne muscular dystrophy: pathogenesis and prospect of membrane sealants as a new therapeutic approach. *Expert Rev Cardiovasc Ther.* **2007**;5(1):99-109.

42. Falzarano MS, Scotton C, Passarelli C, Ferlini A. Duchenne Muscular Dystrophy: From Diagnosis to Therapy. *Molecules.* **2015**;20(10):18168-84.
43. Monaco AP, Bertelson CJ, Liechti-Gallati S, Moser H, Kunkel LM. An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. *Genomics.* **1988**;2(1):90-5.
44. Takeshima Y, Nishio H, Sakamoto H, Nakamura H, Matsuo M. Modulation of in vitro splicing of the upstream intron by modifying an intra-exon sequence which is deleted from the dystrophin gene in dystrophin Kobe. *J Clin Invest.* **1995**;95(2):515-20.
45. Wilton SD, Lloyd F, Carville K, Fletcher S, Honeyman K, Agrawal S, et al. Specific removal of the nonsense mutation from the mdx dystrophin mRNA using antisense oligonucleotides. *Neuromuscul Disord.* **1999**;9(5):330-8.
46. van Deutekom JC, Bremmer-Bout M, Janson AA, Ginjaar IB, Baas F, den Dunnen JT, et al. Antisense-induced exon skipping restores dystrophin expression in DMD patient derived muscle cells. *Hum Mol Genet.* **2001**;10(15):1547-54.
47. Aartsma-Rus A, Bremmer-Bout M, Janson AA, den Dunnen JT, van Ommen GJ, van Deutekom JC. Targeted exon skipping as a potential gene correction therapy for Duchenne muscular dystrophy. *Neuromuscul Disord.* **2002**;12 Suppl 1:S71-7.
48. Mann CJ, Honeyman K, Cheng AJ, Ly T, Lloyd F, Fletcher S, et al. Antisense-induced exon skipping and synthesis of dystrophin in the mdx mouse. *Proc Natl Acad Sci U S A.* **2001**;98(1):42-7.
49. Lu QL, Rabinowitz A, Chen YC, Yokota T, Yin H, Alter J, et al. Systemic delivery of antisense oligoribonucleotide restores dystrophin expression in body-wide skeletal muscles. *Proc Natl Acad Sci U S A.* **2005**;102(1):198-203.
50. Alter J, Lou F, Rabinowitz A, Yin H, Rosenfeld J, Wilton SD, et al. Systemic delivery of morpholino oligonucleotide restores dystrophin expression bodywide and improves dystrophic pathology. *Nat Med.* **2006**;12(2):175-7.
51. Hoffman EP, Connor EM. Orphan drug development in muscular dystrophy: update on two large clinical trials of dystrophin rescue therapies. *Discov Med.* **2013**;16(89):233-9.
52. van Deutekom JC, Janson AA, Ginjaar IB, Frankhuizen WS, Aartsma-Rus A, Bremmer-Bout M, et al. Local dystrophin restoration with antisense oligonucleotide PRO051. *N Engl J Med.* **2007**;357(26):2677-86.
53. Mendell JR, Rodino-Klapac LR, Sahenk Z, Roush K, Bird L, Lowes LP, et al. Eteplirsen for the treatment of Duchenne muscular dystrophy. *Ann Neurol.* **2013**;74(5):637-47.
54. Kinali M, Arechavala-Gomeza V, Feng L, Cirak S, Hunt D, Adkin C, et al. Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study. *Lancet Neurol.* **2009**;8(10):918-28.
55. Lu QL, Cirak S, Partridge T. What Can We Learn From Clinical Trials of Exon Skipping for DMD? *Mol Ther Nucleic Acids.* 2014;3:e152.
56. Pearn J. Incidence, prevalence, and gene frequency studies of chronic childhood spinal muscular atrophy. *J Med Genet.* **1978**;15(6):409-13.
57. Crawford TO, Pardo CA. The neurobiology of childhood spinal muscular atrophy. *Neurobiol Dis.* **1996**;3(2):97-110.
58. Lefebvre S, Burglen L, Reboullet S, Clermont O, Burlet P, Viollet L, et al. Identification and characterization of a spinal muscular atrophy-determining gene. *Cell.* **1995**;80(1):155-65.
59. Lorson CL, Androphy EJ. An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene SMN. *Hum Mol Genet.* **2000**;9(2):259-65.
60. Porensky PN, Burghes AH. Antisense oligonucleotides for the treatment of spinal muscular atrophy. *Hum Gene Ther.* **2013**;24(5):489-98.
61. Cartegni L, Krainer AR. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat Genet.* **2002**;30(4):377-84.
62. Wirth B, Brichta L, Schrank B, Lochmuller H, Blick S, Baasner A, et al. Mildly affected patients with spinal muscular atrophy are partially protected by an increased SMN2 copy number. *Hum Genet.* **2006**;119(4):422-8.
63. Prior TW, Krainer AR, Hua Y, Swoboda KJ, Snyder PC, Bridgeman SJ, et al. A positive modifier of spinal muscular atrophy in the SMN2 gene. *Am J Hum Genet.* **2009**;85(3):408-13.
64. Skordis LA, Dunckley MG, Yue B, Eperon IC, Muntoni F. Bifunctional antisense oligonucleotides provide a trans-acting splicing enhancer that stimulates SMN2 gene expression in patient fibroblasts. *Proc Natl Acad Sci U S A.* **2003**;100(7):4114-9.

65. Madocsai C, Lim SR, Geib T, Lam BJ, Hertel KJ. Correction of SMN2 Pre-mRNA splicing by antisense U7 small nuclear RNAs. *Mol Ther.* **2005**;12(6):1013-22.
66. Singh NK, Singh NN, Androphy EJ, Singh RN. Splicing of a critical exon of human Survival Motor Neuron is regulated by a unique silencer element located in the last intron. *Mol Cell Biol.* **2006**;26(4):1333-46.
67. Hua Y, Liu YH, Sahashi K, Rigo F, Bennett CF, Krainer AR. Motor neuron cell-nonautonomous rescue of spinal muscular atrophy phenotypes in mild and severe transgenic mouse models. *Genes Dev.* **2015**;29(3):288-97.
68. Rigo F, Chun SJ, Norris DA, Hung G, Lee S, Matson J, et al. Pharmacology of a central nervous system delivered 2'-O-methoxyethyl-modified survival of motor neuron splicing oligonucleotide in mice and nonhuman primates. *J Pharmacol Exp Ther.* **2014**;350(1):46-55.
69. Chiriboga CA, Swoboda KJ, Darras BT, Iannaccone ST, Montes J, De Vivo DC, et al. Results from a phase 1 study of nusinersen (ISIS-SMNRx) in children with spinal muscular atrophy. *Neurology.* **2016**;86(10):890-7.
70. Ausems MG, Verbiest J, Hermans MP, Kroos MA, Beemer FA, Wokke JH, et al. Frequency of glycogen storage disease type II in The Netherlands: implications for diagnosis and genetic counselling. *Eur J Hum Genet.* **1999**;7(6):713-6.
71. Martiniuk F, Chen A, Mack A, Arvanitopoulos E, Chen Y, Rom WN, et al. Carrier frequency for glycogen storage disease type II in New York and estimates of affected individuals born with the disease. *Am J Med Genet.* **1998**;79(1):69-72.
72. Gungor D, Reuser AJ. How to describe the clinical spectrum in Pompe disease? *Am J Med Genet A.* **2013**;161A(2):399-400.
73. van den Hout HM, Hop W, van Diggelen OP, Smeitink JA, Smit GP, Poll-The BT, et al. The natural course of infantile Pompe's disease: 20 original cases compared with 133 cases from the literature. *Pediatrics.* **2003**;112(2):332-40.
74. Kishnani PS, Hwu WL, Mandel H, Nicolino M, Yong F, Corzo D, et al. A retrospective, multinational, multicenter study on the natural history of infantile-onset Pompe disease. *J Pediatr.* **2006**;148(5):671-6.
75. Gungor D, de Vries JM, Brusse E, Kruijshaar ME, Hop WC, Murawska M, et al. Enzyme replacement therapy and fatigue in adults with Pompe disease. *Mol Genet Metab.* **2013**;109(2):174-8.
76. Wokke JH, Escolar DM, Pestronk A, Jaffe KM, Carter GT, van den Berg LH, et al. Clinical features of late-onset Pompe disease: a prospective cohort study. *Muscle Nerve.* **2008**;38(4):1236-45.
77. Umapathysivam K, Hopwood JJ, Meikle PJ. Determination of acid alpha-glucosidase activity in blood spots as a diagnostic test for Pompe disease. *Clin Chem.* **2001**;47(8):1378-83.
78. Ausems MG, Lochman P, van Diggelen OP, Ploos van Amstel HK, Reuser AJ, Wokke JH. A diagnostic protocol for adult-onset glycogen storage disease type II. *Neurology.* **1999**;52(4):851-3.
79. Taniguchi N, Kato E, Yoshida H, Iwaki S, Ohki T, Koizumi S. alpha-glucosidase activity in human leucocytes: choice of lymphocytes for the diagnosis of Pompe's disease and the carrier state. *Clin Chim Acta.* **1978**;89(2):293-9.
80. Shin YS, Endres W, Unterreithmeier J, Rieth M, Schaub J. Diagnosis of Pompe's disease using leukocyte preparations. Kinetic and immunological studies of 1,4-alpha-glucosidase in human fetal and adult tissues and cultured cells. *Clin Chim Acta.* **1985**;148(1):9-19.
81. Matern D, Gavrilov D, Oglesbee D, Raymond K, Rinaldo P, Tortorelli S. Newborn screening for lysosomal storage disorders. *Semin Perinatol.* **2015**;39(3):206-16.
82. Verma J, D CT, Sharma S, Jhingan G, Saxena R, Kohli S, et al. Inherited metabolic disorders: prenatal diagnosis of lysosomal storage disorders. *Prenat Diagn.* **2015**;35(11):1137-47.
83. Hesselink RP, Wagenmakers AJ, Drost MR, Van der Vusse GJ. Lysosomal dysfunction in muscle with special reference to glycogen storage disease type II. *Biochim Biophys Acta.* **2003**;1637(2):164-70.
84. Hagemans ML, Stigter RL, van Capelle CI, van der Beek NA, Winkel LP, van Vliet L, et al. PAS-positive lymphocyte vacuoles can be used as diagnostic screening test for Pompe disease. *J Inherit Metab Dis.* **2010**;33(2):133-9.
85. de Barsy T, Jacquemin P, Van Hoof F, Hers HG. Enzyme replacement in Pompe disease: an attempt with purified human acid alpha-glucosidase. *Birth Defects Orig Artic Ser.* **1973**;9(2):184-90.
86. Hug G, Schubert WK. Lysosomes in type II glycogenosis. Changes during administration of extract from Aspergillus niger. *J Cell Biol.* **1967**;35(1):C1-6.

87. Fuller M, Van der Ploeg A, Reuser AJ, Anson DS, Hopwood JJ. Isolation and characterisation of a recombinant, precursor form of lysosomal acid alpha-glucosidase. *Eur J Biochem.* **1995**;234(3):903-9.
88. Bijvoet AG, Van Hirtum H, Kroos MA, Van de Kamp EH, Schoneveld O, Visser P, et al. Human acid alpha-glucosidase from rabbit milk has therapeutic effect in mice with glycogen storage disease type II. *Hum Mol Genet.* **1999**;8(12):2145-53.
89. Van den Hout JM, Reuser AJ, de Klerk JB, Arts WF, Smeitink JA, Van der Ploeg AT. Enzyme therapy for pompe disease with recombinant human alpha-glucosidase from rabbit milk. *J Inherit Metab Dis.* **2001**;24(2):266-74.
90. Kishnani PS, Nicolino M, Voit T, Rogers RC, Tsai AC, Waterson J, et al. Chinese hamster ovary cell-derived recombinant human acid alpha-glucosidase in infantile-onset Pompe disease. *J Pediatr.* **2006**;149(1):89-97.
91. van Gelder CM, Hoogeveen-Westerveld M, Kroos MA, Plug I, van der Ploeg AT, Reuser AJ. Enzyme therapy and immune response in relation to CRIM status: the Dutch experience in classic infantile Pompe disease. J *Inherit Metab Dis.* **2015**;38(2):305-14.
92. Toscano A, Schoser B. Enzyme replacement therapy in late-onset Pompe disease: a systematic literature review. *J Neurol.* **2013**;260(4):951-9.
93. Regnery C, Kornblum C, Hanisch F, Vielhaber S, Strigl-Pill N, Grunert B, et al. 36 months observational clinical study of 38 adult Pompe disease patients under alglucosidase alfa enzyme replacement therapy. *J Inherit Metab Dis.* **2012**;35(5):837-45.
94. Stepien KM, Hendriksz CJ, Roberts M, Sharma R. Observational clinical study of 22 adult-onset Pompe disease patients undergoing enzyme replacement therapy over 5 years. *Mol Genet Metab.* **2016**.
95. Maga JA, Zhou J, Kambampati R, Peng S, Wang X, Bohnsack RN, et al. Glycosylation-independent lysosomal targeting of acid alpha-glucosidase enhances muscle glycogen clearance in pompe mice. *J Biol Chem.* **2013**;288(3):1428-38.
96. Shimada Y, Nishimura E, Hoshina H, Kobayashi H, Higuchi T, Eto Y, et al. Proteasome Inhibitor Bortezomib Enhances the Activity of Multiple Mutant Forms of Lysosomal alpha-Glucosidase in Pompe Disease. *JIMD Rep.* **2015**;18:33-9.
97. van Til NP, Stok M, Aerts Kaya FS, de Waard MC, Farahbakhshian E, Visser TP, et al. Lentiviral gene therapy of murine hematopoietic stem cells ameliorates the Pompe disease phenotype. *Blood.* **2010**;115(26):5329-37.
98. Clayton NP, Nelson CA, Weeden T, Taylor KM, Moreland RJ, Scheule RK, et al. Antisense Oligonucleotide-mediated Suppression of Muscle Glycogen Synthase 1 Synthesis as an Approach for Substrate Reduction Therapy of Pompe Disease. *Mol Ther Nucleic Acids.* **2014**;3:e206.
99. Moraes F, Goes A. A decade of human genome project conclusion: Scientific diffusion about our genome knowledge. *Biochem Mol Biol Educ.* **2016**.
100. Turunen JJ, Niemela EH, Verma B, Frilander MJ. The significant other: splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA*. **2013**;4(1):61-76.
101. Moore MJ, Sharp PA. Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature.* **1993**;365(6444):364-8.
102. Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **1987**;15(17):7155-74.
103. Rogozin IB, Milanesi L. Analysis of donor splice sites in different eukaryotic organisms. *J Mol Evol.* **1997**;45(1):50-9.
104. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol.* **2004**;11(2-3):377-94.
105. Brunak S, Engelbrecht J, Knudsen S. Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J Mol Biol.* **1991**;220(1):49-65.
106. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol.* **1997**;4(3):311-23.
107. Jian X, Boerwinkle E, Liu X. In silico tools for splicing defect prediction: a survey from the viewpoint of end users. *Genet Med.* **2014**;16(7):497-503.
108. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* **2003**;31(13):3568-71.
109. Liu HX, Zhang M, Krainer AR. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.* **1998**;12(13):1998-2012.

110. Liu HX, Chew SL, Cartegni L, Zhang MQ, Krainer AR. Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol Cell Biol.* **2000**;20(3):1063-71.
111. Fairbrother WG, Yeh RF, Sharp PA, Burge CB. Predictive identification of exonic splicing enhancers in human genes. *Science.* **2002**;297(5583):1007-13.
112. Sironi M, Menozzi G, Riva L, Cagliani R, Comi GP, Bresolin N, et al. Silencer elements as possible inhibitors of pseudoexon splicing. Nucleic Acids Res. **2004**;32(5):1783-91.
113. Zhang XH, Leslie CS, Chasin LA. Computational searches for splicing signals. *Methods.* **2005**;37(4):292-305.
114. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **2009**;37(9):e67.
115. Schwartz S, Hall E, Ast G. SROOGLE: webserver for integrative, user-friendly visualization of splicing signals. *Nucleic Acids Res.* **2009**;37(Web Server issue):W189-92.

Chapter 2

Identification and Characterization of aberrant *GAA* pre-mRNA splicing in Pompe Disease Using a Generic Approach

Atze J. Bergsma[1,2,3]
Marian Kroos[1,3]
Marianne Hoogeveen-Westerveld[1,3]
Dicky Halley[4]
Ans T. van der Ploeg[2,3]
W.W.M. Pim Pijnappel[1,2,3]

[1] Molecular Stem Cell Biology, Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, the Netherlands
[2] Department of Pediatrics, Erasmus MC University Medical Center, Rotterdam, the Netherlands
[3] Center for Lysosomal and Metabolic Diseases, Erasmus MC University Medical Center, Rotterdam, the Netherlands
[4] Molecular Diagnostics, Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, the Netherlands

Human Mutation 2015; 36(1):57-68

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 α-glucosidase (*GAA*) gene causing Pompe disease enabled detection of a total of eleven 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.

INTRODUCTION

Identification of variants involved in human inherited disease is important for diagnosis, prediction of disease severity, and genetic counseling. Around 7000 rare genetic diseases are known, and in approximately 50% of cases the gene involved has been identified (1). At present, over 148000 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 (2)). A number of splicing prediction programs exist (3-6), 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, while a change of the reading frame results in a premature termination codon leading to mRNA degradation via the Nonsense-Mediated Decay (NMD) pathway (7).

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* α*-glucosidase* (*GAA*) gene (MIM# 606800, reviewed in (8)). 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 (9-13). 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 (14,15). In general, a higher residual enzyme activity in primary fibroblasts is associated with later onset of Pompe disease (16). Enzyme replacement therapy (ERT) has been developed for Pompe disease, in which recombinant human GAA protein is administered intravenously (17,18). This treatment can rescue the lives of patients with classic infantile Pompe disease (19) and delay disease progression of patients with later onset (20-22), 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 (23). 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). The activity of GAA in fibroblasts was measured with 4-methylumbelliferone (4-MU)-α-D-glucopyranoside as substrate as described previously (14). A more detailed description of Materials and Methods can be found in the Supporting Information.

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 (Figure 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 semi-quantitative and yields primarily qualitative information on the identity of the aberrantly spliced products. Second (Figure 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 β-*Actin* mRNA and for expression in a healthy control. The 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 (8). 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 (Figure 2A and 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 (24). Individual coding exons 2-20 (the translation start codon is localized in exon 2) were quantified using exon-internal qPCR (Figure 2E). Values were normalized

for β-*Actin* expression (as 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, S3, and S4

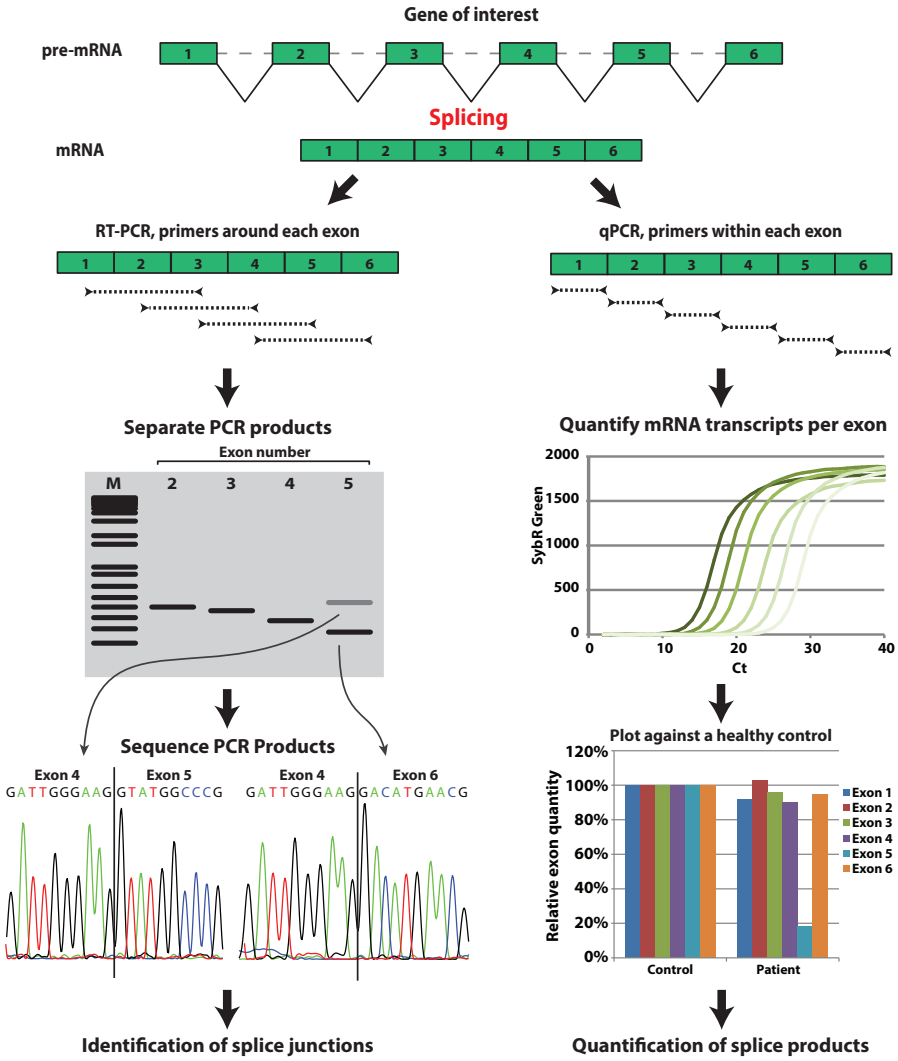


Figure 1. Workflow for the generic analysis of splice site variants. *Qualitative analysis of changes in splice site usage is performed by PCR using primers annealing to the flanking exons (flanking exon PCR), followed by sequencing (left part). Quantitative analysis of aberrant splicing products is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

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 (10,12,13) (Supp. Figure S1 A-C, 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. Figure S1D,E). All five splicing prediction programs used here (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. Figure S2A).

Patient 2 was homozygous for the c.525delT variant, which undergoes NMD (25). Flanking exon PCR detected low levels of products for all exons, which were not derived from genomic DNA contamination (Supp. Figure S3B). This enabled identification of the c.525delT variant by sequence analysis (Supp. Figure S3A, D). Exon-internal qPCR analysis showed expression of 3-9% for all exons relative to the healthy control (Supp. Figure 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 (26), a nonsense variant which 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 (27) (Supp. Figure S4A, B, 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. Figure S4C). GAA enzyme activities were consistent with leaky wild type splicing and adult disease onset in patient 1, and 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

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
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

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 (Figure 2C). This variant is suspected to affect splicing of exon 2

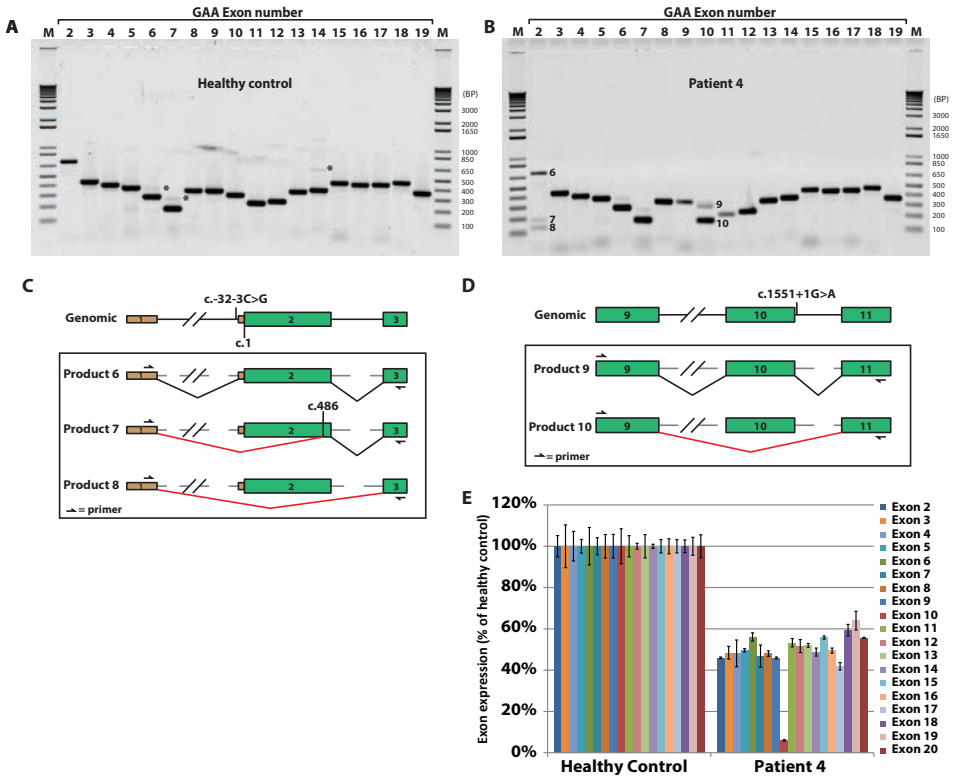


Figure 2. Splicing analysis of a healthy control and Pompe patient 4. *A) Flanking exon PCR analysis of a healthy control. Exon numbers are indicated above the lanes. PCR products were separated by electrophoresis on an agarose gel. Asterisks indicate alternative splicing events detected in both healthy 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, 7 and 8 (lower cartoons, boxed) detected for patient 4. The translation start site is indicated as c.1. Exons are indicated as boxes. Non-coding 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 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. β-actin was used for normalization. Values obtained from the healthy control were set to 100%. Error bars indicate SD (n=3).*

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	(12)	leaky wt splicing	r.=	in frame	p.=	(10,12,13)
				Partial skipping exon 2	r.-32_486del	out of frame	p.?	
				Perfect skipping exon 2	r.-32_546del	out of frame	p.?	
	c.1636+5G>T	intron 11	(33)	intron 11 inclusion	r.1636_1637ins1636+1_1636+957; 1636+5g>u	out of frame	p.V547Hfs*146	(33)
2	c.525delT	exon 2	(25)	premature stop codon	r.525delu	out of frame	p.E176Rfs*45	(25)
3	c.1548G>A	exon10	(26)	premature stop codon	r.1548g>a	new stop codon	p.W516*	(26)
	c.2481+102_2646+31del (del ex18)	intron 17-intron 18	(12)	deletion of full exon 18	r.2482_2646del	in frame	p.G828_N882del	(12)
4	*c.-32-3C>G*	intron 1	*this study*	leaky wt splicing	r.=	in frame	p.=	*this study*
				Partial skipping exon 2	r.-32_486del	out of frame	p.?	*this study*
				Perfect skipping exon 2	r.-32_546del	out of frame	p.?	*this study*
	1551+1G>A	intron 10	(29)	perfect skipping exon 10	r.1438_1551del	in frame	p.V480_I517del	*this study*
5	*c.1075G>A*	exon 6	(31)	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	(33)	leaky wt splicing	r.=	in frame	p.=	(33)
				full intron 10 inclusion	r.1551_1552ins1551+1_1551+100; 1552-3c>g	out of frame	p.D518Vfs*7	*this study*
				partial inclusion intron 10	r.1438_1551delins1437+71_1437+100; 1552-3c>g	in frame	p.V480_I517delins10	(33)
7	*c.1437G>A*	exon 9	(28)	leaky wt splicing	r.1437g>a	in frame	p.=	(33)
				perfect skipping exon 9	r.1327_1437del	in frame	p.D443_K479del	*this study*
8	*c.1256A>T*	exon 8	(32)	leaky wt splicing	r.1256a>u	in frame	p.D419V	*this study*
				partial skip of exon 8	r.1255_1326del	in frame	p.D419_V442del	*this study*
8	*c.1551+1G>T*	intron 10	(32)	leaky wt splicing	r.=	in frame	p.=	*this study*
				perfect skipping of exon 10	r.1438_1551del	in frame	p.V480_I517del	*this study*

Reference sequences used for cDNA and protein annotation are NM_000152.3 and NP_000143.2, respectively. Intronic sequences are derived chr17:80,101,556-80,119,880 (GRCh38

based on its similarity to the published c.-32-3C>A variant (28). 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. Figure S2C). The second allele contained a previously reported (29) but uncharacterized variant at c.1551+1G>A which is located in intron 10 close to the splice donor site of exon 10 (Figure 2D). Based on the similarity to the published c.1551+1G>C variant (12,30), 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. Figure S2C).

The results of the flanking exon PCR analysis indicated aberrant splicing of two exons: exon 2 and exon 10 (Figure 2B). Amplification of exon 2 resulted in 3 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 (Figure 2C and Supp. Figure 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 (Figure 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 (Figure 2D and Supp. Figure 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%, while all other exons were expressed at ~50% of healthy control levels (Figure 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 to the IVS1 variant, as the IVS1 variant allows a higher level of wild type splicing of 10-20% (Supp. Figure S1C). In contrast to the situation with e.g. the c.525delT allele of patient 2, in which residual mRNA 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 to 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 (Figure 3B) (31). This variant has been classified as presumably nonpathogenic with possible effects on splicing (32). 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. Figure 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 (Figure 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 4 nucleotides upstream at c.1071 was used instead (Figure 3B and Supp. Figure S6B). This explains the absence of a product for exon 7, as the forward primer for exon 7 amplification has 4 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 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. Figure 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 (Figure 3G). This confirmed the notion that leaky wild type splicing levels in this patient are very low or absent, while 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 (Figure 3D). Flanking exon PCR analysis showed aberrant splicing of exon 10 with three major products (12-14; Figure 3C). Sequence analysis indicated that in product 14, exon 10 was completely skipped while a splice acceptor site near exon 11 at

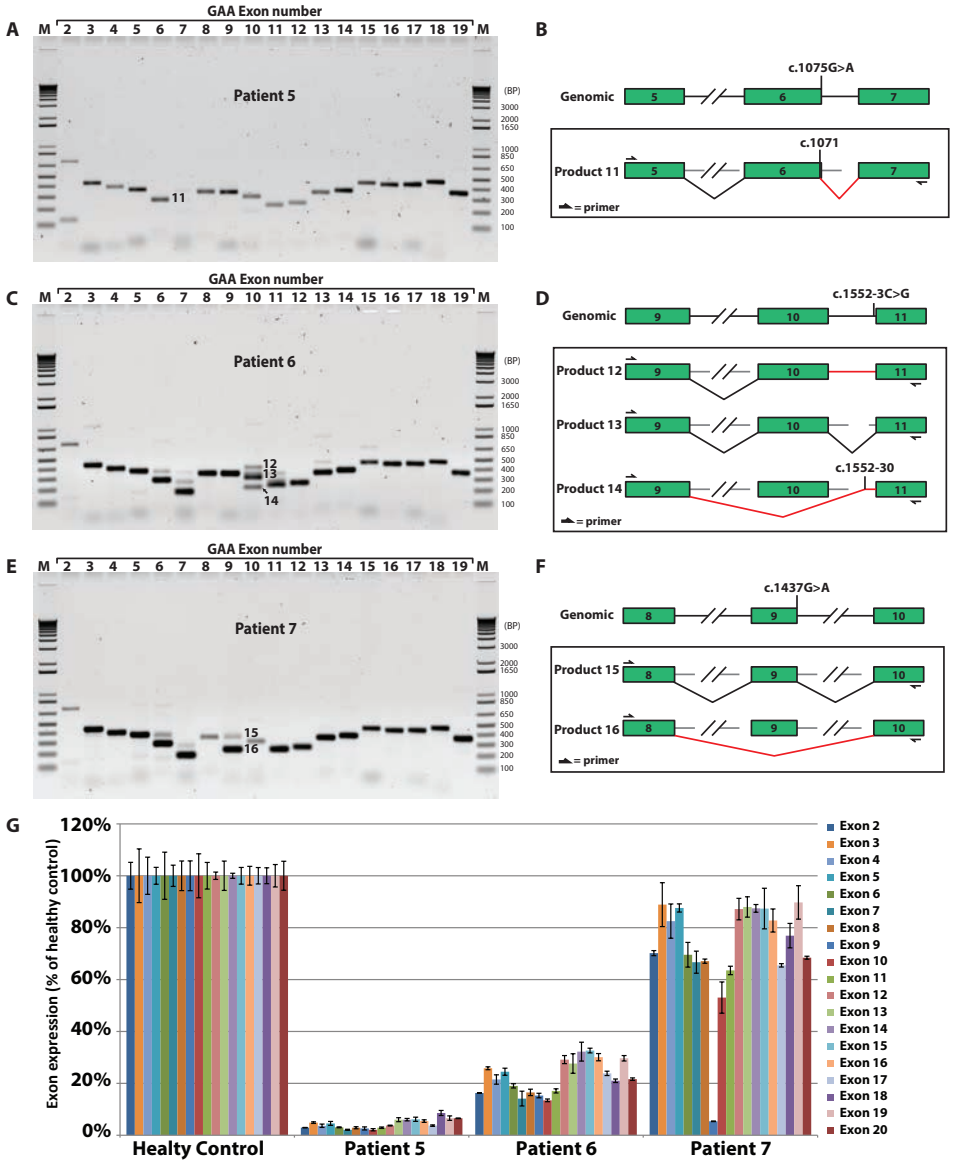


Figure 3. Splicing analysis of Pompe patients carrying homozygous variants. *A) Flanking exon PCR analysis of patient 5. B) Cartoons of the genomic variants and the splicing variant detected for patient 5. C) Flanking exon PCR analysis of patient 6. D) Cartoons of the genomic variant and the splicing 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, 6, and 7. Error bars indicate SD (n=3).*

c.1552-30 was utilized (Figure 3D and Supp. Figure 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

(33), 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. Figure S8F). 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 (Figure 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 (Figure 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 (Figure 3E). Sequence analysis indicated that product 15 represented wild type spliced exon 9, while in product 16, exon 9 was perfectly skipped, resulting in a shorter transcript in which the reading frame was unchanged (Figure 3F and Supp. Figure 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. Figure 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 non-classic 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 (11,34). 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 (Figure 4B,C) (32). The 2nd allele contained a c.1551+1G>T variant, which is located in intron 10 close to the splice donor site of exon 10 (32). It resembles the c.1551+1G>A variant described above for patient 4 (Figure 4D).

Analysis of splicing products

Flanking exon PCR analysis indicated multiple PCR products from amplification of exons 8, 9, and 10 (Figure 4A). All these products were analyzed by sequencing (Supp. Figure 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 Figure 4B and product 20 in Figure 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. Figure 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 (35) 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, while product 22 was the result of perfect exon 10 skipping in which the reading frame remained intact (Figure 4D and Supp. Figure S7). Loss of the exon 10 splice donor site by the c.1551+1G>T variant was consistent with splicing predictions (Supp. Figure 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 the variants via the relative intensities of the products. These can be explained based on the identification of the splicing products (Figure 4B-D) and on the locations of the primers used for amplification (Supp. Figure S7).

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. Figure S8). The 5'end of exon 8 is retained in all cases while 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 (Figure 4A).

Exon 8

Flanking exon PCR primers used for detection of exon 8 anneal to exon 7 and 9 (Supp. Figure S8). Both exons are not affected in this patient predicting that all splicing alterations of exon 8 itself should be detected in a semi-quantitative 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. Figure S8). This complicates detection of exon 9 from these two alleles: a

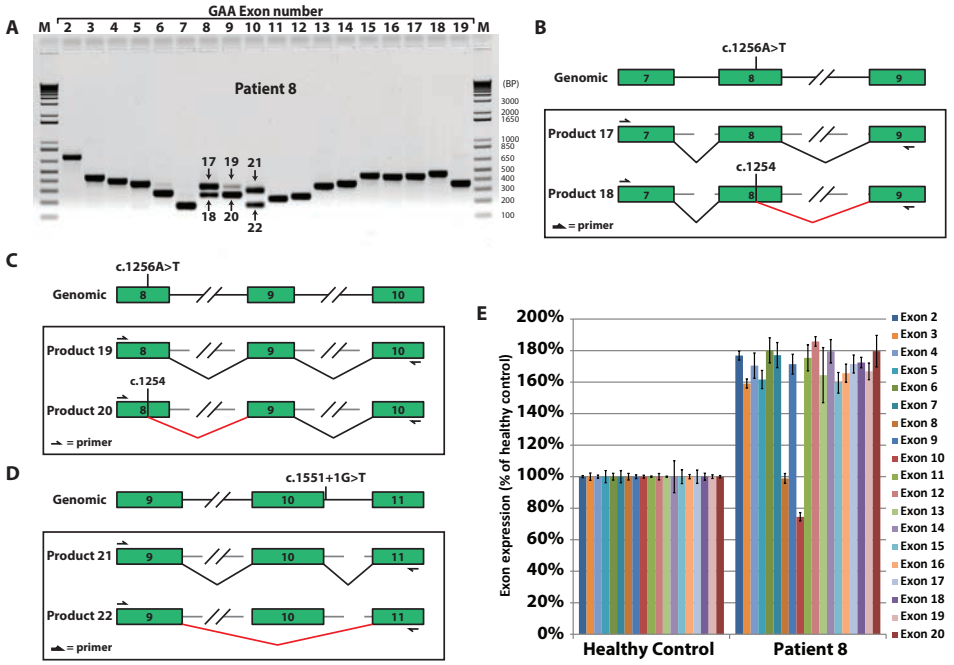


Figure 4. Analysis of complex splicing changes in Pompe patient 8. *A) Flanking exon PCR analysis. B) Cartoons of the genomic variant and the splicing variants from allele 1, detected from flanking exon PCR analysis of exon 8. C) Cartoons of the genomic variant and the splicing variants from allele 1, 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).*

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 which 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 semi-quantitative manner. Product 21 representing wild type splicing was the most abundant, while 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 two-fold higher abundance compared to the healthy control. Exons 8 and 10 were expressed at 2-fold 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 (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 2-fold 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 which has been reported to lower the activity of the corresponding protein (32). 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 to 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 4 nucleotides in patient 5 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 like 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: i) 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. ii) A variant-specific PCR strategy may be developed as shown in Supp. Figure S1D, 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 (36). 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. Figure 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 (37-39). 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 (40-42). 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 (43). 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 (10,12,13), 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 (44). 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). (i) 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 (45). Surprisingly, all full size exons could be detected by flanking exon PCR (Supp. Figure 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 (25). (ii) 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. (iii) 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 (13,46). 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 suggest 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. 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 (24) 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 (47,48) may present a worthwhile therapeutic strategy to treat genetic diseases caused by splicing variants.

ACKNOWLEDGEMENTS

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.

CONFLICT OF INTEREST

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

REFERENCES

1. Boycott, K.M., Vanstone, M.R., Bulman, D.E. and MacKenzie, A.E. (2013) Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet*, **14**, 681-691.
2. Havens, M.A., Duelli, D.M. and Hastings, M.L. (2013) Targeting RNA splicing for disease therapy. *Wiley Interdiscip Rev RNA*, **4**, 247-266.
3. Desmet, F.O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M. and Beroud, C. (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*, **37**, e67.
4. Yeo, G. and Burge, C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*, **11**, 377-394.
5. Reese, M.G., Eeckman, F.H., Kulp, D. and Haussler, D. (1997) Improved splice site detection in Genie. *J Comput Biol*, **4**, 311-323.
6. Pertea, M., Lin, X. and Salzberg, S.L. (2001) GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res*, **29**, 1185-1190.
7. Palacios, I.M. (2013) Nonsense-mediated mRNA decay: from mechanistic insights to impacts on human health. *Brief Funct Genomics*, **12**, 25-36.
8. van der Ploeg, A.T. and Reuser, A.J. (2008) Pompe's disease. *Lancet*, **372**, 1342-1353.
9. Zampieri, S., Buratti, E., Dominissini, S., Montalvo, A.L., Pittis, M.G., Bembi, B. and 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.
10. Boerkoel, C.F., Exelbert, R., Nicastri, C., Nichols, R.C., Miller, F.W., Plotz, P.H. and 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.
11. Herzog, A., Hartung, R., Reuser, A.J., Hermanns, P., Runz, H., Karabul, N., Gokce, S., Pohlenz, J., Kampmann, C., Lampe, C. et al. (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.
12. Huie, M.L., Chen, A.S., Tsujino, S., Shanske, S., DiMauro, S., Engel, A.G. and Hirschhorn, R. (1994) 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.
13. Dardis, A., Zanin, I., Zampieri, S., Stuani, C., Pianta, A., Romanello, M., Baralle, F.E., Bembi, B. and 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.
14. Kroos, M.A., Pomponio, R.J., Hagemans, M.L., Keulemans, J.L., Phipps, M., DeRiso, M., Palmer, R.E., Ausems, M.G., Van der Beek, N.A., Van Diggelen, O.P. et al. (2007) Broad spectrum of Pompe disease in patients with the same c.-32-13T->G haplotype. *Neurology*, **68**, 110-115.
15. Laforet, P., Laloui, K., Granger, B., Hamroun, D., Taouagh, N., Hogrel, J.Y., Orlikowski, D., Bouhour, F., Lacour, A., Salort-Campana, E. 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.

16. Umapathysivam, K., Hopwood, J.J. and Meikle, P.J. (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.
17. Van den Hout, H., Reuser, A.J., Vulto, A.G., Loonen, M.C., Cromme-Dijkhuis, A. and Van der Ploeg, A.T. (2000) Recombinant human alpha-glucosidase from rabbit milk in Pompe patients. *Lancet*, **356**, 397-398.
18. Kishnani, P., Voit, T., Nicolino, M., Tsai, C.H., Herman, G., Waterson, J., Rogers, R.C., Levine, J., Amalfitano, A., Landy, H. et al. (2003) Enzyme replacement therapy with recombinant human acid alpha glucosidase (rhGAA) in infantile Pompe disease (IPD): Results from a Phase 2 study. *Pediatr Res*, **53**, 259a-259a.
19. Kishnani, P.S., Corzo, D., Nicolino, M., Byrne, B., Mandel, H., Hwu, W.L., Leslie, N., Levine, J., Spencer, C., McDonald, M. et al. (2007) Recombinant human acid alpha-glucosidase - Major clinical benefits in infantile-onset Pompe disease. *Neurology*, **68**, 99-109.
20. Gungor, D., Kruijshaar, M.E., Plug, I., D'Agostino, R.B., Hagemans, M.L.C., van Doorn, P.A., Reuser, A.J.J. and van der Ploeg, A.T. (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**.
21. Regnery, C., Kornblum, C., Hanisch, F., Vielhaber, S., Strigl-Pill, N., Grunert, B., Muller-Felber, W., Glocker, F.X., Spranger, M., Deschauer, M. et al. (2012) 36 months observational clinical study of 38 adult Pompe disease patients under alglucosidase alfa enzyme replacement therapy. *J Inherit Metab Dis*, **35**, 837-845.
22. Toscano, A. and Schoser, B. (2013) Enzyme replacement therapy in late-onset Pompe disease: a systematic literature review. *J Neurol*, **260**, 951-959.
23. den Dunnen, J.T. and Antonarakis, S.E. (2000) Mutation nomenclature extensions and suggestions to describe complex mutations: A discussion. *Human Mutation*, **15**, 7-12.
24. Pickrell, J.K., Pai, A.A., Gilad, Y. and Pritchard, J.K. (2010) Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet*, **6**, e1001236.
25. Hermans, M.M., De Graaff, E., Kroos, M.A., Mohkamsing, S., Eussen, B.J., Joosse, M., Willemsen, R., Kleijer, W.J., Oostra, B.A. and Reuser, A.J. (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.
26. Hermans, M.M., van Leenen, D., Kroos, M.A., Beesley, C.E., Van Der Ploeg, A.T., Sakuraba, H., Wevers, R., Kleijer, W., Michelakakis, H., Kirk, E.P. 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.
27. Huie, M.L., Chen, A.S., Brooks, S.S., Grix, A. and Hirschhorn, R. (1994) 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.
28. Pittis, M.G., Donnarumma, M., Montalvo, A.L., Dominissini, S., Kroos, M., Rosano, C., Stroppiano, M., Bianco, M.G., Donati, M.A., Parenti, G. et al. (2008) Molecular and functional characterization of eight novel GAA mutations in Italian infants with Pompe disease. *Hum Mutat*, **29**, E27-36.
29. Orlikowski, D., Pellegrini, N., Prigent, H., Laforet, P., Carlier, R., Carlier, P., Eymard, B., Lofaso, F. and 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.
30. Stroppiano, M., Bonuccelli, G., Corsolini, F. and 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.
31. Muller-Felber, W., Horvath, R., Gempel, K., Podskarbi, T., Shin, Y., Pongratz, D., Walter, M.C., Baethmann, M., Schlotter-Weigel, B., Lochmuller, H. et al. (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.
32. Kroos, M., Hoogeveen-Westerveld, M., Michelakakis, H., Pomponio, R., Van der Ploeg, A., Halley, D., Reuser, A. and Consortium, G.A.A.D. (2012) 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.

33. Kroos, M., Manta, P., Mavridou, I., Muntoni, F., Halley, D., Van der Helm, R., Zaifeiriou, D., Van der Ploeg, A., Reuser, A. and Michelakakis, H. (2006) Seven cases of Pompe disease from Greece. *J Inherit Metab Dis*, **29**, 556-563.
34. Kroos, M., Hoogeveen-Westerveld, M., van der Ploeg, A. and Reuser, A.J. (2012) The genotype-phenotype correlation in Pompe disease. *Am J Med Genet C Semin Med Genet*, **160C**, 59-68.
35. Mucaki, E.J., Shirley, B.C. and Rogan, P.K. (2013) Prediction of mutant mRNA splice isoforms by information theory-based exon definition. *Hum Mutat*, **34**, 557-565.
36. Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R. et al. (2012) The evolutionary landscape of alternative splicing in vertebrate species. *Science*, **338**, 1587-1593.
37. Wang, G.S. and Cooper, T.A. (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet*, **8**, 749-761.
38. Kwan, T., Benovoy, D., Dias, C., Gurd, S., Provencher, C., Beaulieu, P., Hudson, T.J., Sladek, R. and Majewski, J. (2008) Genome-wide analysis of transcript isoform variation in humans. *Nat Genet*, **40**, 225-231.
39. Castle, J.C., Zhang, C., Shah, J.K., Kulkarni, A.V., Kalsotra, A., Cooper, T.A. and Johnson, J.M. (2008) Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat Genet*, **40**, 1416-1425.
40. Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. and Burge, C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470-476.
41. Lappalainen, T., Sammeth, M., Friedlander, M.R., t Hoen, P.A., Monlong, J., Rivas, M.A., Gonzalez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G. et al. (2013) Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, **501**, 506-511.
42. Lalonde, E., Ha, K.C., Wang, Z., Bemmo, A., Kleinman, C.L., Kwan, T., Pastinen, T. and Majewski, J. (2011) RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. *Genome Res*, **21**, 545-554.
43. Sharma, N., Sosnay, P.R., Ramalho, A.S., Douville, C., Franca, A., Gottschalk, L.B., Park, J., Lee, M., Vecchio-Pagan, B., Raraigh, K.S. *et al.* (2014) Experimental Assessment of Splicing Variants Using Expression Minigenes and Comparison with in Silico Predictions. *Hum Mutat.*
44. Wokke, J.H., Ausems, M.G., van den Boogaard, M.J., Ippel, E.F., van Diggelene, O., Kroos, M.A., Boer, M., Jennekens, F.G., Reuser, A.J. and Ploos van Amstel, H.K. (1995) Genotype-phenotype correlation in adult-onset acid maltase deficiency. *Ann Neurol*, **38**, 450-454.
45. Kishnani, P.S., Goldenberg, P.C., DeArmey, S.L., Heller, J., Benjamin, D., Young, S., Bali, D., Smith, S.A., Li, J.S., Mandel, H. *et al.* (2010) Cross-reactive immunologic material status affects treatment outcomes in Pompe disease infants. *Mol Genet Metab*, **99**, 26-33.
46. Lim, K.H., Ferraris, L., Filloux, M.E., Raphael, B.J. and Fairbrother, W.G. (2011) Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc Natl Acad Sci U S A*, **108**, 11093-11098.
47. Fan, L., Lagisetti, C., Edwards, C.C., Webb, T.R. and Potter, P.M. (2011) Sudemycins, novel small molecule analogues of FR901464, induce alternative gene splicing. *ACS Chem Biol*, **6**, 582-589.
48. Webb, T.R., Joyner, A.S. and Potter, P.M. (2013) The development and application of small molecule modulators of SF3b as therapeutic agents for cancer. *Drug Discov Today*, **18**, 43-49.0.

SUPPLEMENTARY MATERIALS AND METHODS

Patients and healthy control

Patients were diagnosed with Pompe disease based on clinical symptoms, GAA enzyme deficiency, and variant analysis.

Nomenclature

Nomenclature is according to HGVS standards (http://www.hgvs.org/mutnomen/), checked on 26 August 2014 (23).

Splicing Prediction

Splicing predictions were performed using Alamut Visual v.2.4.2 (Interactive Biosoftware), which applies five algorithms: SpliceSiteFinder-like (SSF), MaxEntScan (MES), NNSplice (NNS), GeneSplicer (GS) and Human Splicing Finder (HSF). Settings were identical to those used for standard diagnostics by Molecular Diagnostics at the Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands, and are provided in Supp. Table S1. The five algorithms use different maximal scores. For comparison, scores were calculated as % of the maximal score. A change of ≥10% in two or more algorithms was used as threshold for significance. Relative splice site strengths were also predicted using ASSEDA (http://splice.uwo.ca/; version August 2014) with a window range of 200 bases (35).

Cell Culture and cDNA preparation

Fibroblasts were isolated from skin biopsies of patients and a healthy individual. Cells were cultured in DMEM High Glucose (Lonza) + 10% Fetal bovine serum (HyClone, Thermo Scientific) + 1% penicillin / streptomycin (Lonza). Cells were grown in the presence of 10% CO2, harvested with Trypsin/EDTA (Lonza), flash frozen and stored at -80˚C. RNA was isolated using the RNAeasy miniprep kit with on-column DNase treatment according to the manufactor's recommendations (Qiagen). 800 ng of RNA was used for generation of cDNA using the iScript cDNA synthesis kit (Biorad). The cDNA solution was diluted 10 times before use.

Flanking exon PCR analysis

cDNA was amplified using FastStart Taq Polymerase (Roche). Primers were used at a final concentration of 0.333 µM each, dNTPs at 0.333 mM each. The PCR program was performed on a Biorad S1000 thermal cycler (96˚C for 4 min, 35X [96˚C for 20 sec, 60˚C for 30 sec, 72˚C for 1 min], 72˚C for 5 min); 5 µl of each PCR reaction was run on a 1.5% agarose gel containing ethidium bromide. Gels were photographed on a Typhoon FLA 9000 gel imager (G&E Healthcare).

Exon-internal qPCR analysis

To determine the relative concentration of each sample, 4 µl of each cDNA sample (10 times diluted in H2O) was processed in a 15 µl PCR reaction containing IQ Mastermix

(Biorad) and 0.333 µM of each primer. To account for the efficiency of each specific primer set, all samples were related to a standard curve from the healthy control sample. All samples were measured in triplicate and normalized against β-*actin*.

qPCR analysis of splicing isoforms

Primers were designed that specifically amplify splicing variants caused by the IVS1 variant (Supp. Figure S1E and Supp. Table S3). For detection of the normal variant (N), the forward primer annealed to exon 1, and the reverse primer to the beginning of exon 2, which is the part that is retained in this variant but is removed by splicing in the other two main variants. For detection of splice variant (SV) 3, the forward primer partially annealed to exon 1 and partially to exon 2 after the cryptic splice site, while the reverse primer partially annealed to exon 3 and partially to exon 2. For detection of SV2, the forward primer partially annealed to exon 1 and partially to exon 3, and the reverse primer to exon 3.

Quantitation was performed after normalization for β-*actin* using the delta-delta-Ct method. This was the only possible way to compare expression levels of different isoforms with each other. A prerequisite for using this method was that the different qPCR amplifications had similar efficiencies. This was indeed the case: efficiencies were 99% (β-*actin*), 92% (N), 99% (SV2), and 103% (SV3).

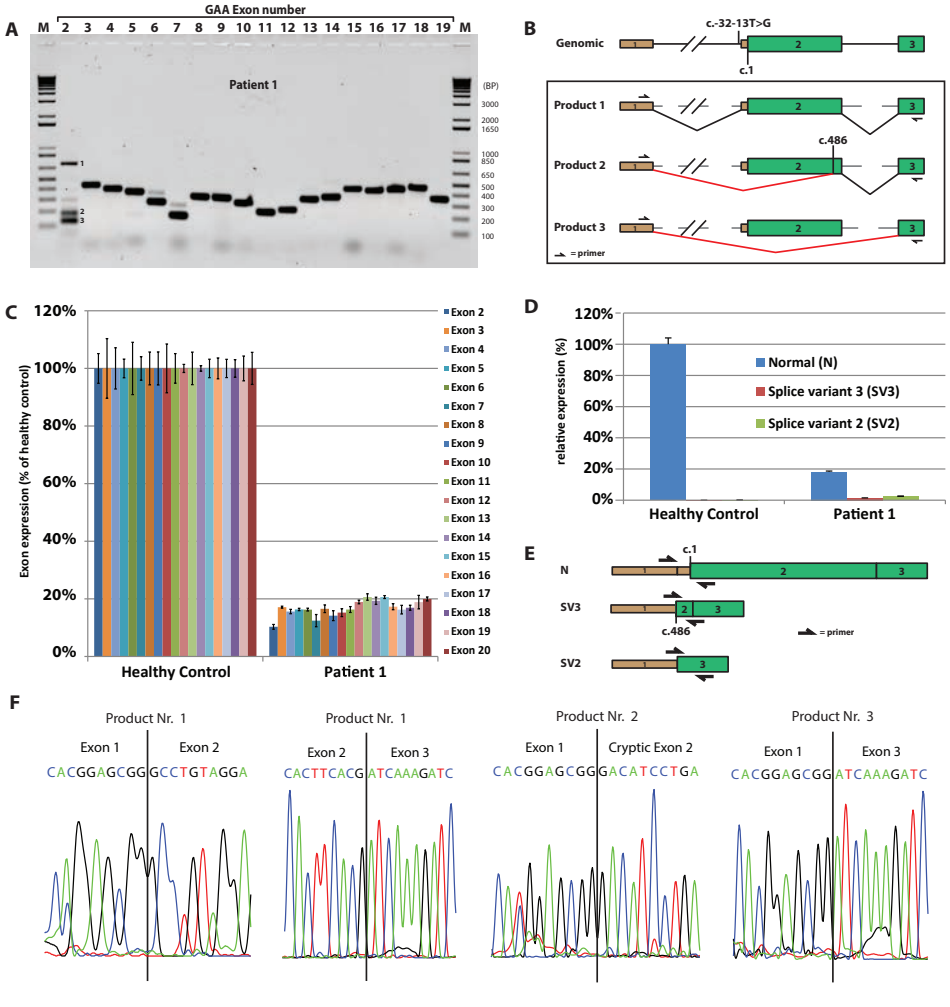
Sequencing

Genomic DNA variants were identified at Molecular Diagnostics at the Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands. Direct sequencing of flanking exon PCR products was performed using the Big Dye Terminator kit v3.1 (Applied Biosystems). To obtain pure DNA samples, PCR products visible on gel were stabbed with a 20 µl pipet tip and DNA on the tip was resuspended in 10 µl H2O. A 1 µl aliquot was subsequently used in a new PCR (as described above) to obtain DNA from a single template. Excess primers and dNTPs were removed using FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific), according to the manufacturer's protocol. Samples were purified with sephadex G-50 (GE Healthcare) and the sequence was determined on an AB3130 Genetic Analyzer (Applied Biosystems, Hitachi).

Time requirements

After development and validation of the approach for the gene of interest, the following guidelines can be used for performing the analysis. Flanking exon PCR analysis can be performed in one day for several patients simultaneously. Exon-internal qPCR and sequence analysis can also be done in one day. When the variant has not been detected by standard sequence analysis (involving the exons only), it may take longer to identify the corresponding variant.

SUPPLEMENTARY FIGURES, TABLES AND FIGURE LEGENDS



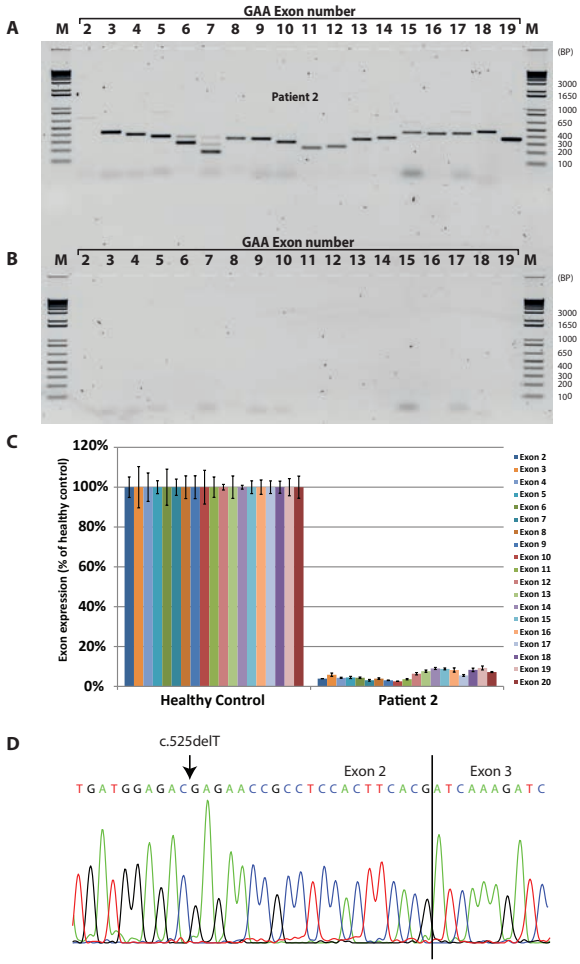
Supplementary Figure S1. Analysis of Patient 1: verification of known splicing events and quantification of splicing products. *This patient was tested to validate whether a well-described splicing variant could be accurately detected in primary fibroblasts using the approach described here. c.-32-13T>G is located in intron 1 close to the splice acceptor site of exon 2, and causes aberrant splicing of exon 2, but also allows leaky wild type splicing (10,12,13). The second allele of this patient carried the c.1636+5G>T variant. This variant is similar to the c.1636+5G>C variant, which is known to be expressed at very low levels due to NMD, caused by intron 11 inclusion and a premature termination codon (28). For this reason, the allele harboring the IVS1 variant is preferentially amplified in the splicing approach described below. Flanking exon PCR analysis of exon 2 to 19 yielded three major products from exon 2 amplification (panel A). Numbers indicate splicing products that were sequenced, which indicated that product 1 represented full-length exon 2 with canonical splice junctions (panels B and F). Product 2 contained partially skipped exon 2 due to the utilization of a cryptic splice acceptor site at c.486, while product 3 represented fully skipped exon 2 (panels B and F). These products correspond to the major splicing variants reported for the IVS1 variant, namely N (product 1), SV3 (product 2) and SV2 (product 3) (10,12,13). The known minor IVS1 splicing variants are expressed at levels too low to allow detection by flanking exon PCR and sequencing, consistent with previous reports (10,13).*

No aberrant flanking exon PCR products around exon 11 were observed, consistent with NMD of the products of the c.1636+5G>T variant.

Exon-internal qPCR analysis for all coding exons showed 10-20% expression of exon 2 and all other exons (panel C). This can be explained as follows. The IVS1 variant allows leaky wild type splicing of exon 2 (product 1 in panel B) yielding a normal mRNA containing all exons, as noted previously (10,12,13). The two other major products 2 and 3 both result in the deletion of the canonical translation start site, which is located in exon 2. This leads to mRNA degradation, resulting in minor contribution in the quantitative exon-internal qPCR assay, and predominant detection of the leaky wild type GAA mRNA from the IVS1 allele.

Whereas the exon-internal qPCR enables quantification of individual exons, it cannot be used to quantify the aberrantly spliced products. This requires the development of an isoform-specific qPCR method. To demonstrate that this is feasible, we have developed a qPCR strategy to quantify the splicing products of exon 2 of patient 1. The results are shown in panel D and the PCR strategy in panel E and Supp. Table S3. This shows that expression in patient 1 of the N, SV3, and SV2 is 18%, 1%, and 3% of the healthy control, respectively. SV2 and SV3 were not detected in the healthy control. The relatively low levels of SV2 and SV3 in patient 1 are most likely caused by NMD.

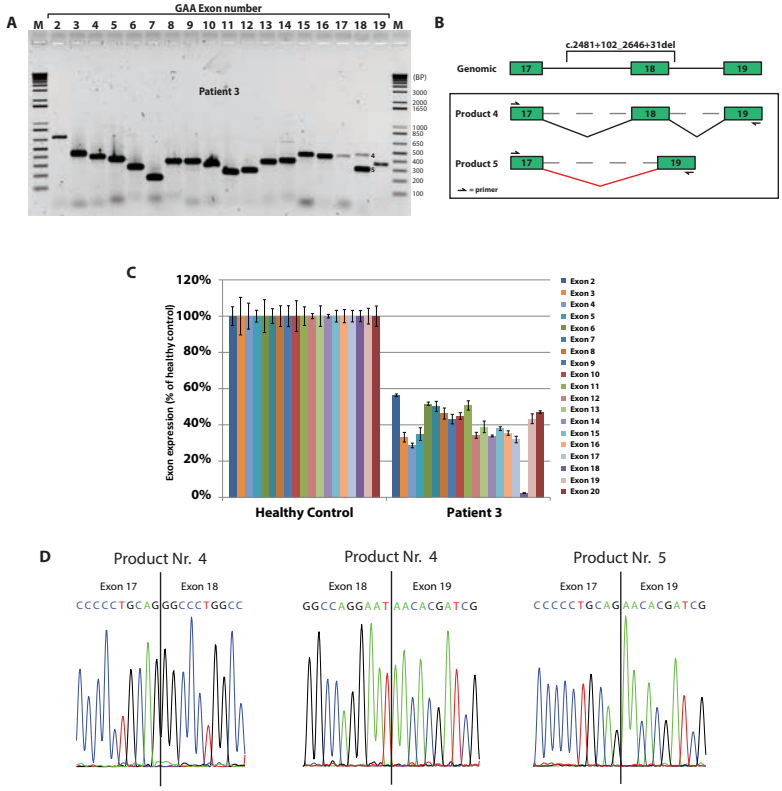
Supplementary Figure S2. Splicing predictions using five algorithms *(SpliceSiteFinder-like (SSF), MaxEntScan (MES), NNSplice (NNS), GeneSplicer (GS) and Human Splicing Finder (HSF)) applied to wild type and variant sequences.*



Supplementary Figure S3. Splicing analysis of patient 2: detection of RNA undergoing NMD and sequencing of the corresponding variant. *This patient was chosen to test the sensitivity of the approach. Due to a homozygous c.525delT variant, GAA mRNA expression is very low as a result of NMD (25). Surprisingly, flanking exon PCR analysis showed that all exons could still be detected at the correct sizes, although at strongly reduced levels compared to the healthy control (compare panel A and Figure 2A). Higher molecular weight products were also observed at even lower levels. These may represent unspliced pre-mRNA species. Control reactions without reverse transcriptase ruled out that these bands were derived from genomic DNA contamination (panel B). The absence of high levels of competing GAA mRNA is the likely cause of PCR detection of both GAA mRNA and GAA pre-mRNA species.*

To quantify the amount of residual mRNA, exon-internal qPCR was performed and showed 3-9 % expression of all exons relative to the healthy control (panel C). To validate that the obtained products in the flanking exon PCR were specific and were not the result of contamination, the product obtained for exon 2 was sequenced. Indeed, the product harbored the c.525delT variant, confirming that the flanking exon PCR products observed were derived from the homozygous c.525delT cells (panel D). Although the exon-internal qPCR showed a minimum of 3% expression of all exons, the reading frame shift in the mRNA that may have escaped degradation is expected to further reduce the amount of wild type protein expression to near background levels. In agreement, GAA enzyme activity was 1.3 nmol/hr/mg, consistent with the classic infantile onset of Pompe disease in this patient (Table 1).

In conclusion, the generic splicing approach for GAA can allow analysis and quantification of very low mRNA expression. This is particularly relevant for mRNAs that are subject to degradation as the result of reading frame alterations. Despite the low expression due to NMD, sequence analysis of the PCR product can still be used to identify the corresponding variant.

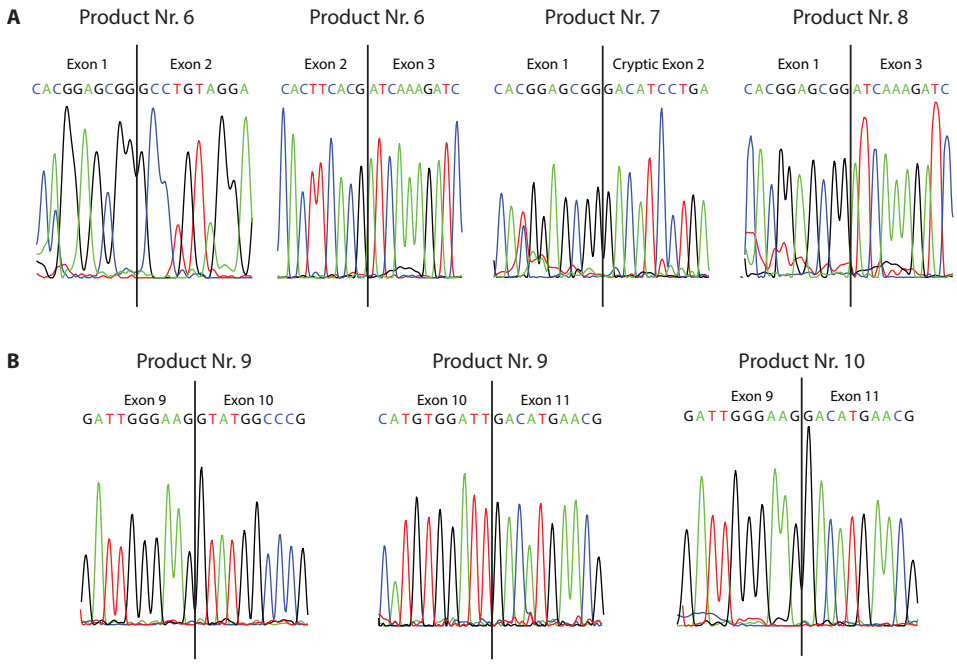


Supplementary Figure S4. Splicing analysis of patient 3 carrying the del ex18 variant. *A third validation was performed on a patient carrying a well-known deletion removing the entire exon 18 plus its flanking sequences (del ex18, or c.2481+102_2646+31del) (27). This variant is interesting because the splice sites of exon 18 are removed. Previous work has shown that a new mRNA is formed in which exon 17 is neatly spliced to exon 19 via canonical splice sites (27). The translation reading frame of the resulting mRNA remains intact, suggesting that this mRNA is not susceptible to degradation via the NMD pathway (Table 2). The second variant in this patient, c.1548G>A, generates a termination codon in exon 10 (26). Its effects on mRNA expression have not been reported so far. The premature termination codon is likely to result in low mRNA abundance from this allele.*

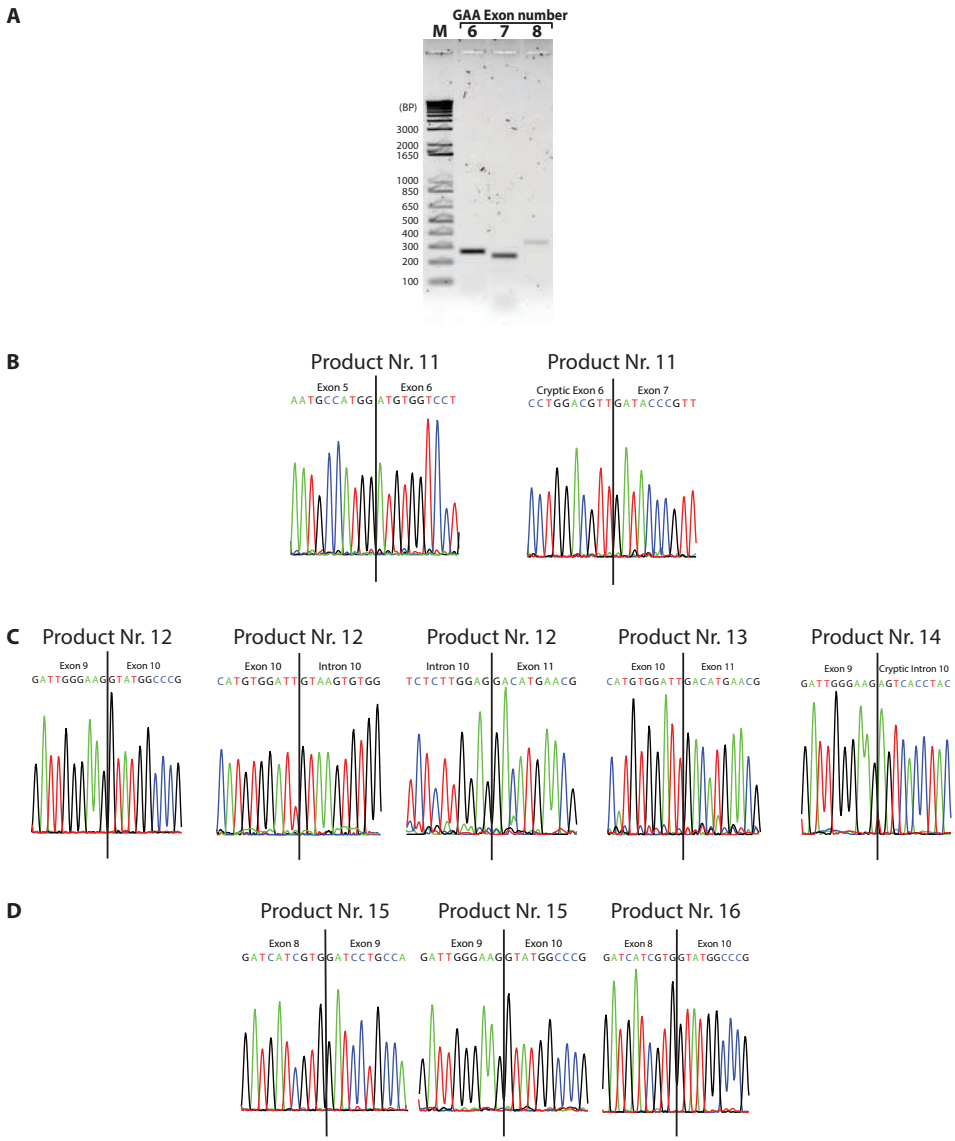
Flanking exon PCR indicated changes for amplification of exons 17, 18, and 19 (panel A). Exon 18 amplification yielded two products instead of one. Sequence analysis indicated that the highest MW product (number 4) represented wild type spliced exon 18, while the lower MW product (number 5) lacked the entire exon 18, and exon 17 and exon 19 were joined via their canonical splice sites (panels B and D). Amplification of exons 17 and 19 yielded lower amounts of the correct products compared to the healthy control. The reverse primer for amplification of exon 17 and the forward primer for amplification of exon 19 both anneal to exon 18, indicating that detection of exons 17 and 19 could not be derived from the delex18 allele but must have come from the c.1548G>A allele. This indicates that the c.1548G>A allele is expressed to some extent, and it explains the detection of low levels of wild type spliced exon 18 by flanking exon PCR.

To quantify expression from the c.1548G>A allele, exon-internal qPCR was performed and indicated 3% expression of exon 18, while all other exons were expressed at ~40-50% of healthy control levels (panel C). This shows that the c.1548G>A variant results in very low mRNA expression, as measured by the low level of exon 18 detection. Expression of all other exons is derived from the del ex18 allele, which produces a stable mRNA in which exon 18 is precisely deleted. The deletion of exon 18 has been shown to result in disruption of enzymatic activity (27). The 3% mRNA expressed from the c.1548G>A allele still carries a premature termination codon and will therefore fail to produce significant levels of wild type GAA protein. Indeed, this patient showed very low GAA enzymatic activity and was diagnosed with classic infantile Pompe disease (Table 1).

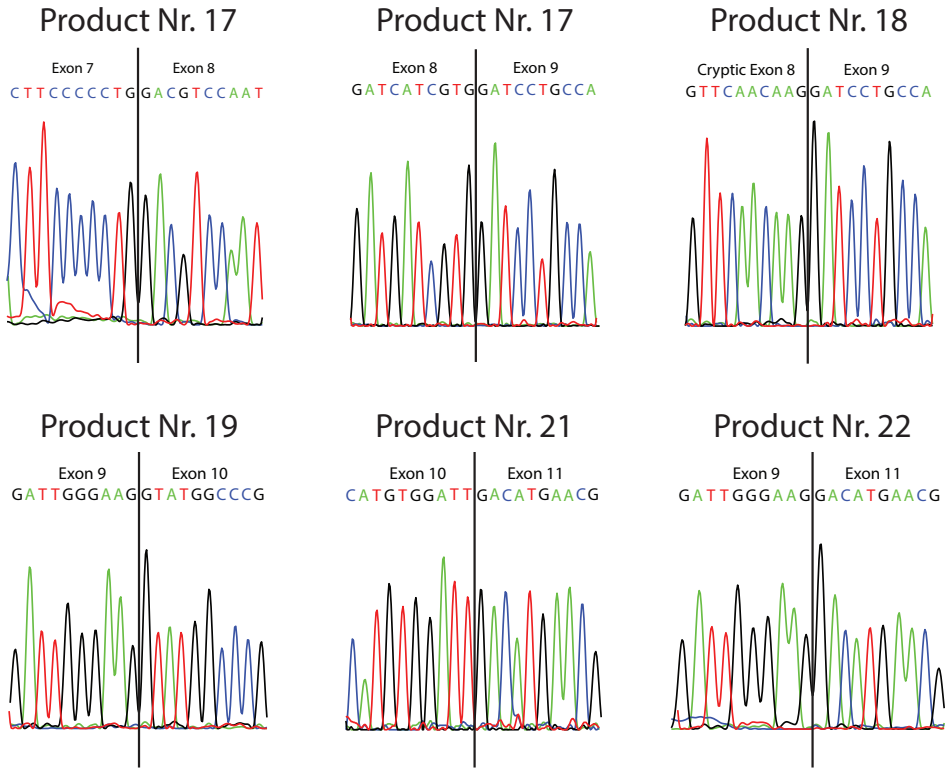
In summary, the generic splicing approach also allows detection and characterization of exonic deletions. A dissection can be made between two alleles by comparing the results of the flanking exon PCR and the exon-internal qPCR assays.



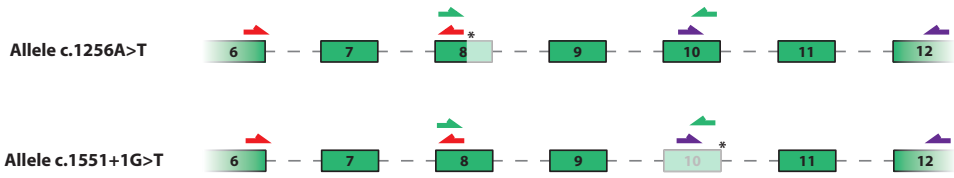
Supplementary Figure S5. Sequence analysis of flanking exon PCR products of patient 4. *Product numbers refer to the numbers indicated next to the bands in Figure 2B.*



Supplementary Figure S6. Further analysis of patient 5 (A-B), patient 6 (C) and patient 7 (D). *A) Flanking exon PCR analysis of patient 5 for exon 7 using an alternative forward primer that anneals to the 5' end of exon 6 (forward qPCR primer for exon 6, see Supp. Table S3) and the normal reverse primer that anneals to exon 8. For comparison, standard flanking exon PCR reactions of exons 6 and 8 are shown. Note that GAA mRNA levels in this patient are low due to NMD. B) Sequence analysis of flanking exon PCR products of patient 5. C) Sequence analysis of flanking exon PCR products of patient 6. D) Sequence analysis of flanking exon PCR products of patient 7. Product numbers refer to the numbers indicated next to the bands in Figure 3A (patient 5), Figure 3C (patient 6) and Figure 3E (patient 7).*



***Supplementary Figure S7. Sequence analysis of patient 8**. Product numbers refer to the numbers indicated next to the bands in Figure 4A.*



***Supplementary Figure S8. Cartoon of exons in patient 8 and the locations of PCR primers used for flanking exon PCR analysis.** Only those primer pairs are shown that anneal to exons affected by the splicing variants.*

SUPPLEMENTARY TABLES

Supplementary Table S1. Settings used for splice predictions

Prediction method	5' donor threshold	Range	3' acceptor threshold	Range	Website
Alamut 2.4.2	-	-	-	-	http://www.interactive-biosoftware.com/doc/alamut-visual/2.4/splicing.html
Splice Site Finder-like	70	[0-100]	70	[0-100]	http://www.interactive-biosoftware.com/doc/alamut-visual/2.4/splicing.html (original site has been removed)
Max Ent Scan	0	[0-12]	0	[0-16]	http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html
NNSPLICE	0,4	[0-1]	0,4	[0-1]	http://www.fruitfly.org/seq_tools/splice.html
Gene Splicer	0	[0-15]	0	[0-15]	http://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml
Human Splicing Finder	65	[0-100]	65	[0-100]	http://www.umd.be/HSF/

Supplementary Table S2. Flanking exon RT-PCR primers used in this study

Target	Forward	Reverse	Full length Product (bp)	Exon size (bp)	product after perfect skipping of exon (bp)
GAA Exon 2	AAACTGAGGCACGGAGCG	GAAGGGCTCCTCGGAGAA	705	578	127
GAA Exon 3	AGCTCCTCTGAAATGGGCTACA	GCAAGGTCCCGGTTCCACA	428	146	282
GAA Exon 4	GCTAACAGGCGCTACGAGGT	TGCTGTTTAGCAGGAACACCC	388	166	222
GAA Exon 5	CTGTTCTTTGCGGACCAGTTC	CCACAACGTCCAGGTACTGCT	361	97	265
GAA Exon 6	GGTCTCACCCTTTCTACCTGG	GTGATAGCGGTGGAGGAGTAG	274	120	154
GAA Exon 7	CAGCAGTACCTGGACGTTGTG	AGTCCATGTAGTCCAGGTCGTT	175	119	56
GAA Exon 8	CGTTCATGCCGCCATACT	GGTCTCGTTGGTGATGAAAAC	335	132	203
GAA Exon 9	GACGTCCAGTGGAACGACCT	ACCTGGTCATGGAACTCAGC	335	111	224
GAA Exon 10	GATCCTGCCATCAGCAGCT	TGGGTTCTCCAGCTCATTGT	297	114	183
GAA Exon 11	AGGACATGGTGGCTGAGTTC	CGTAGAGGTTGTGCAGGTTGTA	228	85	143
GAA Exon 12	AACGAGCCTTCCAACTTCATC	GAGCGGGAGATCACAAATGG	245	118	127
GAA Exon 13	CACCAGTTTCTCTCCACACACTA	GTTCCGCATGAAGGGGTAGA	337	134	202
GAA Exon 14	ACACGCCCATTTGTGATCTC	GTGTAGAGGTGGGGGAGGAGT	356	152	204
GAA Exon 15	AAATCCTGCAGTTTAACCTGCTG	GCAGGTCGTACCATGTGCC	438	149	289
GAA Exon 16	GAGCCGTACAGCTTCAGCGA	ATGTACCCAGCCCGGAGGT	422	142	280
GAA Exon 17	CCTGGACTGTGGACCACCA	CAGGAAGATGACCTGTGTGTAGG	428	150	278
GAA Exon 18	GTGCCAGTAGAGGCCCTTG	GGCTGTAGGTGAAGTTGGAGAC	457	165	292
GAA Exon 19	TCACAACCACAGAGTCCCG	AGAAACTGCTCTCCCATCAACA	352	153	199

Suppplementary Table S3. qPCR primers used in this study

Target	Forward	Reverse	Product size (bp)
ß-Actin	AACCGCGAGAAGATGACCC	GCCAGAGGCGTACAGGGATAG	96
GAA Exon 2	AGCTCCTCTGAAATGGGCTACAC	GGTTCTCAGTCTCCATCATCACG	109
GAA Exon 3	ATCCAGCTAACAGGCGCTAC	GCTCCTCGGAGAACTCCAC	96
GAA Exon 4	CTGTTCTTTGCGGACCAGTT	CTGAGCATCAGGGGACTGAG	95
GAA Exon 5	CGAACCTCTACGGGTCTCAC	TGCTGTTTAGCAGGAACACC	81
GAA Exon 6	CTTAGCTGGAGGTCGACAGG	CACAACGTCCAGGTACTGCT	93
GAA Exon 7	CGTTCATGCCGCCATACT	GGTCATGTTCTCCACCACCT	95
GAA Exon 8	GACGTCCAGTGGAACGACCT	GAAGTCCCGGAAGCCATC	78
GAA Exon 9	ATCCTGCCATCAGCAGCTC	GGTCTCGTTGGTGATGAAAA	89
GAA Exon 10	CACTGCCTTCCCCGACTT	ACCTGGTCATGGAACTCAGC	78
GAA Exon 11	ACATGAACGAGCCTTCCAAC	ACGTAGGGTGGGTTCTCCAG	79
GAA Exon 12	CCTCCAGCCACCAGTTTCTCT	TGTGGGAGGCGATGGCTT	78
GAA Exon 13	GACACGCCCATTTGTGATCT	CCAGGAGCTCCACACGTC	88
GAA Exon 14	CTCAGAGGAGCTGTGTGTGC	CAGACTGAGCAGGCTGTTGT	82
GAA Exon 15	CAGCAGGCCATGAGGAAG	GGCCTGGTGGAACAGTGTG	75
GAA Exon 16	CCCAAGGACTCTAGCACCTG	CAAGGGGAAGTAGCCAGTCA	114
GAA Exon 17	GTGCCAGTAGAGGCCCTTG	GAGGTGGACGTTGATGGTGT	123
GAA Exon 18	GCCTCACAACCACAGAGTCC	TCTCTCCATCGTCCCAGAAC	102
GAA Exon 19	TGCAGAAGGTGACTGTCCTG	GGGCTGTAGGTGAAGTTGGA	88
GAA Exon 20	GGGCGGAGTGTGTTAGTCTC	CTCCAGGTGACACATGCAAC	110
GAA N	AAACTGAGGCACGGAGCG	GAGTGCAGCGGTTGCCAA	129
GAA SV2	GGCACGGAGCGGGACA	CTGTTAGCTGGATCTTTGATCGTG	92
GAA SV3	AGGCACGGAGCGGGATCA	TCGGAGAACTCCACGCTGTA	111

Chapter 3

Inhibition of mRNA Decay for Unbiased Detection of Aberrant Splicing in Pompe Disease

Atze J. Bergsma[1,2,3]
Stijn L.M. in 't Groen[1,2,3]
Manjiro Yamanaka[1,4]
Satoru Takahashi[5]
Toshika Okumiya[6]
Ans T. van der Ploeg[2,3]
W.W.M. Pim Pijnappel[1,2,3]

[1] Molecular Stem Cell Biology, Department of Clinical Genetics, Erasmus Medical Center, 3015 GE Rotterdam, Netherlands
[2] Department of Pediatrics, Erasmus Medical Center, 3015 GE Rotterdam, Netherlands
[3] Center for Lysosomal and Metabolic Diseases, Erasmus Medical Center, 3015 GE Rotterdam, Netherlands
[4] Department of Laboratory Medicine, Shinshu University Hospital, 3-1-1 Asahi, Matsumoto, Nagano 390-8621, Japan
[5] Department of Pediatrics, Asahikawa Medical University, 2-1-1-1 Midorigaoka Higashi, Asahikawa, Hokkaido 078-8510, Japan
[6] Department of Biomedical Laboratory Sciences, Faculty of Life Sciences, Kumamoto University, 4-24-1 Kuhonji, Chuo-ku, Kumamoto 862-0976, Japan

Manuscript in preparation

ABSTRACT

Diagnostic Sanger sequencing of genomic DNA is commonly used for determination of the pathogenic variants which cause monogenic disorders. The identification of pathogenic variants present in a patient is important for estimation of severity of the disease, as well as for genetic counseling and prediction of efficacy of treatment. However, analysis of genomic DNA does not allow to investigate the functional effect of these variants. A recent report describing a method for analysis of aberrant pre-mRNA splicing in primary patient fibroblasts presents a promising tool for functional investigation of any intragenic variant. However, variants which effect the reading frame, and thus induce nonsense-mediated mRNA decay, are difficult to analyze. We investigate if blockage of mRNA degradation in primary fibroblast cultures of a healthy control and four Pompe patients with suspected splicing variants leads to better identification of splice products. Analysis of pre-mRNA splicing after cycloheximide treatment indicates that aberrant splicing of *GAA* pre-mRNA transcripts occurs naturally, regardless of the presence of variants. Aberrant splicing events that could be linked to specific pathogenic variants could be identified in all patients. Moreover, antisense oligonucleotide mediated blockage of a cryptic splice site utilized in the context of the c.546G>T variant resulted in a threefold increase of GAA enzymatic activity, indicating a potential treatment strategy for patients that carry this variant. These data indicate that blockage of nonsense-mediated decay can be a valuable addition to the previously described method for the identification of aberrant splicing events, and can result in the identification of potential drug targets.

INTRODUCTION

Pompe disease (Glycogen storage disease type II, OMIM 232300) is an autosomal inherited recessive disorder which is caused by the partial or complete lack of activity of the enzyme acid α-glucosidase (GAA) (1). Loss of GAA activity results in accumulation lysosomal glycogen, ultimately leading to cell death. The disease mainly effects skeletal muscle tissue, leading to loss of muscle strength. Diagnosis of the disease is established by measurements of residual GAA enzymatic activity in primary fibroblasts, muscle biopsies and/or leukocytes. Genomic DNA is analyzed to identify the genetic basis of the disease by determination of the pathogenic variants present on both alleles of the *GAA* gene. This method has been used for more than 20 years and has resulted in the establishment of a database describing more than 500 *GAA* variants (www.pompecenter.nl). Determination of the pathogenic variants present in Pompe patients is important to predict severity of disease progression, as well as for genetic counseling. Furthermore, the efficacy of ERT treatment can be estimated based on these findings. The effect of each variant on residual GAA enzymatic activity is scored based on multiple aspects, which include comparison to clinical outcome of patients carrying the same variant, functional analysis of missense variants by introduction into a *GAA* cDNA construct, and *in silico* prediction of the effect of the variant on protein activity and/or pre-mRNA splicing. However, these assays do not examine the functional effect on transcription and splicing of the *GAA* mRNA products. Recently, a method was described which analyzes the functional effect of specific variants on pre-mRNA splicing and mRNA expression in primary patient fibroblasts (2). This method can be used to determine the functional effect that coding and non-coding variants have on pre-mRNA splicing and expression. Furthermore, the splicing assay can be used to identify pathogenic variants which are missed in routine diagnostic screening because of their deep-intronic location. However, the assay has one restriction: Analysis of variants which alter canonical splicing can be complicated due the effect of nonsense-mediated decay (NMD). Roughly two-thirds of aberrantly spliced mRNA transcripts are likely to be degraded by NMD due to a shift in the reading frame, making them difficult to identify. A way to circumvent this problem is to treat primary patient fibroblasts with cycloheximide. Cycloheximide inhibits protein translation by binding to the E-site of the ribosome (3). The NMD pathway scans mRNA's for premature termination codons during the first round of translation, thus cycloheximide inhibits RNA degradation by NMD (4). Here, we show that cycloheximide can aid in the identification of aberrant splice products. The splicing assay was performed on fibroblasts with or without cycloheximide treatment from one healthy control and four Pompe patients. *GAA* mRNA was screened for aberrant splicing events which undergo NMD. Eleven new splice events could be identified, from which only one could be detected without cycloheximide treatment. Interestingly, two aberrant splicing events were observed in all cell lines, indicating these variants occur naturally in a wild-type situation. Aberrant splice events specific to the variants present in each patient were

observed for all four patients. One of these events, which constitutes aberrant splicing caused by the common c.1927G>A missense variant, has not been described before. Furthermore, aberrant splice species caused by the frequent Asian c.546G>T variant have been identified. Blockage of the cryptic splice site utilized in these transcripts with an antisense oligonucleotide (AON) resulted in a threefold increase of GAA enzymatic activity. These data indicate that blockage of NMD with cycloheximide complements the splicing assay by preventing decay of out of frame mRNA products. Moreover, cryptic splice sites identified using the assay can be utilized as targets for AONs, which could ultimately lead to antisense based therapies for patients which carry pathogenic variants that alter splicing.

RESULTS

GAA transcripts in healthy control primary fibroblasts are partially degraded by nonsense-mediated decay

Analysis of splicing using the splicing assay is initialized by performing a flanking exon RT-PCR of all spliced exons on cDNA obtained from primary fibroblasts. For identification of abnormal splicing events, all product sizes and sequences obtained from patient material are compared to the product sizes and sequences of a healthy control. Any splice variant that is not identified in the healthy control is considered to be aberrant. For this reason, as a benchmark, primary fibroblasts from a healthy control were analyzed to determine the pattern of splicing seen with and without blockage of NMD. Enzymatic activity of these cells was determined to be in the middle of the control range (between 40-180 nmol 4-MU/mg protein/hr) at 122.4 nmol 4-MU/mg protein/hr. Flanking exon RT-PCR analysis of untreated healthy control showed faint products of larger than normal size in the flanking exon PCR samples of *GAA* exon 6 and 7 (Figure 1A, black asterisk). Due to the size difference between the two products, we hypothesized that this product was caused by the retention of intron 6. Intron 6 retention leads to nonsense-mediated decay due to a shift in the reading frame. As a consequence, this splice variant would be difficult to identify. kTo investigate this hypothesis, primary fibroblasts of a healthy control were treated with cycloheximide to block NMD. Flanking exon RT-PCR of all spliced exons revealed novel splice products present in samples from exons 6, 7, 12 and 13 (Figure 1B). Subsequent sequencing of the products present in these PCR samples shows that products 1 and 3 retain *GAA* intron 6, causing NMD of these mRNA species (Figure 1C, Suppl. Figure S1A). Products 2 and 4 are normally spliced products, and thus do not undergo NMD. To examine if retention of intron 6 is caused by the presence of weak splice donor and acceptor sequences on both sides of the intron, a splice prediction was performed on the splice sites flanking *GAA* intron 6 (Suppl. Figure S1B). The exon 6 donor splice site and the exon 7 acceptor splice site were both recognized by 4 of 5 splice prediction algorithms with an average value of 43% and 47%, respectively. Average values for splice donor and splice acceptor strength in the

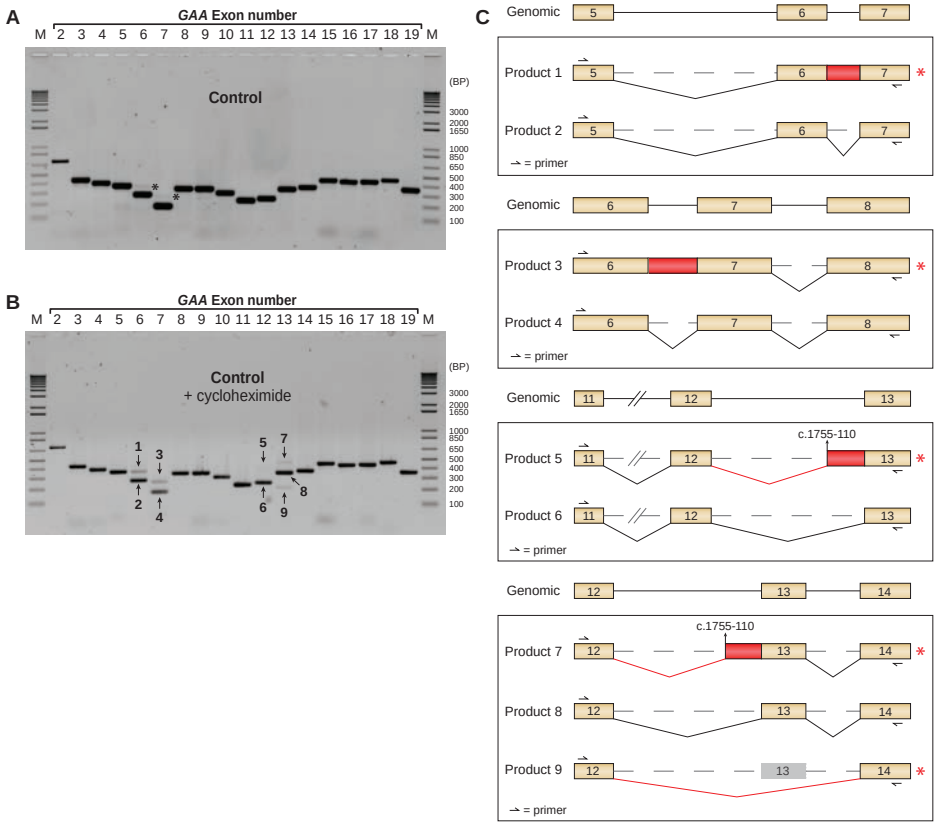


Figure 1. Identification of new splice products after blockage of nonsense mediated decay in fibroblasts from a healthy control. *A) Flanking exon RT-PCR on all spliced GAA exons in primary fibroblasts from a health control. Black asterisks indicate unidentified aberrant splice products. B) Same as in A, but after treatment with cycloheximide. C) Cartoons depicting all products identified after sequencing of TOPO clones generated from the PCR samples of exon 6, 7, 12 and 13. Product numbers refer to the numbers shown in B. Red asterisks indicate products which undergo nonsense mediated decay.*

GAA gene are 67% and 70%, respectively, indicating that the predicted lower strength of the splice sites surrounding intron 6 might be the cause of intron 6 retention in part of the *GAA* mRNA species. Another abnormal splicing event was identified in the flanking exon PCR samples of *GAA* exon 12 and 13. Products 5 and 7 seemed to be of slightly larger size than the canonical product. Sequencing analysis showed that these products were generated by the utilization of a cryptic splice site residing within intron 12 (c.1755-110) (Figure 1C and Suppl. Figure S1C). However, several other splice acceptor sites were identified in between the utilized -110 site and the canonical splice site, indicating that prediction results are difficult to interpret without functional insight (Suppl. Figure S1D). Product 6 and 8 were conventional splice products. One additional product was detected in the flanking exon PCR for exon 13, which was of smaller size compared to the normal splice product. Sequencing analysis showed that this product was the full skip of exon 13

(Suppl. Figure S1E). The prediction of exon 13 splice site strength shows that the splice acceptor strength is average (65%) and that the donor site is high (93%) (Suppl. Figure S1D and S1F). However, the presence of the product which completely skips exon 13 (product 9) demonstrates that predictive values do not always correlate with utilization of specific splice sites. These data indicate that the process of splicing of *GAA* pre-mRNA does not always result in the correct splice form, and can lead to decay of RNA products. Although the extend of expression of aberrant mRNA splicing in the *GAA* transcripts is difficult to predict, these data indicate that a significant amount of *GAA* mRNA products undergo alternative pre-mRNA splicing which results in degradation through the NMD pathway.

Analysis of the effect on splicing of the c.2481+2T>C variant in patient 1

A patient which was diagnosed with Pompe disease presented the disease due to the presence of two previously identified pathogenic variants. The onset of this patient (patient 1, Table 1) was infantile, and residual *GAA* enzymatic activity was barely present at 1.6 nmol 4-MU/mg protein/hr. Allele one harbored the c.1726G>C missense variant which leads to a non-functional protein. The c.2481+2T>C variant present on the other allele is a splicing

Table 1. Patient characteristics.

patient nr.	*GAA* variant on first Allele	*GAA* variant on second Allele	GAA enzymatic activity	Age at diagnosis	Onset
Control	-	-	122,4	-	-
1	c.1726G>C	c.2481+2T>C	1,6	7 months	infantile
2	c.-32-13T>G	c.2331+2T>A	9,1	11 years	juvenile
3	c.1461insC	c.1927G>A	1,5		infantile
4	c.546G>T	c.1798C>T	3,7		-

variant which has not been functionally characterized before, but also shows a severe phenotype. For closer examination of the effect on splicing of the c.2481+2T>C variant, conventional flanking exon RT-PCR was performed on the *GAA* gene (Figure 2A). Aberrant splicing was visually identified in the flanking exon PCR sample of *GAA* exon 17, with the presence of the normal product (product 10) and an aberrant shorter product (product 11). Due to the low abundance of the aberrant product, direct sequencing analysis of all exons did not reveal any aberrant splicing events. TOPO cloning of the PCR sample of exon 17 lead to the identification of the aberrant splice product, which was a complete skip of exon 17 (Figure 2B, Suppl. Figure S2A). The skipping of exon 17 resulted in the loss of 150 bases from the canonical *GAA* mRNA sequence, but did not result in a frameshift, which explains why the product is visible on the gel. Although aberrant splicing was identified, the product could not be the major cause of phenotype in the patient, due to its small contribution visible in the PCR sample. To identify if RNA decay was present, a quantitative exon internal RT-qPCR was performed on all coding exons. With an average value of 21,2% of *GAA* expression compared to the expression in a healthy control, it seemed evident that

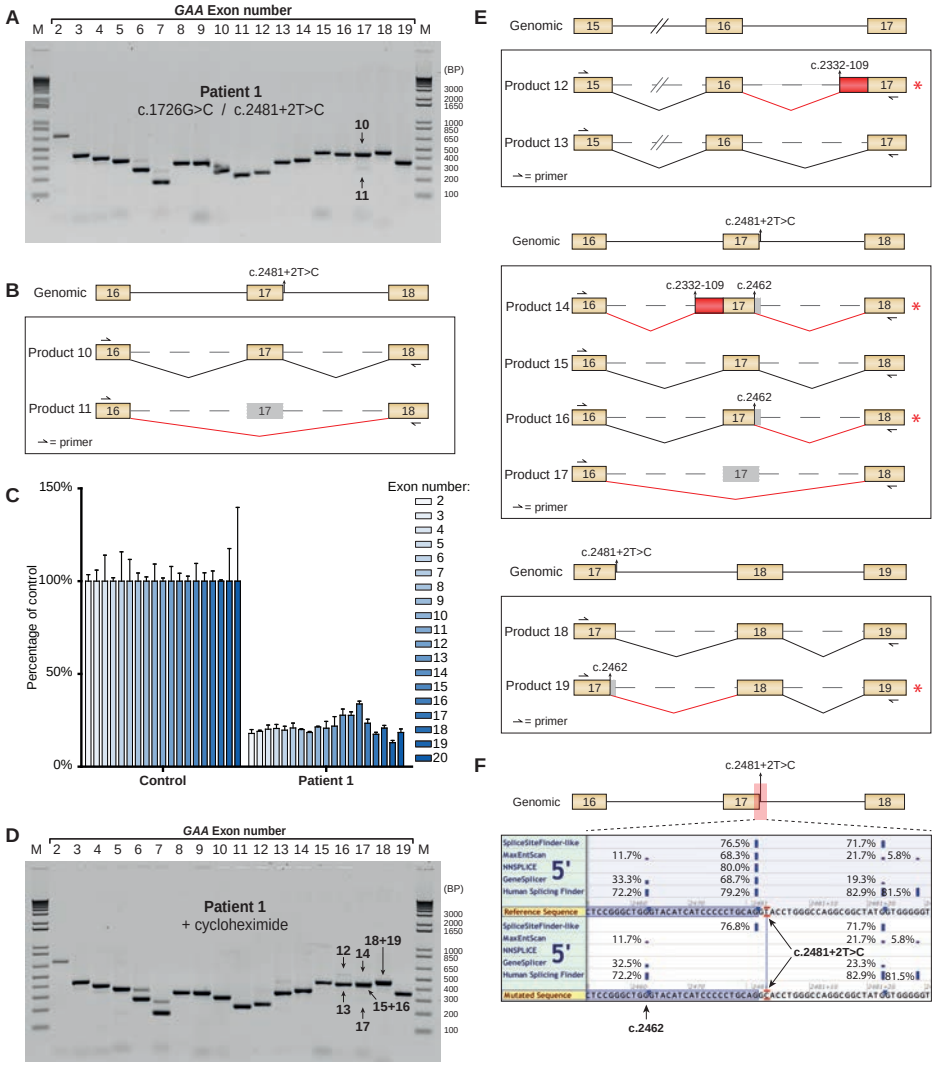


Figure 2. Analysis of aberrant splicing products in patient 1*. A) Flanking exon RT-PCR of patient 1 primary fibroblasts. B) Cartoon of products identified after TOPO cloning and sequencing of specific products from the flanking exon RT-PCR of exon 17. Product numbers refer to products highlighted in A. Red lines indicate aberrant splicing events. Gray area's indicate skipping of canonical GAA mRNA sequence. C) exon-internal RT-qPCR for all coding exons of patient 1 compared to healthy control. Data represents mean +/- S.D. from three technical replicates. D) Same as in A, but after cycloheximide treatment. E) Cartoons of products present in PCR samples for exons 16, 17 and 18. Product numbers refer to product highlighted in D. Red asterisks indicate products undergoing nonsense mediated decay. Red boxes are intronic sequences present in the GAA mRNA transcripts. F) Splice prediction run in Alamut. Five splice site prediction algorithms are shown for the area indicated in red, with or without the presence of the c.2481+2T>C variant. The location of the cryptic donor site used in products 14, 16 and 19 is highlighted.*

low promotor activity and/or RNA-decay was present (Figure 2C). To identify the presence of products undergoing NMD, cycloheximide treatment was performed. mRNA analysis of the treated fibroblasts by flanking exon RT-PCR shows that aberrant splicing events are present in the PCR samples of exon 6, 7, 12 and 13 (Figure 2D). These products correlate

to the same products identified in the healthy control shown above. Furthermore, aberrant products were present for exon 16, 17 and 18. Three aberrant splicing events could be identified after TOPO cloning and subsequent sequence analysis of the products. The flanking exon PCR for exon 16 indicates the presence of a larger product (product 12). The product was determined to be aberrant splicing of intron 16 as sequencing analysis showed utilization of a splice acceptor located at c.2332-109 (Figure 2E , Suppl. Figure S2B). This event was somewhat unexpected, because the investigated variant is not located at the affected exon 17 splice acceptor site. The introduced sequence contains a stop codon, which induces NMD, which suggests why it could not be detected without cycloheximide treatment. Sequencing of the aberrant products present in the exon 17 PCR resulted in identification of product 14. This product also utilizes the acceptor site that was seen in product 12 (c.2332-109). Additionally , a donor site located within exon 17 at c.2462, 19 bases short of the canonical exon 17 donor splice site, is utilized (Suppl. Figure S2B). Product 16 and 19 were identified to be a shorter than normal mRNA species that also utilize the cryptic donor splice at c.2462. From there it splices toward the canonical exon 18 acceptor site. These products induce a frameshift and thus undergo NMD, which explains why they could only be detected after cycloheximide treatment. The third aberrant product (product 17) was identified to be the same product as product 11, which is a complete skip of exon 17. Products 13, 15 and 18 were products that were canonically spliced. To see if the aberrant splice events found in the flanking exon RT-PCR assay could be predicted, splice prediction with Alamut was performed. Analysis of the effect of the c.2481+2T>C variant shows that when the variant is present, four out of five prediction algorithms do not predict the presence of the canonical exon 17 splice donor site, indicating a potentially severe effect on splicing (Figure 2F). Subsequent usage of the c.2462 cryptic splice site is also supported by the splice prediction, as three algorithms predict the presence of the cryptic splice site at this location. Although this cryptic site could be predicted, another cryptic donor splice site present at c.2481+23 was also identified, but no aberrant splice products utilizing this splice site were identified. Furthermore, prediction analysis of the splice sites utilized for the generation of products 12, 16 and 19 shows that all cryptic splice sites were identified (Suppl. Figure S2C). These data indicate that the prediction algorithms used in the Alamut program are useful for predicting the effect of a specific variant on the strength of a canonical splice site. However, functional implications of this effect on splicing are more difficult to predict, which implies that functional investigation still needs to be performed to identify the severity of the variant.

Analysis of splicing in primary fibroblasts from patient 2

A second Pompe patient which was analyzed with the splicing assay was diagnosed at the age of 11 years. Diagnostic measurement of GAA enzymatic activity in primary fibroblasts of this patient showed a residual activity of 9.1 nmol 4-MU/mg protein/

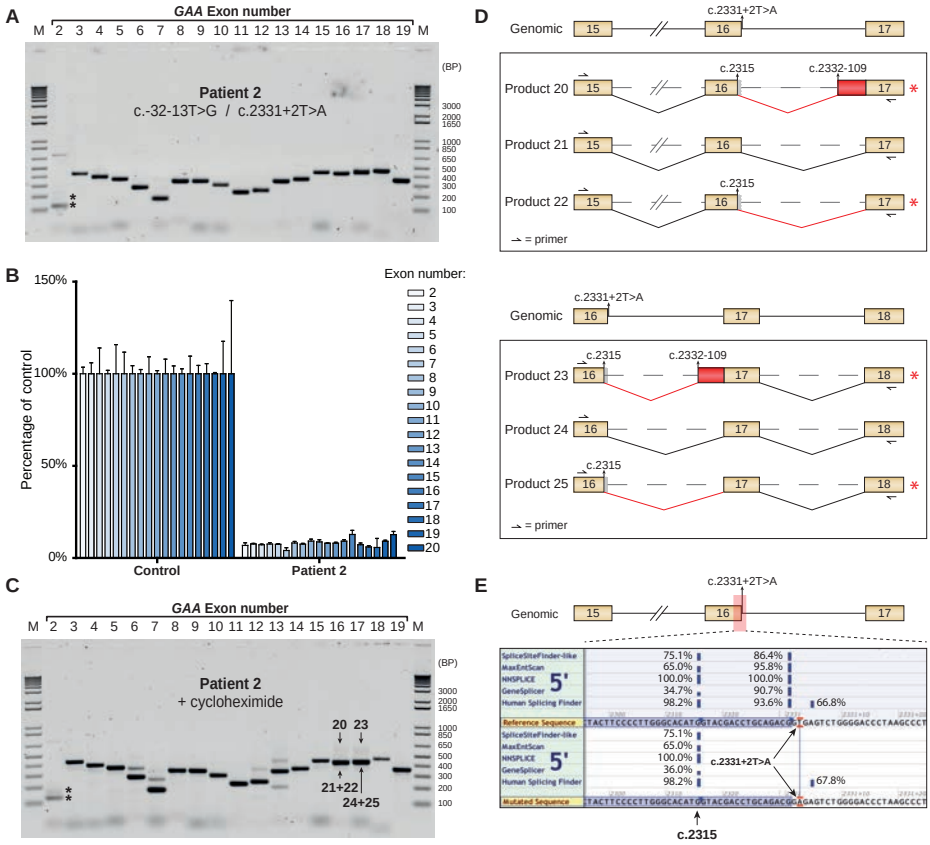


Figure 3. Analysis of aberrant splicing products in patient 2*. A) Flanking exon RT-PCR of patient 2 primary fibroblasts. Black asterisks indicate aberrant products caused by the c.32-13T>G variant. B) exon-internal RT-qPCR for all coding exons of patient 2 compared to healthy control. Data represents mean +/- S.D. from three technical replicates. C) Same as in A, but after cycloheximide treatment. D) Cartoons of products present in PCR samples for exons 16 and 17. Product numbers refer to product highlighted in C. Red asterisks indicate products undergoing nonsense mediated decay. Red boxes indicate non-canonical sequences present in the GAA mRNA transcripts. Gray area's highlight skipping of canonical GAA mRNA sequence. E) Splice prediction run in Alamut. Five splice site prediction algorithms are shown for the area indicated in red, with or without the presence of the c.2331+2T>A variant. The location of the cryptic donor site utilized in products 20, 22, 23 and 25 is highlighted.*

hr, consistent with a juvenile onset of the disease. Genomic DNA analysis detected the presence of two compound heterozygous variants. One allele harbored the well-known c.32-13T>G (IVS1) variant, which causes aberrant splicing of *GAA* exon 2 (5-7). A new variant, c.2331+2T>A, was detected on the second allele. This variant was thought to affect splicing of *GAA* exon 16 because of its close proximity to the canonical exon 16 donor splice site. To analyze if aberrant splicing was present, patient fibroblasts were grown and analysis of splicing was performed using the splicing assay. Flanking exon RT-PCR analysis of all exons demonstrated the presence of the c.32-13T>G variant by the identification of multiple aberrant products in the PCR sample for *GAA* exon 2 (Figure 3A, indicated with black asterisks). The size of these products were consistent with previous

reports describing the effect of the c.32-13T>G variant. Apart from these aberrant splice products, no other aberrant products could be detected after sequencing of all PCR samples. Next, *GAA* expression levels were determined using exon internal RT-qPCR for all coding exons. Expression data indicated that residual *GAA* expression was only 8.0%, suggesting the presence of low promoter activity and/or RNA decay (Figure 3B). For functional analysis of the effect of the c.2331+T>A, primary fibroblasts of patient 2 were treated with cycloheximide. Flanking exon RT-PCR was performed again and showed the presence of aberrant products at the expected locations (exons 6, 7, 12 and 13) but also in PCR samples for exon 16 and 17 (Figure 3C). A larger product was seen in both samples indicated by product 20 and 23. Sequencing analysis of TOPO clones from exon 16 and 17 PCR samples indicate that these products utilize a cryptic splice donor at c.2315 which splices towards a cryptic splice acceptor at c.2332-109, producing an out of frame product that is degraded by NMD (Suppl. Figure S3). Furthermore, the canonical splice products (products 21 and 24) show a band which is unusually thick, indicating that these samples likely contain two PCR products. Sequencing analysis confirmed the presence of a splice product which utilizes the cryptic splice donor site at c.2315, causing the generation of mRNA species that skip the last 16 base pairs of exon 16. From this site, it directly splices towards the splice acceptor site of exon 17 (products 22 and 25) (Suppl. Figure S3). This is in agreement with the presence of the ticker band seen for canonical splicing of exon 16 and 17, which actually consist of two products each (product 21 and 22 for exon 16, and product 24 and 25 for exon 17). Splicing prediction of the c.2331+2T>A variant implies that recognition of the exon 16 splice donor site is completely lost in the presence of this variant (Figure 3E). Both cryptic splice sites utilized in the presence of the variant could also be recognized by splice prediction algorithms (Suppl. Figure S2C). Based on the low GAA enzymatic activity measured for this patient and the presence of the relatively mild c.32-13T>G on the other allele, the c.2331+2T>A is expected to be severe in nature.

Analysis of splicing in primary fibroblasts from patient 3

The splicing assay was performed on a third patient for the purpose of identification of an unknown variant on the second allele. This patient presented with the severe infantile form of Pompe disease, with a residual GAA enzymatic activity of less than 2.0 nmol 4-MU/mg protein/hr. However, diagnosis for this patient was carried out in 1982, suggesting that a variant could be missed because of technical restraints. The variant on one allele was identified and harbored the c.1927G>A missense variant. It should be noted that this variant is labeled as a potentially less severe in the pompecenter.nl database, which is in contradiction with the infantile onset in this patient. The variant is frequently observed in Pompe patients (8). Splicing analysis was performed on primary fibroblasts from this patient. Apart from the presence somewhat high presence of products retaining intron 6, no qualitative differences were seen in the flanking exon RT-PCR analysis compared to the

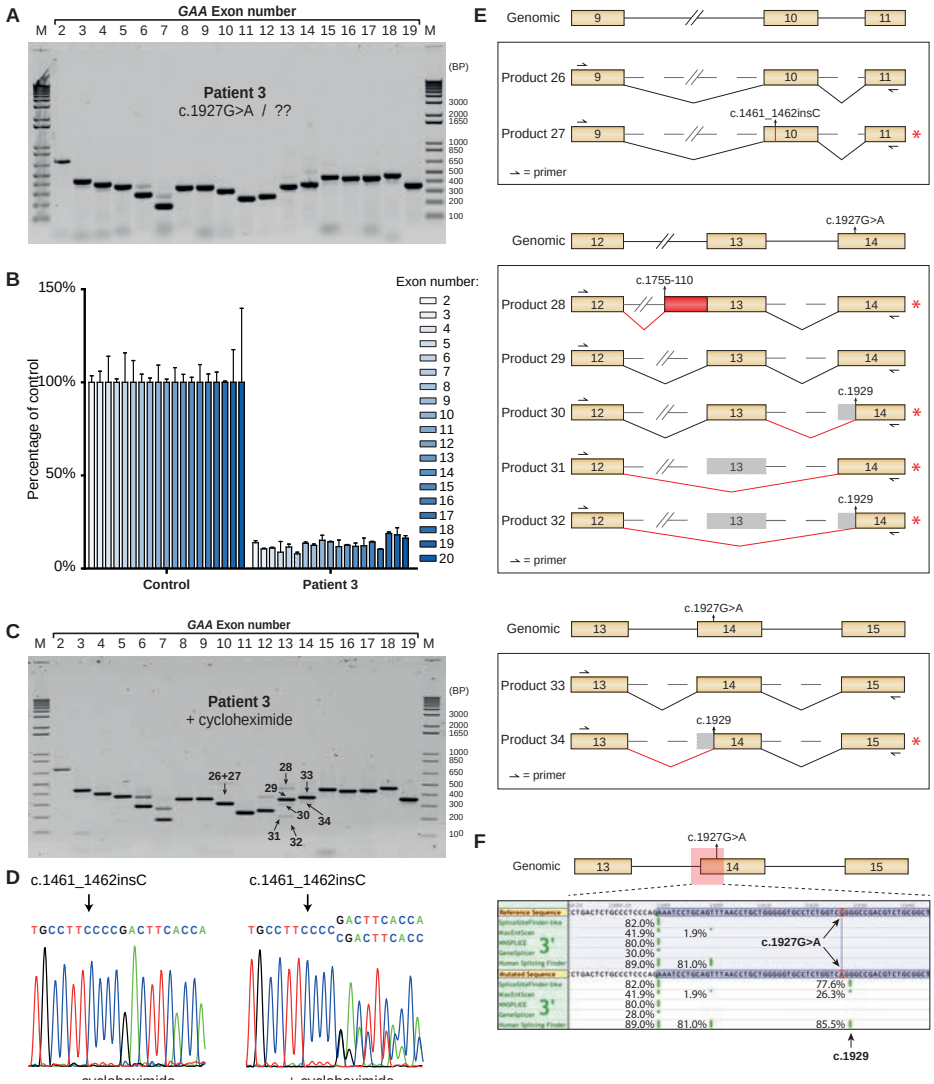


Figure 4. Analysis of aberrant splicing products in patient 3. *A) Flanking exon RT-PCR of patient 3 primary fibroblasts. B) exon-internal RT-qPCR for all coding exons of patient 3 compared to healthy control. Data represents mean +/- S.D. from three technical replicates. C) Same as in A, but after cycloheximide treatment. D) Sequencing results of flanking exon RT-PCR exon 10 sample with and without cycloheximide treatment. E) Cartoons of products present in PCR samples for exons 10, 13 and 14. Product numbers refer to product highlighted in C. Red asterisks indicate products undergoing nonsense mediated decay. Red boxes indicate non-canonical sequences present in the GAA mRNA transcripts. Gray area's highlight skipping of canonical GAA mRNA sequence. F) Splice prediction run in Alamut. Five splice site prediction algorithms are shown for the area indicated in red, with or without the presence of the c.1927G>A variant. The location of the cryptic donor site utilized in products 30, 32 and 34 is highlighted.*

healthy control. Direct sequencing of all products also showed no abnormal splicing events. Quantitative analysis was then performed by exon internal RT-qPCR. Results indicate an average expression level for all coding exons of 12% compared to healthy control (Figure 4B). This suggests that both alleles produce reduced levels of mRNA, either through RNA

decay or reduced promoter activity. For identification of products degraded by NMD, fibroblasts from patient 3 were subjected to treatment with cycloheximide. Flanking exon RT-PCR and subsequent sequencing analysis shows the presence of multiple aberrant products. The first aberrant product, highlighted as product 27 in Figure 4C, contained an insertion of one cytosine, indicated as c.1461_1462insC, and was only seen after cycloheximide treatment as indicated by the sequencing results (Figure 4D and 4E). The presence of this variant resolves the issue of the variant on the second allele. However, multiple other splice variants were identified in PCR samples for exon 13 and 14. TOPO cloning and subsequent sequencing analysis of these variants indicated that products 31, 33 and 35 utilize a cryptic splice acceptor site at c. 1929, close to the c.1927G>A variant present on allele one. Products 28 and 31 are the aberrant splice products identified in all patients and controls (Figure 4E, Suppl. Figure S4). Splice prediction with Alamut indicates that the c.1927G>A variant generates a cryptic splice site at the c.1929 site (Figure 4F), further supporting the fact that the c.1927G>A variant does not only act as a missense variant, but also has an effect on splicing. These findings indicate that the splicing assay complemented with cycloheximide treatment is able to identify variants that shift the reading frame, while at the same time functionally investigates the effect of certain variants on splicing.

Analysis of splicing in primary fibroblasts from patient 4

The last patient studied using the splicing assay was a Japanese patient with a relatively low, but not absent, residual GAA enzymatic activity of 3.7 nmol 4-MU/mg protein/hr. Diagnostic sequencing analysis identified two pathogenic variants. On one allele, the c.546G>T variant was present. This synonymous variant is located at the last base of *GAA* exon 2, and is known to cause aberrant splicing (9). The c.1798C>T pathogenic missense variant was identified on the second allele. This missense variant causes a p.Arg600Cys amino acid change and is indicated as less severe in the Pompe database. The variant frequently occurs in Japanese Pompe patients (10). Splicing analysis was performed on cDNA of patient 4 fibroblasts by flanking exon RT-PCR of all spliced *GAA* exons. Interestingly, very faint aberrant products for the exon 2 PCR could be observed in the gel (Figure 5A). TOPO cloning and subsequent sequencing analysis resulted in the identification of three distinct products. Product 35 was canonical splicing of exon 2, whereas products 36 and 37 were aberrant products identical to the transcripts identified when the common c.-32-13T>G variant is present (Figure 5B, Suppl. Figure S5A) (5-7). No other aberrant splice products could be identified. Next, exon internal RT-qPCR was performed. Results indicate that mRNA expression is on average 59,1% compared to healthy control levels (Figure 5C). These mRNA levels are consistent with the presence of one allele that contains the c.1798C>T missense variant, and thus has no abnormal transcription levels, and the presence of an allele which shows abnormal pre-mRNA splicing caused by the c.546G>T

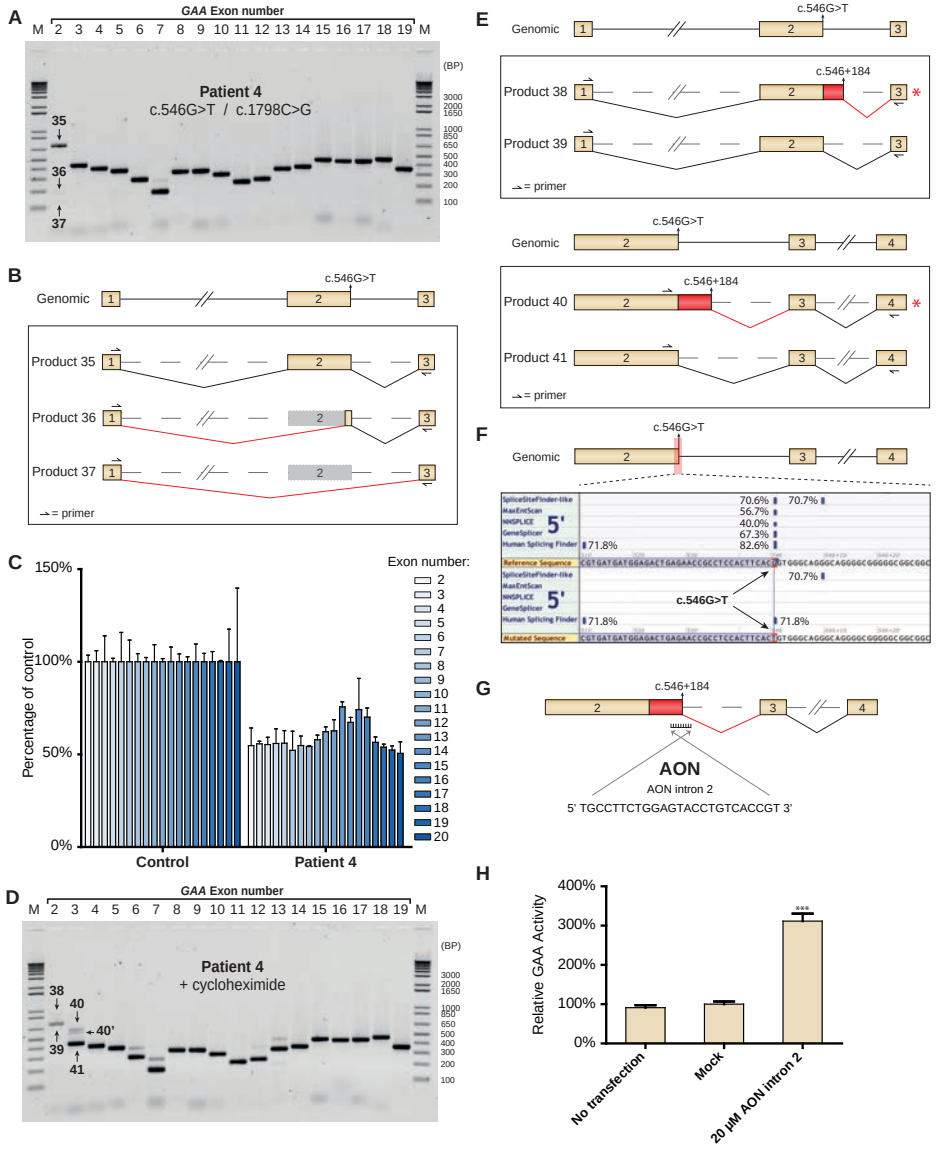


Figure 5. Analysis of aberrant splicing products in patient 4. *A) Flanking exon RT-PCR of patient 4 primary fibroblasts. B) Cartoons of products identified to be present in PCR samples for exon 2. Product numbers refer to product highlighted in A. C) exon-internal RT-qPCR for all coding exons of patient 4 compared to healthy control. Data represents mean +/- S.D. from three technical replicates. D) Same as in A, but after cycloheximide treatment. Product 40' indicates the same product as product 40 E) Cartoons of products present in PCR samples for exon 3. Product numbers refer to product highlighted in D. Red asterisks indicate products undergoing nonsense mediated decay. Red boxes indicate non-canonical sequences present in the GAA mRNA transcripts. F) Splice prediction run in Alamut. Five splice site prediction algorithms are shown for the area indicated in red, with or without the presence of the c.546G>T variant. The location of the cryptic donor site utilized in product 33 is highlighted. G) Cartoon depicting the target site and sequence of AON intron 2. H) GAA enzymatic activity of patient 4 fibroblasts after transfection with AON intron 2. Data are normalized for mock transfection and represent the mean +/- SD of three biological replicates (*** p = 0,001).*

variant. The low levels of aberrant exon 2 transcripts could indicate that other aberrant splice products which undergo NMD are generated as a consequence of the c.546G>T. To determine whether this is the case, primary fibroblast from patient 4 were treated with cycloheximide. Flanking exon RT-PCR analysis of all spliced *GAA* exons indicates that aberrant splice products are indeed present (Figure 5C, products 38 and 40). TOPO cloning and sequencing analysis of exon 2 and 3 PCR samples shows that an aberrant transcript is generated which does not utilize the canonical c.546 exon 2 donor site, a donor site located 184 bases downstream of this site in intron 2 (Figure 5E, Suppl. Figure S5A). This product undergoes nonsense-mediated decay due to frameshift caused by the insertion. The products observed in the gel which correspond to this transcript are highlighted as product 38 and 40. Product 40' indicates a product for which no sequence could be identified. Further investigation for identification of this product showed that it represents a conformational variant of product 40 caused by rapid cooling of the PCR product (Suppl. Figure S5B). Interestingly, the aberrant products identified in flanking exon RT-PCR analysis without cycloheximide treatment could not be detected. *In silico* investigation of the effect of the c.546G>T variant on the predicted strength of the exon 2 donor site indicates that recognition of the splice site is strong in wildtype genotype, but becomes poorly recognized if the T is present (Figure 5F). This shows that the variant likely has an effect on splicing, although the extend of aberrant splicing is difficult to predict. The presence of cryptic splice sites identified using the functional assay could also be identified using Alamut (Suppl. Figure S5C). We hypothesized that blockage of the cryptic splice site at c.546+184 with the use of an AON could at least in part restore canonical exon 2 splicing from the c.546G>T allele. In order to test this, an AON overlapping the cryptic splice site was designed (Figure 5G). Treatment of primary fibroblasts from patient 4 with this AON resulted in a 3-fold increase of enzymatic activity (Figure 5H). These data indicate that antisense treatment would be a potential new type of therapy for patients with this particular variant.

DISCUSSION

The functional consequences that specific variants have on pre-mRNA splicing are often overlooked in current diagnostics of monogenic disorders. Recently, a method was described which could identify the functional effect of variants which alter splicing. This assay is based on the identification of aberrant splice products in mRNA species transcribed in human primary fibroblasts. An Inherent pitfall of this assay is that products which undergo nonsense-mediated decay are difficult to identify due to their low abundance. Here, we describe the use of cycloheximide for blockage of NMD in human primary fibroblasts to identify spliced mRNA products that have a premature termination codon. Our study shows that blockage of NMD resulted in the identification of aberrant mRNA species in all patients tested. These products could not be identified when NMD

was not blocked. Furthermore, the assay shows that aberrant splicing even occurs in a healthy control which does not harbor any pathogenic variants, indicating that splicing is not always a perfect process.

The Human gene mutation database (HGMD) is a database which collects all known disease causing variants. 9% of these variants are annotated to effect splicing. However, because of a lack of basic functional screening in diagnostic labs, many variants which are determined to be missense or nonsense variants could also have an effect on splicing. It has been estimated that as much as 62% of disease causing variants can have an effect on splicing, highlighting the need for better functional evaluation of pathogenic variants (11).
General diagnostic screening for the presence of pathogenic variants in patients with a known monogenic disorder is typically done by Sanger sequencing of the coding parts of the gene of interest on genomic DNA. To test the effect of a specific variant on splicing, *in silico* prediction can be performed. Multiple splice prediction algorithm exist which can be implemented to give an indication of the effect a variant has on the strength of nearby splice elements. However, interpretation of the outcomes of these predictions remain challenging. *In silico* predicted changes in splicing do not necessarily reflect the situation *in vivo* (12). The way a variant effects splicing is also difficult to predict. However, prediction programs are becoming more reliable. An example is the prediction algorithm termed Automated Splice Site And Exon Definition Analyses, or ASSEDA for short. It has become available as an online tool that can be used to determine which splice sites will be used in the presence and absence of a given variant (13).

For Pompe disease, a comprehensive *in silico* prediction analysis of all *GAA* variants reported in the pompecenter.nl database that are associated with aberrant splicing has previously been performed, and results indicated that almost all investigated variants are predicted to effect splicing (14). Here we show that not all *in silico* predictions can be directly linked to aberrant splicing. A good example is the *GAA* missense variant c.1927G>A present in patient 3. This variant is often identified in Pompe patients, and has first been described in 1993 (15). It is described in the pompecenter.nl database as a potentially less severe missense variant, however many patients have a severe phenotype when this variant is present. Here we describe for the first time that the pathogenic effect of this variant might not only be caused by a mutation in the protein sequence, but also by its effect on splicing of *GAA* exon 14. Splice prediction software indicated that a weak splice acceptor site is generated in the presence of the variant at c.1929, but direct evidence of aberrant splicing could not be deducted from this prediction. This highlights the importance of sufficient functional screening of any variant regardless of their position within the gene.
With the advent of next generation sequencing, it has become evident that NGS techniques will become increasingly important in disease diagnostics. Whole exome and even whole genome sequencing will likely be implemented for diagnostic purposes in the near future (16). Although these techniques aid in identification of new disease-causing genes, these

assays are still based on the sequencing of genomic DNA, and therefore do not provide functional data on variants affecting gene expression and/or pre-mRNA splicing. RNA-sequencing would give insight to both of these parameters. Although this technique is still too expensive for implementation in diagnostic labs, it is a promising technique that could potentially give a better understanding of functional consequences of pathogenic variants. The addition of cycloheximide treatment to the original splicing assay promotes the identification of aberrant splice products. Here we show that ten out of eleven newly identified splice products could not have been identified without blockage of NMD using cycloheximide. The identification of these splice products leads to a better understanding of the modes of pathogenicity exerted by any variant. Furthermore, cryptic splice sites utilized in these aberrant products can be targeted with the use of AONs to restore canonical splicing, which highlights the potential of this treatment strategy.

MATERIALS AND METHODS:

Patients and controls

Patients were diagnosed with Pompe disease based on phenotype, GAA enzymatic activity and genomic DNA sequence analysis. The control is a random healthy individual from which fibroblasts were obtained. Patients and control gave informed consent for diagnostic purposes.

Nomenclature

Annotations used follow guidelines set by HGVS (http://www.hgvs.org/mutnomen/) (17).

Cell culture and cycloheximide treatment

Primary fibroblasts were obtained from skin biopsy and cultured in DMEM high glucose (Lonza) with 10% FBS (Hyclone) and Penicillin/Streptomycin/Glutamine in (Gibco) 5% CO2 at 37°C. Cycloheximide treatment was carried out at a concentration of 100 µg/ml cycloheximide (Sigma) for 48 hours.

mRNA analysis

mRNA analysis was performed as described previously (2). In short, RNA was harvested and purified using the RNAeasy miniprep kit (Qiagen) according to manufacturer's protocol. cDNA synthesis was performed on 800 ng RNA using iScript (Biorad) according to manufacturer's instructions. cDNA was diluted 5x before further analysis. RT-PCR was performed on samples using FastStart Taq Polymerase (Roche).Sequence analysis was performed directly on PCR samples using BigDye Terminator v3.1 (Thermo). If products were present at very low levels, TOPO cloning was performed using the TOPO® TA Cloning Kit (Thermo). RT-qPCR was carried out using ITaq universal SYBR Green Supermix (Biorad). ß-*Actin* was used as an internal control for qPCR analysis.

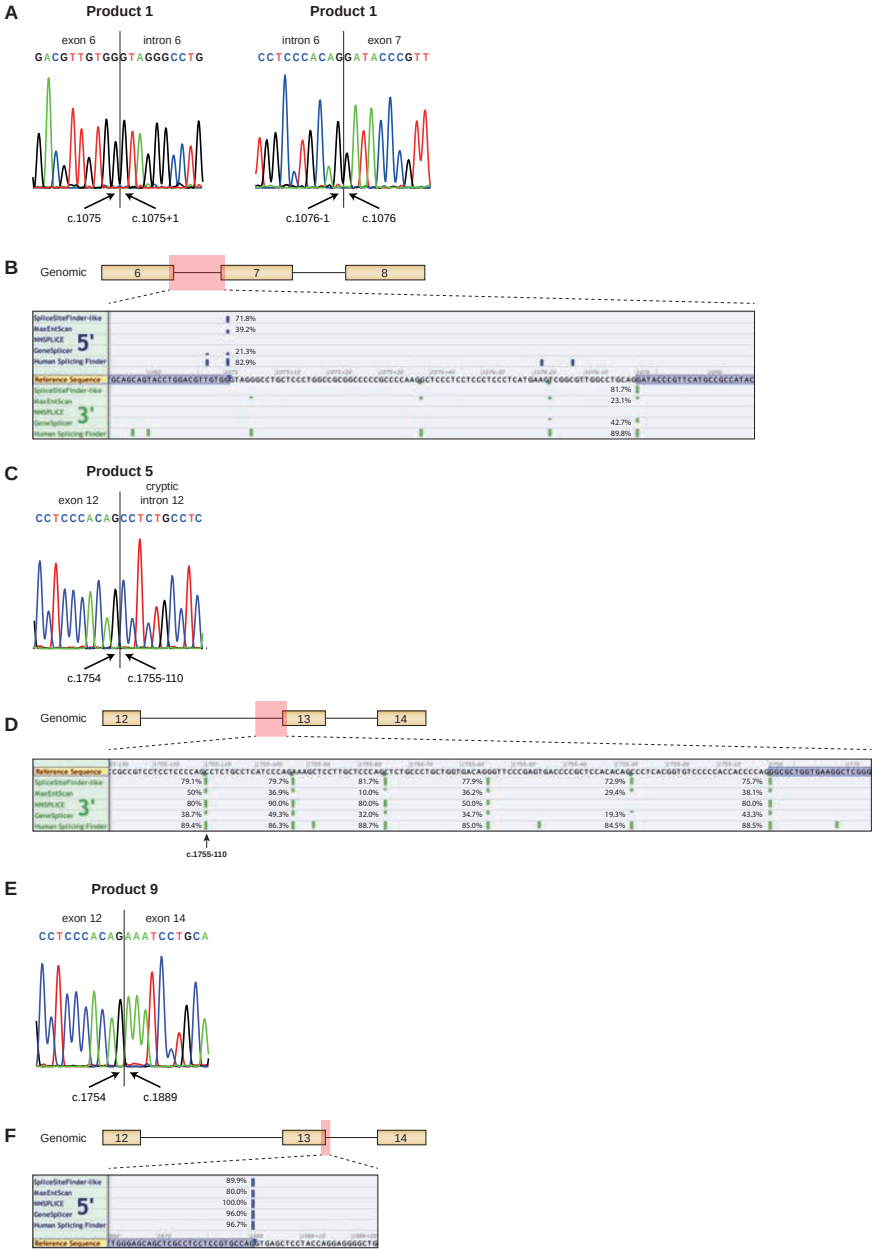
Protein Activity Assay

Cells were lysed in standard lysis buffer (50mM Tris (pH 7.5), 100mM NaCl, 50mM NaF, 1% Tx-100, Protease inhibitor). Protein quantification was done using the BCA Protein Assay (Thermo). 4-methylumbelliferone (4-MU)-α-D-glucopyranoside assay was performed as describes previously (18) to measure GAA enzymatic activity.

REFERENCES

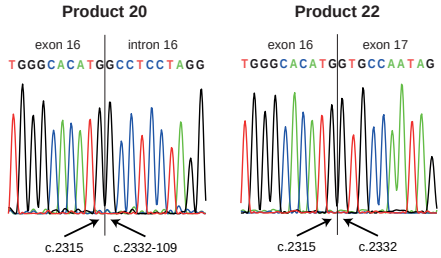
1. van der Ploeg, A.T. and Reuser, A.J. (**2008**) Pompe's disease. *Lancet*, 372, 1342-1353.
2. Bergsma, A.J., Kroos, M., Hoogeveen-Westerveld, M., Halley, D., van der Ploeg, A.T. and Pijnappel, W.W. (**2015**) Identification and characterization of aberrant GAA pre-mRNA splicing in pompe disease using a generic approach. *Hum Mutat*, 36, 57-68.
3. Schneider-Poetsch, T., Ju, J., Eyler, D.E., Dang, Y., Bhat, S., Merrick, W.C., Green, R., Shen, B. and Liu, J.O. (**2010**) Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nat Chem Biol,* 6, 209-217.
4. Peccarelli, M. and Kebaara, B.W. (**2014**) Regulation of natural mRNAs by the nonsense-mediated mRNA decay pathway. *Eukaryot Cell*, 13, 1126-1135.
5. Huie, M.L., Chen, A.S., Tsujino, S., Shanske, S., DiMauro, S., Engel, A.G. and Hirschhorn, R. (**1994**) 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.
6. Boerkoel, C.F., Exelbert, R., Nicastri, C., Nichols, R.C., Miller, F.W., Plotz, P.H. and 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.
7. Dardis, A., Zanin, I., Zampieri, S., Stuani, C., Pianta, A., Romanello, M., Baralle, F.E., Bembi, B. and 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.
8. Kroos, M., Hoogeveen-Westerveld, M., van der Ploeg, A. and Reuser, A.J. (**2012**) The genotype-phenotype correlation in Pompe disease. *Am J Med Genet C Semin Med Genet*, 160C, 59-68.
9. Maimaiti, M., Takahashi, S., Okajima, K., Suzuki, N., Ohinata, J., Araki, A., Tanaka, H., Mukai, T. and Fujieda, K. (**2009**) Silent exonic mutation in the acid-alpha-glycosidase gene that causes glycogen storage disease type II by affecting mRNA splicing. *J Hum Genet*, 54, 493-496.
10. Tsujino, S., Huie, M., Kanazawa, N., Sugie, H., Goto, Y., Kawai, M., Nonaka, I., Hirschhorn, R. and Sakuragawa, N. (**2000**) Frequent mutations in Japanese patients with acid maltase deficiency. *Neuromuscul Disord*, 10, 599-603.
11. Lopez-Bigas, N., Audit, B., Ouzounis, C., Parra, G. and Guigo, R. (**2005**) Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett*, 579, 1900-1903.
12. Jian, X., Boerwinkle, E. and Liu, X. (**2014**) In silico tools for splicing defect prediction: a survey from the viewpoint of end users. *Genet Med*, 16, 497-503.
13. Mucaki, E.J., Shirley, B.C. and Rogan, P.K. (**2013**) Prediction of mutant mRNA splice isoforms by information theory-based exon definition. *Hum Mutat*, 34, 557-565.
14. Zampieri, S., Buratti, E., Dominissini, S., Montalvo, A.L., Pittis, M.G., Bembi, B. and 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.
15. Hermans, M.M., Kroos, M.A., de Graaff, E., Oostra, B.A. and Reuser, A.J. (**1993**) Two mutations affecting the transport and maturation of lysosomal alpha-glucosidase in an adult case of glycogen storage disease type II. *Hum Mutat*, 2, 268-273.
16. Katsanis, S.H. and Katsanis, N. (**2013**) Molecular genetic testing and the future of clinical genomics. *Nat Rev Genet*, 14, 415-426.
17. den Dunnen, J.T. and Antonarakis, S.E. (**2000**) Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat*, 15, 7-12.
18. Kroos, M.A., Pomponio, R.J., Hagemans, M.L., Keulemans, J.L., Phipps, M., DeRiso, M., Palmer, R.E., Ausems, M.G., Van der Beek, N.A., Van Diggelen, O.P. et al. (**2007**) Broad spectrum of Pompe disease in patients with the same c.-32-13T->G haplotype. *Neurology*, 68, 110-115.

SUPPLEMENTARY FIGURES AND FIGURE LEGENDS

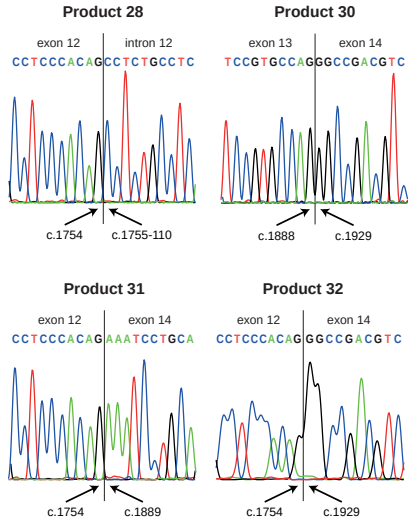


Supplementary Figure 1. In silico and sequencing analysis of splicing in healthy control primary fibroblasts after cycloheximide treatment. *A) Sanger sequencing results of products 1 and 3. B) In silico prediction of the region containing the 5' and 3' splice sites of intron 6 using Alamut. The pink block in the cartoon highlights the screened region. C) Sanger sequencing results of products 5 and 7. D) In silico prediction of the region containing the 3' splice site of intron 12 and upstream region using Alamut. E) Sanger sequencing results of product 9. F) In silico prediction of the 5' splice site of exon 13 using Alamut.*

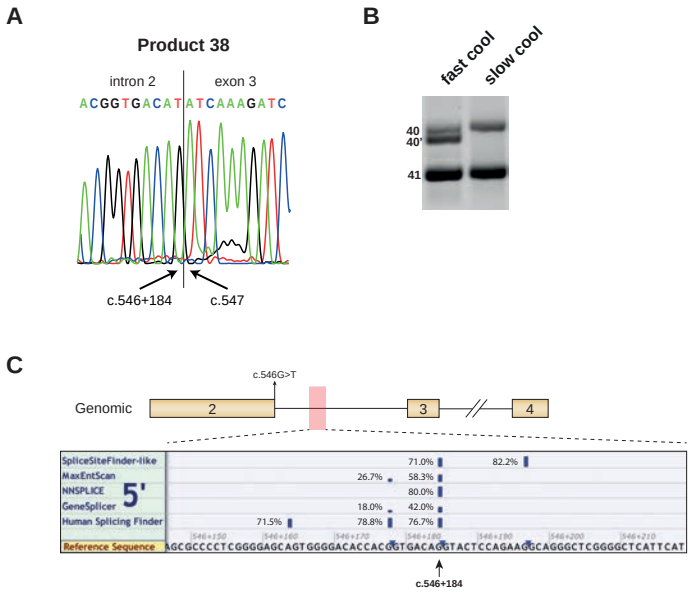
Supplementary Figure 2. In silico and sequencing analysis of splicing in healthy control primary fibroblasts after cycloheximide treatment. *A) Sanger sequencing results of splice junctions present in products 11, 12, 14, 16, 17 and 19. B) In silico prediction of the region from the cryptic splice site utilized by products 12 and 14 to the canonical splice site of exon 17.*



Supplementary Figure 3. Sequencing analysis of aberrant products identified in patient 2*. Sequencing data at splice junctions present in products 20, 22, 23 and 25 are shown.*



Supplementary Figure 4. Sequencing analysis of aberrant products identified in patient 3*. Sequencing data at splice junctions present in products 28-32 and 34 are shown.*



Supplementary Figure 5. Analysis of utilized splice sites in patient 4 primary fibroblasts identified after cycloheximide treatment*. A) Sequencing analysis of splice junctions in products 38 and 40. B) Analysis of the products 40 and 40'. Product 40' is visible when the PCR sample is cooled down quickly, but disappears when the sample is slowly cooled down, indicating secondary structure formation. C) In silico predictions of the region containing the natural cryptic splice site present in intron 2 using Alamut.*

Chapter 4

From cryptic towards canonical pre-mRNA splicing in Pompe disease: a pipeline for the development of antisense oligonucleotides

Atze J. Bergsma[1,2,3]
Stijn L.M. in 't Groen[1,2,3]
Frans W. Verheijen[4]
Ans T. van der Ploeg[2,3]
W.W.M. Pim Pijnappel[1,2,3]

[1] Molecular Stem Cell Biology, Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, the Netherlands
[2] Department of Pediatrics, Erasmus MC University Medical Center, Rotterdam, the Netherlands
[3] Center for Lysosomal and Metabolic Diseases, Erasmus MC University Medical Center, Rotterdam, the Netherlands
[4] Molecular Diagnostics, Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, the Netherlands

Molecular Therapy, Nucleic Acids 2016; in press

ABSTRACT

While 9% of human pathogenic variants have an established effect on pre-mRNA splicing, it is suspected that an additional 20% of otherwise classified variants also affect splicing. Aberrant splicing includes disruption of splice sites or regulatory elements, or creation or strengthening of cryptic splice sites. For the majority of variants, it is poorly understood to what extent and how these may affect splicing. We have identified cryptic splicing in an unbiased manner. Three types of cryptic splicing were analyzed in the context of pathogenic variants in the *acid α-glucosidase* gene causing Pompe disease. These involved newly formed deep intronic or exonic cryptic splice sites, and a natural cryptic splice that was utilized due to weakening of a canonical splice site. Antisense oligonucleotides that targeted the identified cryptic splice sites repressed cryptic splicing at the expense of canonical splicing in all three cases, as shown by RT-qPCR analysis and by enhancement of acid α-glucosidase enzymatic activity. This argues for a competition model for available splice sites, including intact or weakened canonical sites and natural or newly formed cryptic sites. The pipeline described here can detect cryptic splicing and correct canonical splicing using antisense oligonucleotides to restore the gene defect.

INTRODUCTION

The Human Gene Mutation Database, with >7 000 monogenic disorders, reports that approximately 9% of all known disease causing variants (numbering >170 000) are located at or near splice sites (http://www.hgmd.cf.ac.uk/) (1). In addition, it has been estimated that an additional ~20% of pathogenic variants affect splicing but are located at more distant locations in introns or in exons (2-4). Consequences of aberrant splicing include exon skipping, intron retention, cassette exon inclusion, and use of cryptic splice sites (5-7). There are several ways in which cryptic splicing may be induced by a pathogenic variant. One possibility is the generation of a new intronic splice donor or acceptor site that outcompetes a canonical splice site. Alternatively, a natural cryptic splice site may be present that is normally suppressed by a canonical splice site, but becomes the dominant splice site due to weakening of the canonical splice site by a pathogenic variant. Weakening of a canonical splice site can be a consequence of pathogenic variants that disrupt conventional splicing elements, including the splice site itself (8), the polypyrimidine tract (9), branch point sequence (10). or variants that modulate the activity of exonic and intronic splicing silencer or enhancer elements (reviewed by Scotti et al. (5)) (11,12). Another mechanism represents formation of a newly formed cryptic splice site at an exonic rather than intronic location. The location of a cryptic splice site in coding or noncoding RNA is likely important because exons and introns are defined within the pre-mRNA via interaction with distinct RNA binding proteins that may differentially affect cryptic splice site utilization. These include members of the SR and hnRNP families (13,14). However, for many disorders and gene variants, those that may affect splicing are poorly characterized.

Pre-mRNA splicing is regulated at several levels. Splice sites have short consensus sequences that are recognized by the spliceosome complex members, while additional, poorly defined sequences present in the introns and exons can modulate splice site choice. In addition, other mechanisms can affect splicing including the speed of RNA pol II transcription, chromatin structure, histone modifications, alternative transcription start and termination sites, GC content, mRNA export, mRNA stability, and expression levels of critical splicing proteins (13,14). These are likely reasons why *in silico* prediction of splice site choice is difficult and experimental testing is required to elucidate the effect of pathogenic gene variants.

We have previously described a splicing assay that can be used for the detection and quantification of aberrant splicing in Pompe disease (OMIM232300) (15). It is based on the analysis of all exons of the *acid* α*-glucosidase* (*GAA*) gene. Besides the detection of aberrant splicing that may be linked to the disease, this approach provides quantitative information on the extent of leaky wild type splicing, which is informative for disease severity, and on the splicing mechanism, which is required to design methods for splicing correction. Antisense oligonucleotides (AONs) provide a way to interfere with splicing in a sequence-specific manner (16). AONs can be targeted to the pre-mRNA region of interest

and they can repress the activity of a splicing motif. The clinically most advanced examples are enhancement of exon inclusion of *survival of motor neuron 2* (*SMN2*) pre-mRNA in spinal muscular atrophy (SMA) (17), and exon skipping of the *dystrophin* pre-mRNA in Duchenne muscular dystrophy (DMD) (18). Other pre-clinical examples include modulation of splicing in Hutchinson-Gilford progeria syndrome (19) and type I Usher syndrome (20). The developments for enhancing cellular uptake using conjugation of AONs to cell penetrating peptides (21-23) and via co-administration of hexose (24) is expected to further stimulate the clinical development of AON-based splicing modulation.

Here, we present a pipeline in which characterization of aberrant splicing in Pompe disease is first performed using a generic splicing assay (15). The information obtained is then used to design an AON based on inhibition of cryptic splicing, and the AON is tested in patient-derived fibroblasts. Pompe disease is an autosomal recessive disorder caused by variants in the *acid α-glucosidase* (25) gene and results in lysosomal glycogen accumulation that predominantly affects skeletal muscle in the childhood/adult onset form of the disease (26). Currently, enzyme replacement therapy (ERT) is available, but there are several reasons to develop alternative therapies, including the heterogenic clinical response, the inability to completely counteract the disease, and the extremely high costs (27-30). Three pathogenic *GAA* variants with different effects on cryptic splicing were analyzed, and this enabled the successful design of AONs that promoted splicing correction. The pipeline from splicing analysis to AON-based splicing correction may provide a basis for personalized medicine in human disease.

RESULTS

Unbiased splicing analysis of all exons identifies aberrant splicing from an unknown GAA allele in patient 1

Patient 1 was diagnosed with Pompe disease based on a deficiency of GAA enzymatic activity in fibroblasts and leukocytes (Table 1). Standard diagnostic DNA analysis, which includes Sanger sequencing of exons and short flanking intronic regions, identified the c.-32-13T>G (IVS1) *GAA* variant on one allele (9), but the pathogenic variant on the second allele was not identified (Figure 1a). We then applied our splicing assay to test whether the unknown variant may be a regulatory or deep intronic splicing variant. Flanking exon PCR analysis of all spliced *GAA* exons showed the expected aberrant splice products for exon 2 caused by the IVS1 variant, including a full exon 2 skip, a partial exon 2 skip, and leaky wild type splicing (Figure 1b) (15,31,32). Amplification of exons 15 and 16 revealed several higher molecular weight PCR products of low abundance that were not normally observed in cells from healthy controls (15). To identify these, the exon 15 PCR products were directly processed by Topo cloning, and 93 clones were analyzed by Sanger sequencing. Four mRNAs were identified (Figure 1c and Suppl. Figure S1a). All aberrantly spliced products (1-3) utilized a cryptic splice acceptor in intron 15 at

c.2190-344, and showed inclusion of various parts of intron 15. Product 1 included the downstream intronic region up to the splice acceptor of exon 16. This product contains a premature stop codon and is likely to undergo mRNA decay, in agreement with the low abundance of this product. Product 2 and 3 included a small part of intron 15 by using the cryptic splice donor sites at c.2190-282 and c.2190-300, respectively. Product 2

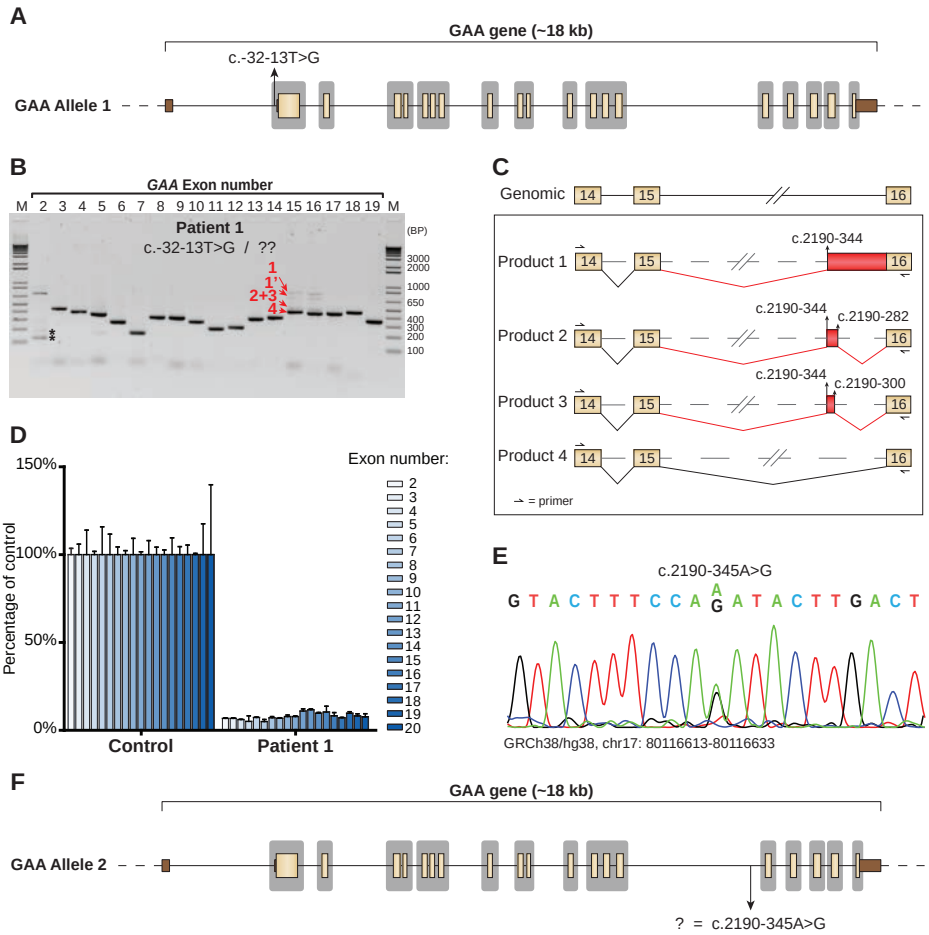


Figure 1. Identification of a deep intronic pathogenic variant in patient 1 using the splicing assay. *(a) Scaled cartoon of GAA allele 1 from patient 1 carrying the c.-32-13T>G (IVS1) variant. Coding exons are indicated in yellow, untranslated regions in brown. Lines represent introns. Areas in grey indicate the regions that are sequenced as part of standard diagnostic practice. (b) Flanking exon RT-PCR of all GAA coding exons. Asterisks indicate known aberrant splicing products caused by the IVS1 variant. Products 1-4 indicate novel mRNA products. Product 1' refers to a secondary structural variant of product 1 (see Suppl. Figure S1B). (c) Cartoon of splicing products 1-4. Boxes indicate exonic regions, (dashed) lines represent intronic regions. Aberrant splicing events are indicated in red. (d) Exon internal RT-qPCR analysis of GAA exons 2-20. Data are normalized for ß-Actin and for a healthy control. Data represent means of three technical replicates +/- SD. (e) Sequence analysis of genomic DNA of the region surrounding the c.2190-344 cryptic splice acceptor site. (f) Cartoon of GAA allele 2 showing the location of the c.2190-345A>G variant.*

contained the same premature stop codon as product 1, and was likely degraded as well. In product 3, the reading frame remained intact. Therefore, the low abundance of the PCR product suggests minimal usage of the c.2190-300 splice site. Product 4 contained the canonical splice junctions and was the most abundant product of exon 15 amplification. It should be noted that the canonically spliced exon 15 product is also expressed from the IVS1 allele. No unique sequence of product 1' could be identified. However, this product disappeared by adapting the final cooling step of the PCR reaction, suggesting that it represented a conformational variant (Suppl. Figure S1b). The distribution of clones was: product 4 (canonical): 75 clones; product 1 and 1': 8 clones; product 3: 6 clones; product 2: 4 clones.

Next, exon internal RT-qPCR analysis of exons 2 to 20 was performed to quantify aberrant splicing in cells from patient 1. All exons were expressed at similarly low levels of up to 12% compared to a healthy control (Figure 1d). The IVS1 *GAA* variant is known to allow 10-15% leaky wild type splicing (9,15,31-33). This suggests that the majority of *GAA* expression is derived from the IVS1 allele, and that the 2nd allele containing the unknown *GAA* variant was expressed at much lower levels. Taken together, splicing analysis of all exons explained why patient 1 was diagnosed with Pompe disease, despite the lack of identification of the 2nd *GAA* allele.

Table 1. Patient characteristics.

	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 (years)	Onset
Patient 1	c.-32-13T>G	? = c.2190-345A>G	11.2	62	Adult
Patient 2	c.-32-13T>G	? = c.2190-345A>G	12.5	66	Adult
Patient 3	c.1552-3C>G	c.1552-3C>G	12.6	16	Juvenile
Patient 4	c.1256A>T	c.1551+1G>T	5.4	1,3	Juvenile

reference *GAA* transcript for cDNA annotation is RefSeq NM_000152.3

The newly identified deep intronic variant c.2190-345A>G causes aberrant splicing in patient 1

To identify which *GAA* DNA variant could be responsible for aberrant splicing of exon 15 and 16, we first analyzed the sequences of the aberrantly spliced mRNA products 1-4. This failed to show any potentially pathogenic *GAA* variant. We then analyzed the genomic DNA surrounding the cryptic splices utilized in products 1-3. This revealed the presence of a heterozygous A>G variant at -1 relative to the cryptic splice acceptor site at c.2190-344 (Figure 1e). The intronic variant was missed by standard diagnostic sequence analysis

because of its deep intronic location (Figure 1f, regions sequenced by standard diagnostics are indicated in grey). Its location close to the cryptic splice acceptor site suggested that it was involved in mediating the aberrant splicing observed in cells from patient 1.

To test the effect of the c.2190-345A>G *GAA* variant on splicing, we used splicing prediction programs and minigenes. Alamut software, which includes five different algorithms to identify either 3' or 5' splice sites, predicted the generation of a strong splice acceptor site at c.2190-344 in the presence of the c.2190-345A>G variant (Figure 2a). This acceptor site was identical to the site identified in the aberrantly spliced products 1-3 of patient 1. Its predicted strength was similar to the strength of the canonical exon 16 splice acceptor site (compare Figure 2a and Suppl. Figure S2a). The region around the cryptic donor sites identified in products 1-3 was also scrutinized and this resulted in a moderately strong prediction of a splice donor site at c.2190-300 (used in product 3); a strong prediction of a splice donor at c.2190-287 (not detected *in vivo*); and a weak prediction (below the diagnostic threshold) of a splice donor site at c.2190-282 (used in product 2) (Suppl. Figure S2b). Some of these potential splice donor sites were indeed utilized *in vivo* in a subset of mRNAs. In summary, splicing prediction programs uniformly predicted the generation of the new splice acceptor at c.2190-344 in response to the c.2190-345A>G variant, while prediction of splice donor sites was variable.

Next, minigenes were used to test the pathogenic nature of the c.2190-345A>G variant. Two minigenes including the genomic DNA region spanning *GAA* exon 15 to 17 were generated, one wild type (c.2190-345A) and one containing the variant (c.2190-345A>G) (Figure 2b). These constructs were transfected in HEK293T cells, and *GAA* exon 16 splicing was analyzed using flanking exon RT-PCR. Endogenous HEK293T transcripts were circumvented by usage of a reverse primer specific to the minigene backbone. In the wild type minigene, a single product for amplification of exon 16 was detected that represented canonical exon 16 splicing (product 8, Figure 2c,d and Suppl. Figure S2c). This product was absent from untransfected cells, demonstrating that it was derived from the minigene. In contrast, the minigene containing the c.2190-345A>G variant failed to show expression of the canonical exon 16 splicing product, but showed expression of multiple alternative products (Figure 2c). The exon 16 PCR product was Topo cloned, and sequence analysis of 16 clones confirmed absence of the wild type product. Similar to endogenous exon 16 splicing in primary fibroblasts from patient 1, all aberrant products utilized the c.2190-344 splice acceptor in combination with various cryptic splice donor sites (Figure 2d and Suppl. Figure S2c). Product 5 contained intron retention from c.2190-344 towards exon 16 (three clones); product 6 showed usage of the cryptic splice donor at c.2190-282 (eight clones); product 7 used a cryptic splice donor at c.2190-287 (two clones) (Figure 2d and Suppl. Figure S2c). These findings indicate that the c.2190-345A>G variant causes aberrant splicing of *GAA* intron 15 and represents the new pathogenic *GAA* variant in patient 1. The unbiased analysis of splicing in cells from patient 1 has enabled

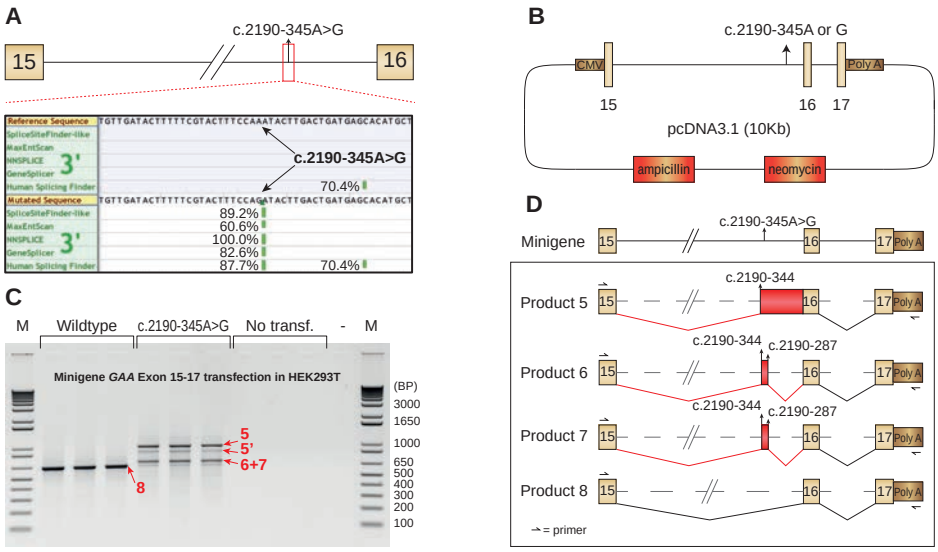


Figure 2. Assessment of the pathogenic nature of the c.2190-345A>G variant. *(a) In silico prediction of the effect of the c.2190-345A>G variant on the generation of a 3' cryptic splice site. Percentages indicate relative scores of strength per algorithm. (b) Cartoon of the minigenes containing the c.2190-345A or c.2190-345G variants. (c) RT-PCR analysis following transfection of minigenes into HEK293T cells. The reverse primers is specific for minigene mRNA as indicated in Figure (d). Products 5-8 indicate sequenced PCR products. Product 5' refers to a secondary structural variant of product 5. Three biological replicates are shown. (d) Cartoon indicating the splicing products from (c) as identified by sequence analysis.*

the identification of aberrant splicing and the associated genomic DNA variant that was missed by standard diagnostic sequencing.

Splicing correction by AONs in cells derived from patient 1 and a sibling, patient 2

Previously, AONs have been used to modulate pre-mRNA splicing. A relatively straightforward approach has been to block canonical splice sites to circumvent a variant hotspot and restore the reading frame, such as in DMD. AONs that bind to a splice site at the pre-mRNA can promote skipping of the splice site, which often results in utilization of another available (cryptic) splice site. We reasoned that a similar approach may be used to prevent utilization of a cryptic splice site such as the one generated by the c.2190-345G>A variant in patient 1. To test this, we first analyzed the feasibility of AON-mediated splicing modulation of pre-mRNA in primary fibroblasts. A control phosphorodiamidate morpholino oligonucleotide (PMO)-based AON was designed that targets the splice donor site of exon 4 in *Cyclophilin A* (*CypA*) pre-mRNA (AON *CypA*) (Suppl. Figure S3a; based on (34)). AON *CypA* efficiently promoted skipping of *CypA* exons 3 and 4 in fibroblasts from patient 1 as analyzed by RT-PCR (Suppl. Figure S3b) and RT-qPCR (Suppl. Figure S3c). This demonstrated robust splicing modulation using PMO-based AONs in primary fibroblasts. Next, we designed an AON that targeted the cryptic splice site at c.2190-

344 (AON intron 15; Figure 3a). Transfection of AON intron 15 in fibroblasts from patient 1 resulted in lower levels of endogenous aberrant splice products 1, 2 and 3, and higher abundance of canonical splice product 4, as shown by semi-quantitative exon flanking RT-PCR analysis (Figure 3b). Mock transfection or transfection with AON *CypA* showed no effect. Quantitative RT-qPCR analysis was then performed to quantify changes in *GAA* splicing. Changes in wild type splicing were quantified using primers spanning the *GAA* exon 1-exon 2 splice junction. Low-abundant aberrantly spliced products were quantified using a forward primer annealing to exon 15 and a reverse primer that anneals to the common intronic region included in all aberrant splice products 1-3 (Figure 3c, lower panel). Treatment with AON intron 15 promoted splicing towards canonical *GAA* mRNA, while it inhibited aberrant *GAA* splicing (Figure 3c). Mock transfected cells or cells transfected with AON *CypA* showed no effect on *GAA* splicing. The effect of AON intron 15 treatment on GAA enzymatic activity was then assessed, and this showed a more than 2-fold enhancement (Figure 3d). Importantly, this increase would result in a residual enzymatic activity that is above the disease threshold of 20% of control, suggesting that the AON intron 15-mediated splicing correction was sufficient to alleviate Pompe disease in cells from patient 1.

A sibling of patient 1, patient 2, was also diagnosed with Pompe disease. This patient contained the IVS1 *GAA* variant but also lacked identification of the second pathogenic *GAA* allele using standard diagnostic analysis. Genomic DNA analysis showed that the unknown

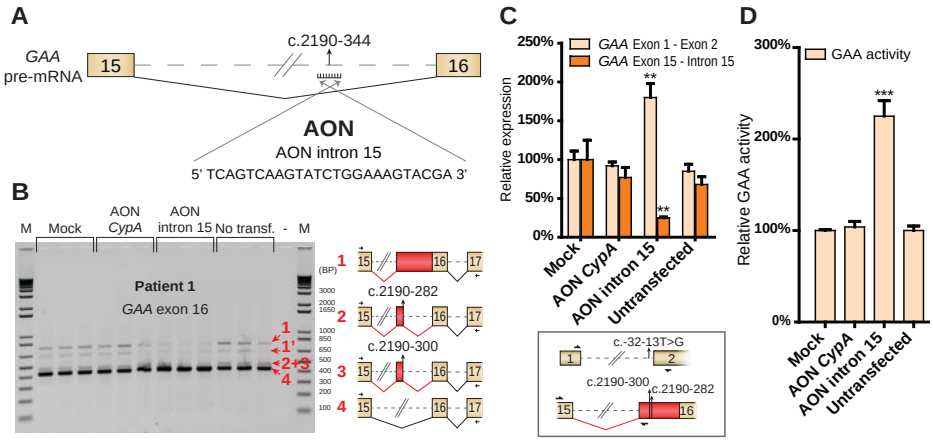


Figure 3. Correction of aberrant splicing in fibroblasts from patient 1 using an AON. *(a) Cartoon depicting the region in the GAA pre-mRNA that was targeted with AON intron 15. (b) Flanking exon RT-PCR analysis of GAA exon 16 in fibroblasts from patient 1 treated with AON intron 15. AON CypA was used as a control for AON treatment (see Supplementary Figure S3). Cartoons depict spliced mRNAs. Primer locations are indicated. Three biological replicates are shown. (c) RT-qPCR analysis of the experiment outlined in (b). Primers used for specific amplification of canonical (exon 1 - exon 2) or aberrant (exon 15 - intron 15) GAA mRNA are indicated in the cartoon below the graph. Data are normalized for ß-Actin and for mock transfection. Data represent means +/- SD of three biological replicates (** p = 0,01). (d) GAA enzymatic activity of patient 1 fibroblasts transfected with AONs. Data are normalized for mock transfection and represent the mean +/- SD of three biological replicates (*** p = 0,001).*

allele of patient 2 was c.2190-345A>G, identical to the second allele of patient 1 (Suppl. Figure S4a). Flanking exon RT-PCR analysis of exon 16 in primary fibroblasts indicated the same aberrant splicing products compared to patient 1 (Suppl. Figure S4b). Treatment with AON intron 15 resulted in a similar correction of splicing compared to patient 1, as shown by RT-PCR (Suppl. Figure S4b) and RT-qPCR (Suppl. Figure S4c). Also the GAA enzymatic activity was enhanced by AON intron 15 treatment to 250% compared to mock transfection (Suppl. Figure S4d), similar to patient 1. The results with patient 2 confirm the feasibility of splicing correction by repression of the newly formed, deep intronic cryptic splice site at c.2190-344.

Inhibition of a natural intronic cryptic splice acceptor to correct splicing in cells derived from patient 3

To test whether the blocking of cryptic splice sites by AONs may represent a general approach to correct splicing in Pompe disease, we examined two additional patients.

Patient 3 (Table 1) represents an example in which an intronic *GAA* variant (homozygous c.1552-3C>G, (15,35)) weakens a canonical splice site (splice acceptor of exon 11), which promotes utilization of a nearby cryptic splice site (c.1552-30, Figure 4a). We hypothesized that blockage of this cryptic splice site with an AON may enhance the probability of utilizing the canonical splice site. The details of aberrant splicing in fibroblasts of this patient have been described by us previously (patient 6 in (15)). We now performed Topo cloning of cDNA products derived from flanking exon PCR of exon 10 (S1 Table, -AON), and this resulted in the identification of two additional splicing products (10 and 11 in Figure 4b). All identified splicing products include: retention of intron 10 (product 9; reading frame disrupted), splicing from the canonical splice donor site at exon 10 to a natural cryptic splice acceptor at c.1552-30 in intron 10 (product 10, reading frame intact), splicing from a natural splice acceptor at c.1537 in exon 10 to a cryptic splice acceptor at c.1552-30 (product 11; reading frame intact), leaky wild type splicing (product 12), and skipping of exon 10 plus utilization of the cryptic splice acceptor at c.1552-30 (product 13; reading frame intact) (Figure 4b cartoons and Suppl. Figure S5a).

AON intron 10 was designed to target the cryptic splice site at c.1552-30 (Figure 4a). Transfection of this AON in fibroblasts from patient 3 caused a reduction of expression of aberrantly spliced mRNAs (products 9-11 and 13), while canonical spliced mRNA (product 12) seemed enhanced, as analyzed by flanking exon PCR of exon 10 (Figure 4b). This was also evident from the numbers of Topo clones derived from this PCR: without AON treatment, 49% of clones contained wild type exon 10 spliced cDNA, while treatment with AON intron 10 yielded 80% wild type cDNA clones (Suppl. Table S1). It should be noted that product 9 (full intron 10 retention) is out of frame and is likely subject to mRNA degradation, which prevents an estimation of the total amount of transcripts produced. Both flanking exon RT-PCR and Topo cloning can only be considered semi-quantitative methods. To quantify the effect of

AON intron 10, RT-qPCR analysis was performed using a primer that anneals to the boundary of exon 10/11, as this distinguishes wild type form aberrant products (Figure 4c, lower panel). This confirmed that AON intron 10 enhanced canonical exon 10 splicing with 50% (Figure 4c). In agreement, a modest but significant increase of 1.3 fold of GAA enzymatic activity was detected in cells treated with AON intron 10 (Figure 4d). This patient has a considerable residual GAA activity (12.6 nmol/hr/mg, representing ~10% of healthy control values), and an enhancement of 130% elevates GAA activity to levels closer to the disease threshold of ~20% of healthy control. Taken together, these results show that aberrant splicing caused by a splice site variant can be at least partially corrected by preventing the utilization of a nearby located cryptic splice site using AONs.

Correction of aberrant splicing from a newly formed exonic splice donor site in cells derived from patient 4

To test whether it is also possible to modulate splicing by blocking a newly formed cryptic splice site formed at an exonic location, we examined patient 4. This patient carries the c.1256A>T missense *GAA* variant on allele 1, which leads to the generation of an exonic cryptic splice donor site at c.1254 in exon 8 (patient 8 in (15); Figure 4e, Suppl. Figure S5b). Products from this allele include skipping of the 3' part of exon 8. The second allele of this patient contains another *GAA* splicing variant, c. 1551+1G>T, which causes skipping of exon 10 ((15); Suppl. Figure S5b). We hypothesized that blocking of the cryptic splice site at c.1254 with an AON may restore normal splicing of the c. 1256A>T allele. First, we tested the possibility that the c.1256A>T missense (p.D419V) variant, which would still be present in the splice-corrected mRNA, affects GAA enzymatic activity. To test this, the *GAA* cDNA was cloned into an expression vector, and the c.1256A>T variant was introduced by site directed mutagenesis. Transfection into HEK293T cells followed by measurement of GAA enzymatic activity showed that the c.1256A>T variant did not impair GAA enzymatic activity (Suppl. Figure S5c). This suggests that its pathogenic effect can be solely attributed to its effects on pre-mRNA splicing. Next, AON exon 8 was designed complementary to the cryptic splice site including the variant at c.1256 (Figure 4e, variant sequence in red). AON exon 8 was transfected in primary fibroblasts from patient 4, and the effect on *GAA* mRNA expression and GAA enzymatic activity was determined. Flanking exon PCR was performed for exon 9 using primers designed to specifically detect splicing correction from the c.1256A>T allele with minimal interference from the c.1551+1G>T allele. This was achieved by using a forward primer that anneals to the 3' part of exon 8, which is largely skipped in mRNA from the c.1256A>T allele, and a reverse primer that anneals to exon 10, which is completely skipped in mRNA from the c.1551+1G>T allele (Figure 4f). AON exon 8 enhanced the expression of canonically spliced exon 8 from the c.1256A>T allele (product 14, Figure 4f). Quantitative analysis using RT-qPCR with similar allele-specific primers confirmed that AON exon 8 enhanced

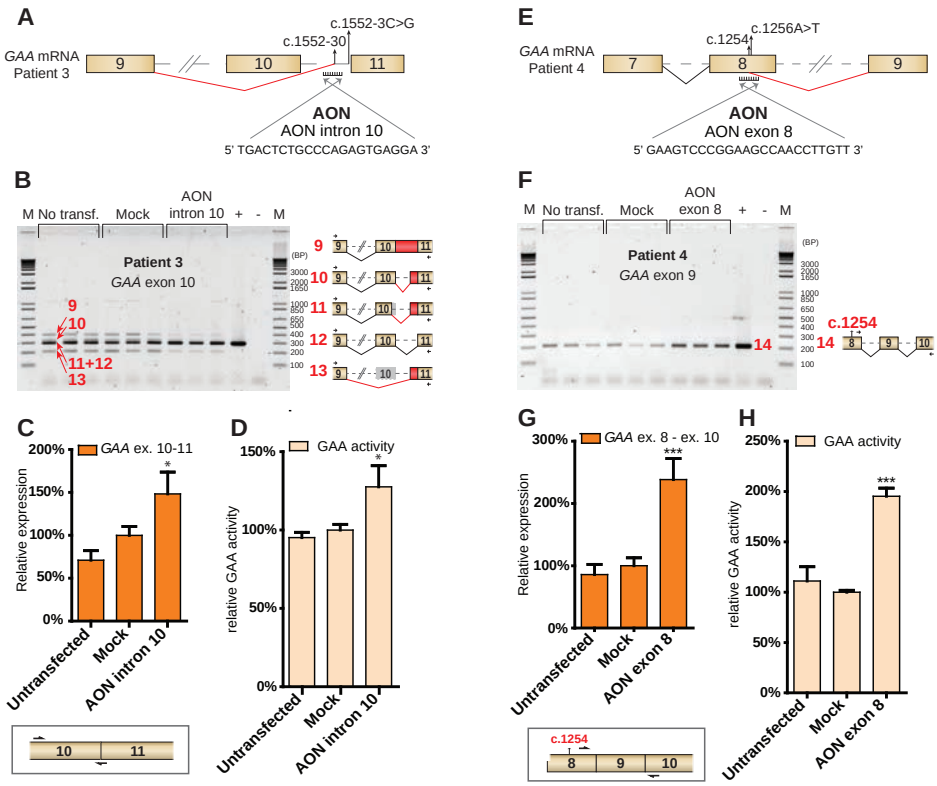


Figure 4. AON-mediated correction of cryptic splicing in fibroblasts from two additional patients. *(a) Cartoon depicting the region in the GAA pre-mRNA of patient 3 that was targeted with an AON. This patient utilizes a cryptic splice site at c.1552-30 due to a homozygous variant c.1552-3C>G as described previously (15,35). The sequence of AON intron 10 is shown. (b) Flanking exon RT-PCR analysis of GAA exon 10 in fibroblasts from patient 3 that were treated with AON intron 10. Cartoons depict spliced mRNAs. Primer locations are indicated. Three biological replicates are shown. + represents analysis of a healthy control. (c) RT-qPCR analysis of GAA exon 10 – exon 11 expression in fibroblasts from patient 3 treated with AON intron 10. The forward primer anneals to GAA exon 10 and the reverse primer anneals to the GAA exon 10-11 junction for specific detection of canonical exon 10-11 splicing. Data are normalized for ß-Actin and for mock transfection and represent means +/- SD of three biological replicates (* p = 0,05). (d) GAA enzymatic activity in patient 3 fibroblasts transfected with AON intron 10. Data represent means +/- SD of three biological replicates (* p = 0,05). (e) Cartoon depicting the region in the GAA pre-mRNA of patient 4 that was targeted with an AON. This patient utilizes a cryptic splice site at c.1254 due to the c.1256A>T variant as described previously (15). The sequence of AON exon 8 is shown. (f) Flanking exon RT-PCR analysis of GAA exon 9 on fibroblasts from patient 4 transfected with AON exon 8. Primers indicated in the cartoon specifically amplify canonically spliced GAA exon 8 mRNA from the allele harboring the c.1256A>T variant. Three biological replicates are shown. + represents analysis of a healthy control. (g) RT-qPCR analysis of GAA exon 8 – exon 10 expression in fibroblasts from patient 4 treated with AON exon 8. Primers only amplify canonically spliced GAA exon 8 mRNA from the allele harboring the c.1256A>T variant. Data represent means +/- SD of three biological replicates (*** p = 0,001). (h) GAA enzymatic activity in patient 4 fibroblasts transfected with AON exon 8. Data represent means +/- SD of three biological replicates (*** p = 0,001).*

expression of the c.1256A>T allele with 2.3 fold (Figure 4g). Besides promoting canonical exon 8 splicing, AON exon 8 also caused skipping of exon 8, as shown by flanking exon PCR of exon 8 (Suppl. Figure S5d). The net result of this effect was a partial correction

of the splicing effect of the c.1256A>T variant. This is further discussed in the legend to Suppl. Figure S5d.

In agreement with the mRNA analysis, GAA enzymatic activity was increased 2 fold as the result of AON exon 8 treatment (Figure 4h). This patient showed a relatively low GAA enzymatic activity of 5.4 nmol/hr/mg (4.4% of healthy control) and a juvenile disease onset. Elevation of the GAA enzymatic activity with a factor 2 is expected to attenuate the onset and severity of Pompe disease. In summary, it is possible to redirect aberrant splicing towards canonical splicing by targeting a cryptic splice site formed by a variant at an exonic location.

DISCUSSION

Assay for splicing analysis and genotyping

The assessment of a gene of interest for possible splicing defects can provide both qualitative and quantitative information on aberrant splicing (15). This is important for establishing the pathogenic nature of a variant, with implications for diagnostics, genetic counseling, prediction of disease severity, and the development of novel treatment options. The splicing assay can also be used to identify deep intronic variants that have been missed by standard diagnostic DNA sequencing. Especially in countries in which a diagnosis requires the identification of two pathogenic variants to be eligible for therapy, as is the case with ERT for Pompe disease, this information can have a serious impact on treatment options for the patient. A rather straightforward flow has been applied here to identify a deep intronic variant. This started with the detection of aberrant splicing using unbiased, PCR-based analysis of all exons (Figure 5). It is important to note that also small amounts of aberrant mRNA products should be analyzed as these may be out of frame and subject to mRNA decay, as is the

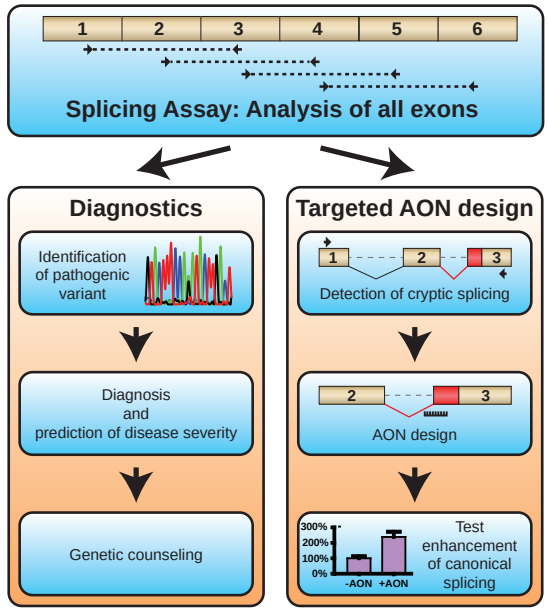


Figure 5. Pipeline for the identification and targeting of aberrant splicing events as applied to Pompe disease. *The splicing assay consists of unbiased mRNA analysis of all coding exons by RT-(q)PCR and Sanger sequencing (15). The results aid in diagnostics (left panel), but also provide the basis for the development of AONs that repress the utilization of natural or newly formed cryptic splice sites (right panel).*

case in patient 1. Next, the region surrounding the novel splice sites was sequenced at the genomic DNA level. This revealed a potential splicing variant. Third, the pathogenic effect of the variant was confirmed by *in silico* analysis and using a minigene construct (Figure 5). This approach improves standard diagnostic practice and provides a potential basis for therapeutic development.

The use of RNA sequencing-based methods for splicing analysis of patients may play an increasingly important role in the future (36-38). However, at present, there are several disadvantages of using RNA sequencing-based techniques for diagnostics. In particular, to fully assess aberrant splicing of a particular gene in a quantitative manner, deep sequencing combined with bioinformatics analysis is required, which is rather expensive. The present PCR-based assay can be standardized and is fast, cheap, and sensitive. A useful addition may be the implementation of new techniques such as single molecule real-time (SMRT) DNA sequencing (39). This could be applied to cDNA, for example to determine whether pathogenic variants reside on the same allele or are compound heterozygous, a question that could not be addressed for patients 1 and 2 due to the absence of DNA from the parents.

Disturbance of the splicing equilibrium by point variants

It is intriguing that a single point variant can generate a strong splice site as is the case for patient 1. This is in line with the fact that the sequence requirements for the generation of a potential splice site are relatively limited (13). The decision to utilize a splice site can depend on many aspects of gene expression, including the speed of RNA pol II transcription, cis-acting sequences in the pre-mRNA, pre-mRNA structure, chromatin modifications, and expression levels of splicing proteins (13,14). It is therefore very difficult to predict the outcome of splicing based on *in silico* analysis of splice site strength. In patient 1 and 2 the formation of a new 3' splice site not only results in retention of the downstream intronic sequence, but also in the utilization of two natural cryptic 5' splice sites (Suppl. Figure S2b). The 10-15% leaky wild type splicing detected in these patients is likely derived from the IVS1 allele, which is known to have this level of normal splicing (9,15,31,32). It should be noted that we could not formally prove that the IVS1 and c.2190-345A>G were compound heterozygous variants, as genomic DNA from the parents of patients 1 and 2 could not be obtained. However, the combined evidence obtained here strongly suggests that this is the case. In patient 3, the weakening of the canonical 3' splice site of exon 11 is straightforward, but the outcome with at least 3 different splicing products and the involvement of a nearby cryptic splice site was difficult to anticipate (15). Patient 4 illustrates that missense variants can also affect splicing. In fact, recent evidence suggests that this phenomenon is rather frequent and has been estimated to be the case in ~20% of all missense variants (2-4). In this patient, aberrant splicing appeared to be solely responsible for the pathogenic effect of the c.1256A>T variant, as the enzymatic activity of the mutated protein was similar to

wild type *GAA* in cDNA expression analysis. Taken together, point variants can induce complex changes in AS that are difficult to predict but can be characterized using PCR-based analysis of all exons.

Targeted splicing correction using AONs

The information obtained with the splicing assay can directly be used to design AONs directed towards inhibition of cryptic splicing by targeting the identified cryptic splice site (Figure 5). This was done in the case of patient 1 and 2. In primary fibroblasts from these patients, the enhancement of GAA enzymatic activity after treatment with AON was sufficient to reach levels above the Pompe disease threshold of 20% activity of healthy controls. This is further illustrated by the experience in our diagnostic center, at which all Pompe patients in the Netherlands are diagnosed. A combination of clinical and biochemical parameters applied to >200 patients and >300 healthy individuals showed a patient range to be 0-20 nmol/hr/mg, and a normal range from 40-180 nmol/hr/mg ((40) and diagnostics department, Erasmus MC). This suggests that the AON designed to correct splicing in both patients has the potential to restore GAA enzyme levels towards healthy control levels. It should be noted that the correction of glycogen storage by AONs in cells in culture cannot be studied at present, because cells from juvenile/adult onset Pompe patients do not show glycogen accumulation *in vitro*. This may imply that additional factors are required that are involved in lysosomal pathology in juvenile/adult onset Pompe disease. In agreement, several studies including our own report on the strongly heterogeneous disease onset and progression of juvenile/adult Pompe patients, also when they have identical *GAA* genotypes (41-43).

A priori, it was not always obvious that AON mediated repression of cryptic splicing would enhance canonical splicing. In the case of patients 1 and 2, we anticipated that repression of a deep intronic, newly formed splice site would restore canonical splicing, as was indeed observed. However, patients 3 and 4 represented more challenging cases.

In patient 3, the point variant was located near a 3' splice site, which resulted in weakening of the canonical splice site and utilization of the c.1552-30 cryptic splice site. Another observed event as a consequence of the c.1552-3C>G variant was retention of intron 10. It was not clear whether blocking of the cryptic splice site with an AON would promote canonical splicing or intron retention. This depended on the question whether the canonical splice site was still partially functional despite the nearby c.1552-3C>G variant. In this case, the AON blocked utilization of the cryptic splice site without promoting other detectable forms of aberrant splicing. This showed that the canonical splice site was still partially functional, suggesting a competition model in which the strongest nearby splice site is preferred. Competition for splice site usage is thought to play an important role in the outcome of pre-mRNA splicing, as is the distance to the next available splice site (13,14). The fact that repression of the natural upstream cryptic

splice site promoted splicing to a weakened canonical splice site is encouraging and may suggest a general strategy for similar cases.

In patient 4, the targeting of an AON to an exonic region represented a challenge, as this has the potential to interfere with protein translation (44). We found no evidence for this, as canonical splicing was increased with a concomitant increase in GAA protein activity. However, the AON targeted to exon 8 resulted in a new equilibrium of AS, in which both canonical splicing and aberrant splicing were altered. This illustrates that AON targeting of cryptic splice sites may induce new splice forms, and it underscores the importance of testing for this possibility. In this patient, the AON-induced changes in AS were advantageous as the c.1256A>T allele showed an improvement to 195%, while the effect on exon 8 splicing of the 2nd allele was irrelevant as this already was affected by aberrant splicing of exon 10.

Application of AONs in human disease

To translate the present results to a clinical setting, more preclinical testing is required. Extensive work has been performed for a number of disorders including SMA and DMD (at the level of clinical trials), and Hutchinson-Gilford progeria syndrome (19) and type I Usher syndrome (20) (using animal models). These studies have shown that AONs can enhance expression of the gene of interest following various routes of administration. It is important to consider the disease-specific aspects for AON-based treatment options. For instance, in the case of DMD, the maximal effect that can be reached using an exon skipping strategy is the milder Becker's Muscular Dystrophy (BMD) phenotype due to the expression of a truncated dystrophin protein, and this is still a serious condition. In addition, in DMD or BMD, not only skeletal muscle is affected, but also cardiac muscle and the CNS, which are more difficult to target using AONs. Cellular uptake can be further enhanced by conjugation with cell penetrating peptides or octaguanidine dendrimers, which have been applied to PMO-based AONs and showed good efficacy in animal models (45-47). In contrast to the situation in DMD, AONs in Pompe disease have the potential to restore wild type *GAA* expression towards levels present in healthy individuals. In the childhood/adult form of the disease, skeletal muscle cells are the major cells affected without involvement of cardiac or neuronal cells (48,49). This contrasts with classic infantile Pompe disease, in which hypertrophic cardiomyopathy is present at birth (26), and cognitive decline may progressively develop (50,51). The present study identifies AONs for childhood/adult onset Pompe patients. Further studies are required to test the efficacy of these AONs to target skeletal muscle cells *in vitro* and *in vivo*. Interestingly, a recent study used AONs with a PMO backbone coupled to an arginine-rich cell penetrating peptide to target muscle glycogen synthase in a mouse model for Pompe disease, suggesting that it is feasible to use AONs for Pompe disease *in vivo* (47). The ongoing development of methods to enhance tissue delivery of AONs *in vivo* by changing the chemistry (52), charge (53), by

coupling to cell penetrating peptides (21-23,45-47), or by using carriers such as exosomes or nanoparticles (54,55), is expected to further stimulate clinical testing of AONs. Current knowledge from recent clinical trials on the safety of AON backbones should help to facilitate further development of AONs, and to test whether AONs provides a valuable alternative or addition to ERT in Pompe disease.

MATERIAL & METHODS

Materials

Bovine serum albumin was purchased from GE Healthcare (UT, USA). Penicillin/Streptomycin/Glutamine (p/s/g) and TryplE were purchased from Thermo Fisher Scientific (MA, USA). DMEM High Glucose was purchased from Lonza (CH). DMSO Endoporter reagent and PMO-based AONs were purchased from Gene-Tools (OR, USA) and all other chemicals were purchased from Sigma Aldrich (UK) unless otherwise stated.

Patients

Patients were diagnosed with Pompe disease at the Center for Lysosomal and Metabolic Diseases of the Erasmus MC, Rotterdam, The Netherlands. Diagnosis was based on GAA enzymatic activity in leukocytes and/or fibroblasts, *GAA* variants, and clinical symptoms. Analysis was performed on anonymous patient material using informed consent.

Nomenclature

All references towards locations of variants and splice sites are made according to HGVS standards (http://www.hgvs.org/mutnomen/) (56). The reference *GAA* transcript was RefSeq NM_000152.3.

Generation of minigene and cDNA construct

For generation of the minigene containing the genomic *GAA* DNA region of *GAA* exon 15-17 (chr17:80113219-80117749, GRCh38/hg38), genomic DNA from a healthy control was amplified with PFU Ultra Hotstart polymerase (Agilent Technologies) and cloned in the pcDNA3.1(-)Myc-His A vector using the XbaI and NotI restriction sites. The c.2190-345A>G variant was introduced using the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies). The *GAA* cDNA expression construct (RefSeq NM_000152.3) was generated in the same vector, using restriction sites NheI and AflII, and the c.1256A>T variant was introduced using site directed mutagenesis as above. All constructs were verified by sequence analysis (all primers used in Supplementary Table S2).

Cell culture and transfections

HEK293T cells and primary human fibroblasts were cultured in 10% FBS, 1x p/s/g and

DMEM High Glucose. Transfection of minigenes and cDNA constructs was performed using Lipofectamine 2000 according to the manufactures' protocol. Cells were harvested 48 hours after transfection. *GAA* cDNA expression was corrected for mRNA expression of the *Neomycin* cassette present on the pcDNA3.1 backbone using RT-qPCR analysis. Transfection of AONs was performed using 4,5 µl/ml endoporter in the medium at a concentration of 20 µM AON. RNA and protein were harvested 3 and 5 days after transfection, respectively.

Splicing assay

The splicing assay was performed as described before (15). In short, RNA was isolated using the RNAeasy miniprep kit (Qiagen). RT-PCR was performed with 800 ng RNA input using iScript (Biorad) and FastStart Taq Polymerase (Roche). qPCR was carried out using iTaq SYBR green supermix (Biorad), and was performed on a cfx96rts cycler (Biorad). Primers are shown in Supplementary Table S2. ß-*Actin* was used as a reference gene. All primer sets used showed high efficiency and specificity based on melting-curve analysis and standard curve measurements.

In silico splice prediction

In silico prediction was carried out using Alamut Visual version 2.6.1, which uses five algorithms for predicting 5' and 3' splice junctions (description of algorithms at http://www.interactive-biosoftware.com/doc/alamut-visual/2.6/splicing.html) as described (15).

GAA enzymatic activity

Enzymatic activity of the GAA protein was determined as previously described (33). In short, cell lysates were incubated with 4-methyl-umbelliferyl-α-d-glucopyranoside (4-MU, Sigma) in citrate–phosphate buffer for 1 h at 37°C, after which fluorescence was measured at 365/448 nm with the Varioskan system (Thermo Fisher). The GAA enzymatic activity is dependent on cell culture conditions. To correct for this, a control cell line was included in each experiment, and the GAA enzymatic activity was normalized based on this control.

ACKNOWLEDGMENTS

We would like to thank Laurieke Harlaar and Pieter A. van Doorn (Department of Neurology, Erasmus MC, Rotterdam, The Netherlands)) for biopsies and clinical evaluation. This work was supported by the Sophia Children's Hospital Foundation [grant number S-687]. ATvdP has provided consulting services for various industries in the field of Pompe disease under an agreement between these industries and Erasmus MC, Rotterdam, the Netherlands. All the other authors declare no conflict of interest.

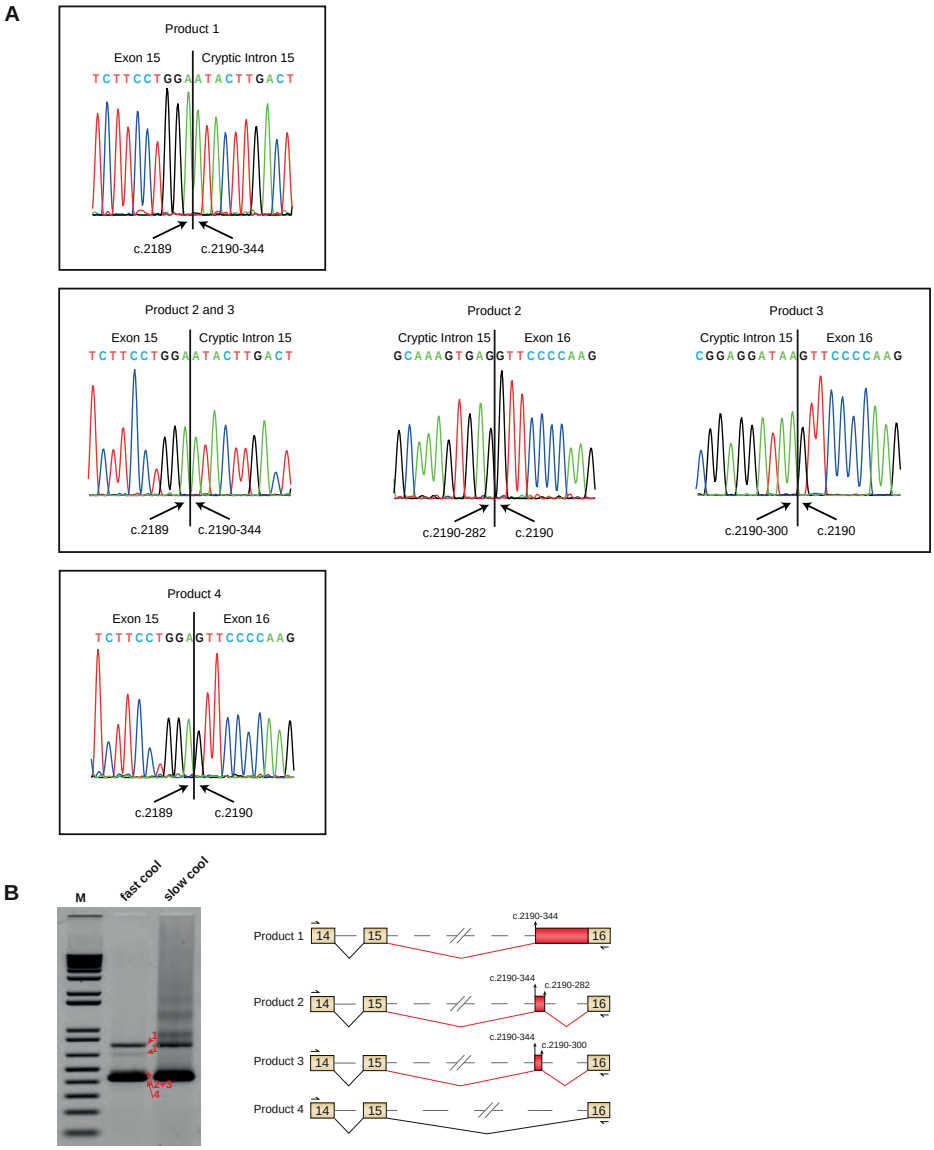
REFERENCES

1. Stenson, P.D., Mort, M., Ball, E.V., Shaw, K., Phillips, A. and Cooper, D.N. (**2014**) The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet,* 133, 1-9.
2. Lim, K.H., Ferraris, L., Filloux, M.E., Raphael, B.J. and Fairbrother, W.G. (**2011**) Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc Natl Acad Sci U S A,* 108, 11093-11098.
3. Soukarieh, O., Gaildrat, P., Hamieh, M., Drouet, A., Baert-Desurmont, S., Frebourg, T., Tosi, M. and Martins, A. (**2016**) Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. *PLoS Genet,* 12, e1005756.
4. Sterne-Weiler, T., Howard, J., Mort, M., Cooper, D.N. and Sanford, J.R. (**2011**) Loss of exon identity is a common mechanism of human inherited disease. *Genome Res,* 21, 1563-1571.
5. Scotti, M.M. and Swanson, M.S. (**2016**) RNA mis-splicing in disease. *Nat Rev Genet,* 17, 19-32.
6. Daguenet, E., Dujardin, G. and Valcarcel, J. (**2015**) The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing inform novel therapeutic approaches. *EMBO Rep,* 16, 1640-1655.
7. Chabot, B. and Shkreta, L. (**2016**) Defective control of pre-messenger RNA splicing in human disease. *J Cell Biol,* 212, 13-27.
8. Krawczak, M., Reiss, J. and Cooper, D.N. (**1992**) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet,* 90, 41-54.
9. Huie, M.L., Chen, A.S., Tsujino, S., Shanske, S., DiMauro, S., Engel, A.G. and Hirschhorn, R. (**1994**) 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.
10. Aten, E., Sun, Y., Almomani, R., Santen, G.W., Messemaker, T., Maas, S.M., Breuning, M.H. and den Dunnen, J.T. (**2013**) Exome sequencing identifies a branch point variant in Aarskog-Scott syndrome. *Hum Mutat,* 34, 430-434.
11. Cartegni, L. and Krainer, A.R. (**2002**) Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat Genet,* 30, 377-384.
12. Kashima, T. and Manley, J.L. (**2003**) A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nat Genet,* 34, 460-463.
13. Lee, Y. and Rio, D.C. (**2015**) Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu Rev Biochem, 84, 291-323.
14. Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R. and Blencowe, B.J. (**2013**) Dynamic integration of splicing within gene regulatory pathways. *Cell,* 152, 1252-1269.
15. Bergsma, A.J., Kroos, M., Hoogeveen-Westerveld, M., Halley, D., van der Ploeg, A.T. and Pijnappel, W.W. (**2015**) Identification and characterization of aberrant GAA pre-mRNA splicing in pompe disease using a generic approach. *Hum Mutat,* 36, 57-68.
16. McClorey, G. and Wood, M.J. (**2015**) An overview of the clinical application of antisense oligonucleotides for RNA-targeting therapies. *Curr Opin Pharmacol,* 24, 52-58.
17. Singh, N.N., Lee, B.M., DiDonato, C.J. and Singh, R.N. (**2015**) Mechanistic principles of antisense targets for the treatment of spinal muscular atrophy. *Future Med Chem,* 7, 1793-1808.
18. Koo, T. and Wood, M.J. (**2013**) Clinical trials using antisense oligonucleotides in duchenne muscular dystrophy. *Hum Gene Ther,* 24, 479-488.
19. Osorio, F.G., Navarro, C.L., Cadinanos, J., Lopez-Mejia, I.C., Quiros, P.M., Bartoli, C., Rivera, J., Tazi, J., Guzman, G., Varela, I. et al. (**2011**) Splicing-directed therapy in a new mouse model of human accelerated aging. *Sci Transl Med,* 3, 106ra107.
20. Lentz, J.J., Jodelka, F.M., Hinrich, A.J., McCaffrey, K.E., Farris, H.E., Spalitta, M.J., Bazan, N.G., Duelli, D.M., Rigo, F. and Hastings, M.L. (2013) Rescue of hearing and vestibular function by antisense oligonucleotides in a mouse model of human deafness. *Nat Med,* 19, 345-350.
21. Lehto, T., Ezzat, K., Wood, M.J. and El Andaloussi, S. (**2016**) Peptides for nucleic acid delivery. *Adv Drug Deliv Rev.*
22. Abushahba, M.F., Mohammad, H., Thangamani, S., Hussein, A.A. and Seleem, M.N. (**2016**) Impact of different cell penetrating peptides on the efficacy of antisense therapeutics for targeting intracellular pathogens. *Sci Rep,* 6, 20832.

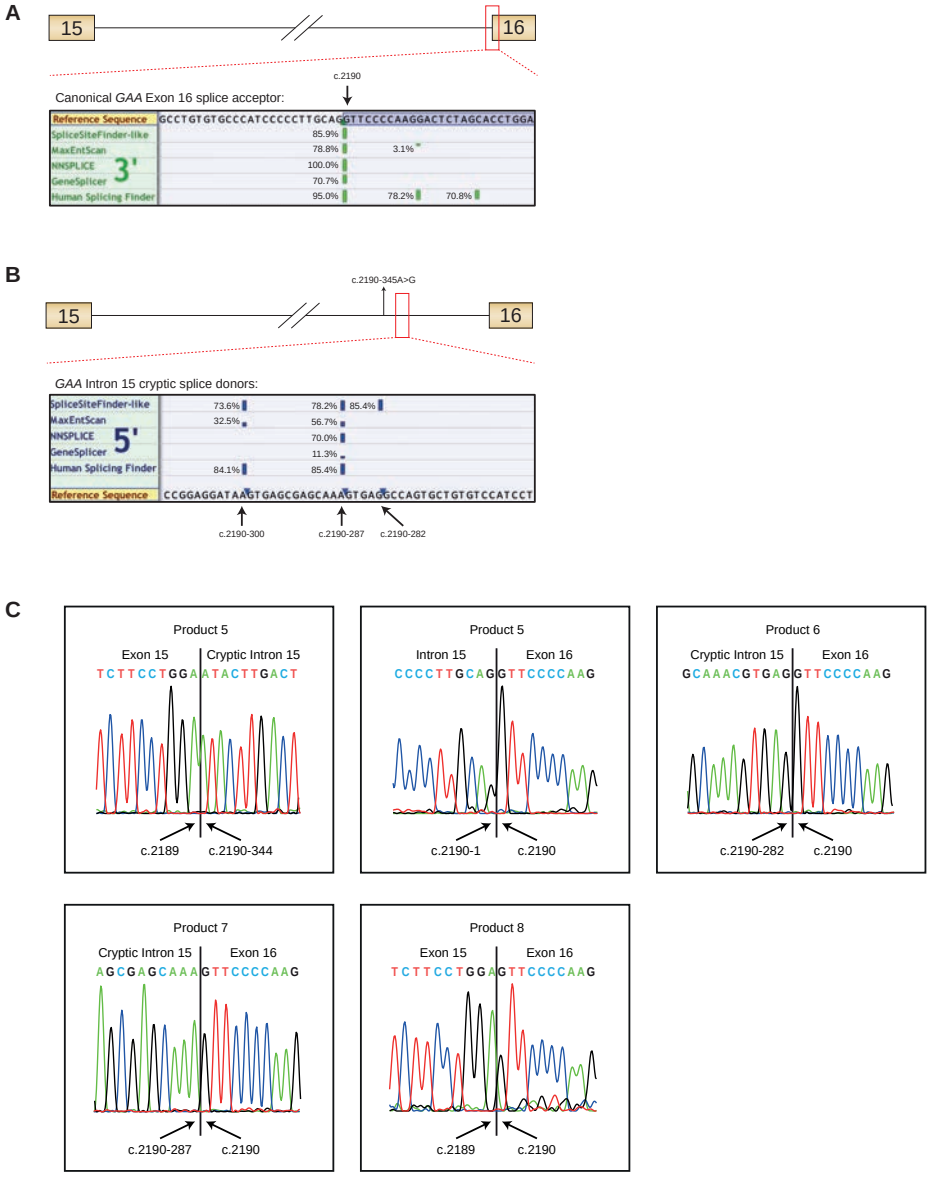
23. Boisguerin, P., Deshayes, S., Gait, M.J., O'Donovan, L., Godfrey, C., Betts, C.A., Wood, M.J. and Lebleu, B. (**2015**) Delivery of therapeutic oligonucleotides with cell penetrating peptides. *Adv Drug Deliv Rev,* 87, 52-67.
24. Han, G., Gu, B., Cao, L., Gao, X., Wang, Q., Seow, Y., Zhang, N., Wood, M.J. and Yin, H. (**2016**) Hexose enhances oligonucleotide delivery and exon skipping in dystrophin-deficient mdx mice. *Nat Commun,* 7, 10981.
25. Palhais, B., Praestegaard, V.S., Sabaratnam, R., Doktor, T.K., Lutz, S., Burda, P., Suormala, T., Baumgartner, M., Fowler, B., Bruun, G.H. et al. (**2015**) Splice-shifting oligonucleotide (SSO) mediated blocking of an exonic splicing enhancer (ESE) created by the prevalent c.903+469T>C MTRR mutation corrects splicing and restores enzyme activity in patient cells. *Nucleic Acids Res,* 43, 4627-4639.
26. van der Ploeg, A.T. and Reuser, A.J. (2008) Pompe's disease. Lancet, 372, 1342-1353.
27. Broomfield, A., Fletcher, J., Davison, J., Finnegan, N., Fenton, M., Chikermane, A., Beesley, C., Harvey, K., Cullen, E., Stewart, C. et al. (**2016**) Response of 33 UK patients with infantile-onset Pompe disease to enzyme replacement therapy. *J Inherit Metab Dis,* 39, 261-271.
28. Kishnani, P.S. and Beckemeyer, A.A. (**2014**) New therapeutic approaches for Pompe disease: enzyme replacement therapy and beyond. *Pediatr Endocrinol Rev,* 12 Suppl 1, 114-124.
29. van Gelder, C.M., Poelman, E., Plug, I., Hoogeveen-Westerveld, M., van der Beek, N.A., Reuser, A.J. and van der Ploeg, A.T. (**2016**) Effects of a higher dose of alglucosidase alfa on ventilator-free survival and motor outcome in classic infantile Pompe disease: an open-label single-center study. *J Inherit Metab Dis.*
30. Toscano, A. and Schoser, B. (**2013**) Enzyme replacement therapy in late-onset Pompe disease: a systematic literature review. *J Neurol,* 260, 951-959.
31. Boerkoel, C.F., Exelbert, R., Nicastri, C., Nichols, R.C., Miller, F.W., Plotz, P.H. and Raben, N. (**1995**) Leaky splicing mutation in the acid maltase gene is associated with delayed onset of glycogenosis type II. A*m J Hum Genet,* 56, 887-897.
32. Dardis, A., Zanin, I., Zampieri, S., Stuani, C., Pianta, A., Romanello, M., Baralle, F.E., Bembi, B. and 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.
33. Kroos, M.A., Pomponio, R.J., Hagemans, M.L., Keulemans, J.L., Phipps, M., DeRiso, M., Palmer, R.E., Ausems, M.G., Van der Beek, N.A., Van Diggelen, O.P. et al. (**2007**) Broad spectrum of Pompe disease in patients with the same c.-32-13T->G haplotype. *Neurology,* 68, 110-115.
34. Ittig, D., Liu, S., Renneberg, D., Schumperli, D. and Leumann, C.J. (**2004**) Nuclear antisense effects in cyclophilin A pre-mRNA splicing by oligonucleotides: a comparison of tricyclo-DNA with LNA. *Nucleic Acids Res,* 32, 346-353.
35. Kroos, M., Manta, P., Mavridou, I., Muntoni, F., Halley, D., Van der Helm, R., Zaifeiriou, D., Van der Ploeg, A., Reuser, A. and Michelakakis, H. (**2006**) Seven cases of Pompe disease from Greece. *J Inherit Metab Dis,* 29, 556-563.
36. Merkin, J., Russell, C., Chen, P. and Burge, C.B. (**2012**) Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science,* 338, 1593-1599.
37. Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R. et al. (**2012**) The evolutionary landscape of alternative splicing in vertebrate species. *Science,* 338, 1587-1593.
38. Carrillo Oesterreich, F., Herzel, L., Straube, K., Hujer, K., Howard, J. and Neugebauer, K.M. (**2016**) Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell,* 165, 372-381.
39. Orkunoglu-Suer, F., Harralson, A.F., Frankfurter, D., Gindoff, P. and O'Brien, T.J. (**2015**) Targeted single molecule sequencing methodology for ovarian hyperstimulation syndrome. *BMC Genomics,* 16, 264.
40. Reuser, A.J., Kroos, M.A., Hermans, M.M., Bijvoet, A.G., Verbeet, M.P., Van Diggelen, O.P., Kleijer, W.J. and Van der Ploeg, A.T. (**1995**) Glycogenosis type II (acid maltase deficiency). *Muscle Nerve Suppl,* 3, S61-69.
41. Kroos, M., Hoogeveen-Westerveld, M., van der Ploeg, A. and Reuser, A.J. (**2012**) The genotype-phenotype correlation in Pompe disease. *Am J Med Genet C Semin Med Genet,* 160C, 59-68.
42. Herzog, A., Hartung, R., Reuser, A.J., Hermanns, P., Runz, H., Karabul, N., Gokce, S., Pohlenz, J., Kampmann, C., Lampe, C. et al. (**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.

43. Wens, S.C., van Gelder, C.M., Kruijshaar, M.E., de Vries, J.M., van der Beek, N.A., Reuser, A.J., van Doorn, P.A., van der Ploeg, A.T. and Brusse, E. (**2013**) Phenotypical variation within 22 families with Pompe disease. *Orphanet J Rare Dis,* 8, 182.
44. Aartsma-Rus, A., van Vliet, L., Hirschi, M., Janson, A.A., Heemskerk, H., de Winter, C.L., de Kimpe, S., van Deutekom, J.C., t Hoen, P.A. and van Ommen, G.J. (**2009**) Guidelines for antisense oligonucleotide design and insight into splice-modulating mechanisms. *Mol Ther,* 17, 548-553.
45. Betts, C., Saleh, A.F., Arzumanov, A.A., Hammond, S.M., Godfrey, C., Coursindel, T., Gait, M.J. and Wood, M.J. (**2012**) Pip6-PMO, A New Generation of Peptide-oligonucleotide Conjugates With Improved Cardiac Exon Skipping Activity for DMD Treatment. *Mol Ther Nucleic Acids,* 1, e38.
46. Widrick, J.J., Jiang, S., Choi, S.J., Knuth, S.T. and Morcos, P.A. (**2011**) An octaguanidine-morpholino oligo conjugate improves muscle function of mdx mice. *Muscle Nerve, 44,* 563-570.
47. Clayton, N.P., Nelson, C.A., Weeden, T., Taylor, K.M., Moreland, R.J., Scheule, R.K., Phillips, L., Leger, A.J., Cheng, S.H. and Wentworth, B.M. (**2014**) Antisense Oligonucleotide-mediated Suppression of Muscle Glycogen Synthase 1 Synthesis as an Approach for Substrate Reduction Therapy of Pompe Disease. *Mol Ther Nucleic Acids,* 3, e206.
48. van der Beek, N.A., de Vries, J.M., Hagemans, M.L., Hop, W.C., Kroos, M.A., Wokke, J.H., de Visser, M., van Engelen, B.G., Kuks, J.B., van der Kooi, A.J. et al. (**2012**) Clinical features and predictors for disease natural progression in adults with Pompe disease: a nationwide prospective observational study. *Orphanet J Rare Dis,* 7, 88.
49. Wens, S.C., Schaaf, G.J., Michels, M., Kruijshaar, M.E., van Gestel, T.J., In 't Groen, S., Pijnenburg, J., Dekkers, D.H., Demmers, J.A., Verdijk, L.B. et al. (**2016**) Elevated Plasma Cardiac Troponin T Levels Caused by Skeletal Muscle Damage in Pompe Disease. *Circ Cardiovasc Genet,* 9, 6-13.
50. Ebbink, B.J., Poelman, E., Plug, I., Lequin, M.H., van Doorn, P.A., Aarsen, F.K., van der Ploeg, A.T. and van den Hout, J.M. (**2016**) Cognitive decline in classic infantile Pompe disease: An underacknowledged challenge. *Neurology,* 86, 1260-1261.
51. Ebbink, B.J., Aarsen, F.K., van Gelder, C.M., van den Hout, J.M., Weisglas-Kuperus, N., Jaeken, J., Lequin, M.H., Arts, W.F. and van der Ploeg, A.T. (**2012**) Cognitive outcome of patients with classic infantile Pompe disease receiving enzyme therapy. *Neurology,* 78, 1512-1518.
52. Goyenvalle, A., Griffith, G., Babbs, A., El Andaloussi, S., Ezzat, K., Avril, A., Dugovic, B., Chaussenot, R., Ferry, A., Voit, T. et al. (**2015**) Functional correction in mouse models of muscular dystrophy using exon-skipping tricyclo-DNA oligomers. *Nat Med,* 21, 270-275.
53. Iversen, P.L., Warren, T.K., Wells, J.B., Garza, N.L., Mourich, D.V., Welch, L.S., Panchal, R.G. and Bavari, S. (**2012**) Discovery and early development of AVI-7537 and AVI-7288 for the treatment of Ebola virus and Marburg virus infections. *Viruses,* 4, 2806-2830.
54. Liu, R., Liu, J., Ji, X. and Liu, Y. (**2013**) Synthetic nucleic acids delivered by exosomes: a potential therapeutic for generelated metabolic brain diseases. *Metab Brain Dis,* 28, 551-562.
55. Cheng, X. and Lee, R.J. (**2016**) The role of helper lipids in lipid nanoparticles (LNPs) designed for oligonucleotide delivery. *Adv Drug Deliv Rev.*
56. den Dunnen, J.T., Dalgleish, R., Maglott, D.R., Hart, R.K., Greenblatt, M.S., McGowan-Jordan, J., Roux, A.F., Smith, T., Antonarakis, S.E., Taschner, P.E. et al. (**2016**) HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat.*

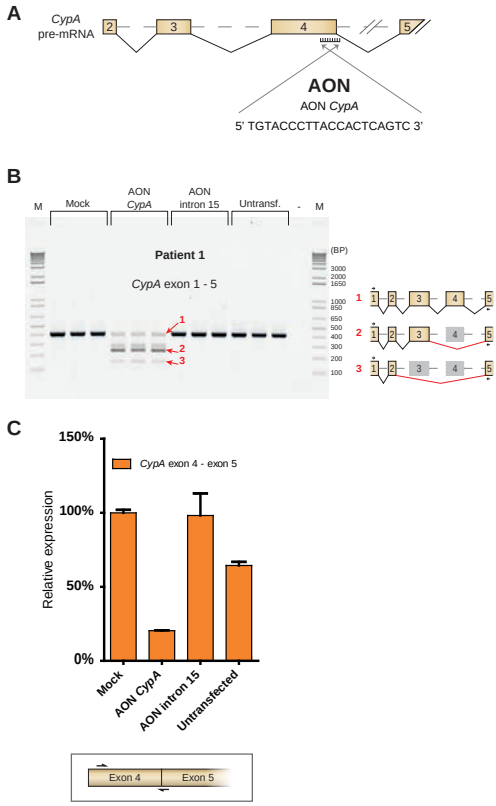
SUPPLEMENTARY FIGURES AND FIGURE LEGENDS



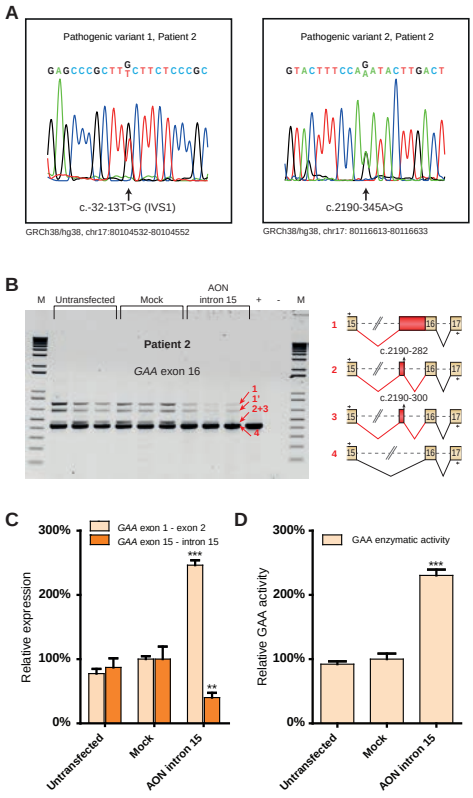
Supplementary Figure 1. Sequence analysis and structural effects of GAA splicing products 1-4 identified in cells from patient 1. *(A) Sequence analysis of splice junctions utilized in splice products 1-4, shown in Fig 1C. (B) Effect of the speed of cooling on formation of secondary structure in PCR products. A sample from the flanking exon RT-PCR analysis of GAA exon 16 on primary fibroblasts from patient 1 is shown. Left lane: fast cooling, right lane: slow cooling. Product 1' disappeared as a consequence of slow cooling, suggesting that it represents a conformational variant. Cartoons of products 1-4 are shown in the right panel.*



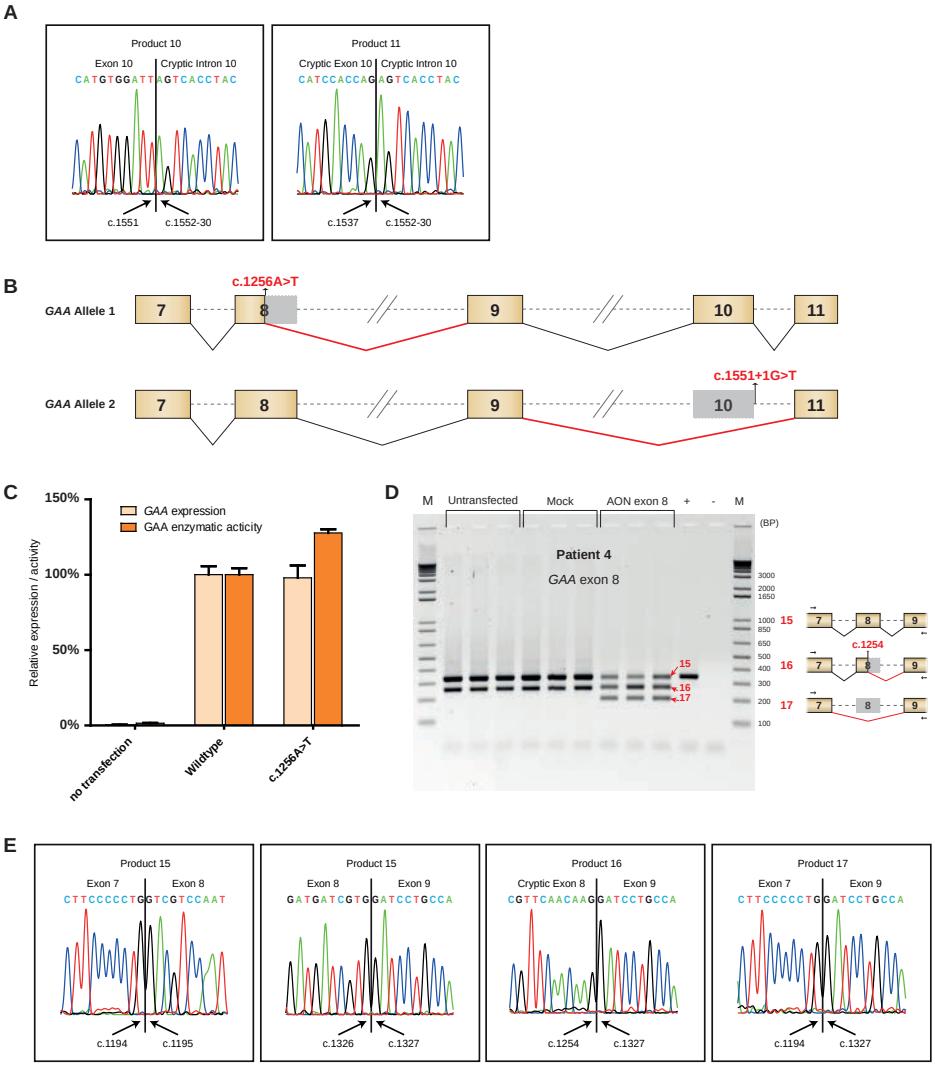
Supplementary Figure 2. In silico prediction and sequence analysis of GAA splicing products 5-8 in cells from patient 1. *(A) In silico prediction of the 3' canonical splice acceptor site located at the GAA intron 15 – exon 16 junction. (B) In silico prediction of the 5' cryptic splice donor sites located in GAA intron 15 that were found to be utilized in the context of the c.2190-345A>G variant. (C) Sequence analysis of splice junctions utilized in products 5-8, shown in Fig 2D.*



Supplementary Figure 3. Exon skipping in CypA pre-mRNA by an AON. *(a) Cartoon depicting the region in the CypA pre-mRNA that was targeted with an AON. The sequence of AON CypA is shown. (b) Flanking exon RT-PCR analysis of fibroblasts from patient 1 treated with AONs. Cartoons depict spliced mRNAs. Primer locations are indicated. Three biological replicates are shown. (v) RT-qPCR analysis of the experiment in (b), using primers spanning the CypA exon 4-5 splice junction (shown in the lower panel). Data are normalized for ß-Actin and for mock transfection and represent means +/- SD of three biological replicates.*



Supplementary Figure 4. AON intron 15 restores GAA splicing in cells from patient 2, a sibling of patient 1. *(A) Sequence analysis of genomic DNA obtained from primary fibroblasts of patient 2. (B) Flanking exon RT-PCR analysis of exon 16 in primary fibroblasts from patient 2 treated with AON intron 15. Products were inferred from the analysis of patient 1 based on electrophoretic mobility, and these are indicated in the cartoons along with primer locations in the right panel. Three biological replicates are shown. (C) RT-qPCR analysis of the experiment in (B). (D) Effect of AON intron 15 on GAA enzymatic activity in fibroblasts from patient 2. Data in (C) and (D) represent means +/- SD of three biological replicates (** p = 0,01 , *** p = 0,001).*


A
Product 10
Exon 10
Cryptic Intron 10
CATGTGGATTTAGTCACCTAC
c.1551
c.1552-30
Product 11
Cryptic Exon 10
Cryptic Intron 10
CATCCACCAGAGTCACCTAC
c.1537
c.1552-30
B
c.1256A>T
GAA Allele 1
7
8
9
10
11
c.1551+1G>T
GAA Allele 2
C
150%
100%
50%
0%
GAA expression
GAA enzymatic acticity
Relative expression / activity
no transfection
Wildtype
c.1256A>T
D
M
Untransfected
Mock
AON exon 8
+
-
M
(BP)
Patient 4
GAA exon 8
3000
2000
1650
1000
850
650
500
400
300
200
100
15
16
17
c.1254
E
Product 15
Exon 7
Exon 8
CTTCCCCCTGGTCGTCCAAT
c.1194
c.1195
Product 15
Exon 8
Exon 9
GATGATCGTGGATCCTGCCA
c.1326
c.1327
Product 16
Cryptic Exon 8
Exon 9
CGTTCAACAAGGATCCTGCCA
c.1254
c.1327
Product 17
Exon 7
Exon 9
CTTCCCCCTGGATCCTGCCA
c.1194
c.1327

Supplementary Figure 5. Further analysis of patients 3 and 4. *(A) Sequence analysis of newly identified products 10 and 11 derived from patient 3 (see Fig. 4B). (B) Cartoons of the GAA alleles of patient 4. (C) Effect of the c.1256A>T variant on GAA cDNA mRNA expression and enzymatic activity. Wild type or mutant cDNA constructs were transfected in HEK293T cells. GAA enzymatic activity was determined using 4-MU as substrate. GAA mRNA expression was also determined (using RT-qPCR analysis) to verify that both constructs were expressed at similar levels, which was the case. Neomycin mRNA expression (using RT-qPCR analysis) was used to normalized for transfection efficiency. Data are expressed relative to wild type GAA mRNA expression/activity. Both expression and protein activity data represent means +/- S.D. of three biological replicates. (D) Flanking exon RT-PCR of patient 4 GAA exon 8 after treatment. Cartoons depict spliced mRNAs. Primer locations are indicated. Three biological replicates are shown. + represents cDNA from a healthy control. Flanking exon RT-PCR analysis of GAA exon 8 showed that AON exon 8 had multiple effects on exon 8 splicing, resulting in similarly abundant expression of three products: mRNA with canonically spliced exon 8 (product 15), mRNA in which the 3' part of exon 8 was skipped via utilization of the cryptic splice site at c.1254 (product 16) and complete skipping of exon 8 (product 17) (for sequence see (E)). To explain why AON exon 8 corrected splicing from the c.1256A>T allele while at the same time it also caused aberrant splicing of total GAA expression, the effects of AON exon 8 on the individual alleles should be considered. Without AON treatment, the c.1256A>T allele shows predominant skipping of the 3' part of exon 8 (assume here 95%), and a small amount of leaky wild type splicing (assume here 5%). After AON exon 8 treatment, the splicing equilibrium shifts in which expression of canonically spliced mRNA (product 15) and aberrantly spliced mRNA (products 16 and 17) are at similar levels. This results in an increase in the expression of the canonically spliced c.1256A>T allele, which has been quantified to be 2.3 fold (Fig 4G). The second allele with the c.1551+1G>T variant is responsible for the majority of canonically spliced exon 8 mRNA. However, in this mRNA, exon 10 is skipped resulting in translation of inactive GAA protein. Without AON exon 8 treatment, the c.1551+1G>T allele is the dominant allele that is preferably amplified in the flanking exon PCR of exon 8. Treatment with AON exon 8 changes the equilibrium of exon 8 splicing, but this has no functional consequences on the c.1551+1G>T allele as its product is nonfunctional in all cases. (E) Sequence analysis of products 15-17 identified in (D).*

Supplementary Table 1. Number of products with and without AON treatment of patient 1.

Product Nr.	Product type	- AON	+ AON
9	Retention intron 10	1	1
10	Utilization cryptic c.1552-30	9	3
11	Utilization cryptic c.1537 and c.1552-30	5	5
12	Canonical splicing	23	37
13	Skip exon 10 and use of cryptic c.1552-30	9	-
	total	47	46

Supplementary Table 2. gDNA: genomic DNA, SDM: Site Directed Mutagenesis.

Primer name	Sequence 5' -> 3'	purpose
GAA intron15 gDNA fw	AGGAAACAGGACAGGGCAGAG	gDNA sequencing
GAA intron15 gDNA rv	CTTGGAGGACTCAGGTCAGACG	gDNA sequencing
Minigene*GAA* ex15-17 fw	GCGCCTCGAGCCCAGGAGCCGTACAGCTT	minigene
Minigene*GAA* ex15-17 rv	GCGCGAATTCCTGCAGGGGGATGATGTACC	minigene
GAA c.2190-345A>G fw	TCGTACTTTCCAGATACTTGACTGATGAGC	SDM
GAA c.2190-345A>G rv	CATCAGTCAAGTATCTGGAAAGTACGAAAA	SDM
GAA exon 15-17 minigene fw	GAGCCGTACAGCTTCAGCGA	RT-PCR
GAA exon 15-17minigene rv	TAGAAGGCACAGTCGAGG	RT-PCR
GAA Exon 15-17 fw	GAGCCGTACAGCTTCAGCGA	RT-PCR
GAA Exon 15-17 rv	ATGTACCCAGCCCGGAGGT	RT-PCR
CyPA Exon 1-5 fw	CACCGTGTTCTTCGACATTG	RT-PCR
CyPA Exon 1-5 rv	CCATGGCCTCCACAATATTC	RT-PCR
CyPA Exon 4-5 fw	GGACCCAACACAAATGGTTC	qPCR
CyPA Exon 4-5 rv	GGCCTCCACAATATTCATGC	qPCR
GAA Exon 1-2 fw	AAACTGAGGCACGGAGCG	qPCR
GAA Exon 1-2 rv	GAGTGCAGCGGTTGCCAA	qPCR
GAA exon15mut fw	CAGCAGGCCATGAGGAAG	qPCR
GAA exon15mut rv	GCATGTGCTCATCAGTCAAGTAT	qPCR
GAA Exon 10 fw	GATCCTGCCATCAGCAGCT	RT-PCR
GAA Exon 10 rv	TGGGTTCTCCAGCTCATTGT	RT-PCR
GAA Exon 20 fw	GGGCGGAGTGTGTTAGTCTC	qPCR
GAA Exon 20 rv	CTCCAGGTGACACATGCAAC	qPCR
GAA exon8mut fw	GGCGCTACATGATGATCGTG	qPCR
GAA exon8mut rv	AGTGGACCCGGGCCATAC	qPCR
GAA Exon 9 fw	GACGTCCAGTGGAACGACCT	RT-PCR
GAA Exon 9 rv	ACCTGGTCATGGAACTCAGC	RT-PCR
*Neomycin*fw	TCATCTCACCTTGCTCCTGC	qPCR
*Neomycin*rv	GTGGTCGAATGGGCAGGTAG	qPCR

Chapter 5

GAA splicing in Pompe disease is restored by exon inclusion in expanded iPS cell-derived myotubes

Erik van der Wal[1,2,3,*]
Atze J. Bergsma[1,2,3,*]
Joon M. Pijnenburg[1,2,3]
Tom J.M. van Gestel[1,2,3]
Stijn L.M. in 't Groen[1,2,3]
Holm Zaehres[4]
Marcos J Araúzo-Bravo[4]
Hans R. Schöler[4]
Ans T. van der Ploeg[2,3]
W.W.M. Pim Pijnappel[1,2,3]

[1] Molecular Stem Cell Biology, Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, the Netherlands
[2] Department of Pediatrics, Erasmus MC University Medical Center, Rotterdam, the Netherlands
[3] Center for Lysosomal and Metabolic Diseases, Erasmus MC University Medical Center, Rotterdam, the Netherlands
[4] Department of Cell and Developmental Biology, Max Planck Institute for Molecular Biomedicine, 48149 Münster, Germany

***** equal contribution

Manuscript submitted

ABSTRACT

Few therapies for human disorders that affect skeletal muscle are available. Correction of pre-mRNA splicing using antisense oligonucleotides (AONs) is a promising therapeutic strategy. Promotion of exon inclusion in Pompe disease caused by the common *GAA* IVS1 splicing variant may provide an alternative to partial effective enzyme replacement therapy. It is poorly understood how exon inclusion can be promoted, and suitable models of patient-derived skeletal muscle cells are scarce. Here, we developed methods for promotion of exon inclusion and for *in vitro* drug testing using human iPS cell-derived myotubes. We used these to restore *GAA* splicing to healthy control levels. Myogenic progenitors were expanded to at least 10^{12} cells while maintaining myogenic differentiation capacity. A screen identified splicing silencers of *GAA* that were part of a natural intronic pseudo exon rather than classical intronic splicing silencer motifs. These findings may have general applicability to modulating exon inclusion in human disease.

INTRODUCTION

Although >700 human disorders that affect skeletal muscle are known (1), very few therapies for skeletal muscle disorders have been developed (2). A notable exception is Enzyme Replacement Therapy (ERT) for Pompe disease (3-5). Pompe disease is an autosomal recessive metabolic myopathy caused by deficiency of the lysosomal enzyme acid α-glucosidase (GAA). In ERT, recombinant human GAA (rhGAA) is supplied intravenously on a (bi)weekly basis. ERT can improve symptoms and prolong life expectancy in both classic infantile and childhood/adult onset patients (6-14). However, in childhood/adult onset patients, evidence accumulates that, on average, ERT can be particularly effective during the first years of treatment while clinical decline may occur thereafter (15,16). In addition, the response to ERT is heterogeneous, with a significant fraction of patients that respond moderately or not at all (15,17). It is poorly understood why this is the case. In adult patients, antibodies to rhGAA can occur but only in a minority of cases these may explain poor clinical outcome (9,18-20). Among the factors that may affect the response to ERT are expression and recycling of the mannose-6-phosphate receptor (M6PR) (21), responsible for cellular uptake of rhGAA, and autophagic flux, which is repressed in Pompe disease (22,23). These disadvantages of ERT led us to search for an alternative treatment option that preferentially would be based on a different mechanism of action.

A common *GAA* variant in childhood/adult Pompe disease is the splicing variant c.-13-32T>G (IVS1), which is present in 60-90% in adults and 30-50% in children with Pompe disease. The IVS1 allele causes aberrant *GAA* pre-mRNA splicing, but it also allows 10-15% leaky wild type splicing (24-27). This explains why patients with the IVS1 allele have a slower disease onset and progression compared to classic infantile patients, who have two fully deleterious alleles. The major known aberrant splicing events caused by the IVS1 allele are a complete and partial skipping of *GAA* exon 2. Because this exon contains the translation initiation codon, its skipping results in a failure to initiate protein synthesis, triggering mRNA degradation. The residual wild type splicing from the IVS1 allele suggested that it may be possible to promote exon 2 inclusion to restore wild type *GAA* splicing.

Antisense oligonucleotides (AONs) have been used previously to modulate splicing (28-31). These can be targeted to pre-mRNA to block splicing regulatory sequences. Sequences that are clearly defined can be targeted using AONs in a relatively straightforward manner. Examples include the blocking of splice junctions, either canonical splice sites or cryptic splice sites that are generated by a gene variant. Such strategies have been applied to restore or disrupt a reading frame, to switch alternative splicing, or to prevent cryptic splicing. We recently inhibited cryptic splicing caused by *GAA* variants, identified using a splicing assay, in cells from Pompe patients (27,32). However, it is not obvious how to promote exon inclusion

when there are no defined splicing sequences that can be blocked. Hence only a few of such cases have been described. In the case of SMA, exon skipping of *SMN2* exon 7 is caused by an exonic nucleotide change, and nearby exonic and intronic splicing silencer sequences (ESS and ISS, respectively) have been identified experimentally as these are difficult to predict *in silico*. These sequences could be blocked with AONs to promote exon inclusion (33-38). In the case of Pompe disease, the IVS1 variant is located in the polypyrimidine (pY) tract preceding exon 2. While splicing correction of pY tract variants have not been reported to date, it seemed possible that ESS or ISS sequences would be potential targets for AONs. A better understanding of the mechanisms involved would be valuable for the design of strategies to correct splicing using AONs that promote exon inclusion.

A drawback for the development of AONs for human skeletal muscle disorders in general is the lack of suitable *in vitro* model systems. This is relevant because target genes may be expressed in muscle cells only, splicing mechanisms can be cell-type specific (39,40), and for clinical testing human-specific AONs rather than animal-specific AONs should be tested. Primary myoblasts can be obtained from patients via a muscle biopsy, but these undergo senescence and show strongly impaired capacity to differentiate into multinucleated myotubes upon passaging (41); (our own unpublished results). Alternatively, induced pluripotent stem (iPS) cells can be generated from patient cells and these can be differentiated to skeletal muscle progenitor cells (42-46). However, up to now it has not been possible to expand purified myogenic cells to large cell numbers while maintaining the capacity for efficient *in vitro* differentiation into multinucleated myotubes. Such properties would greatly facilitate *in vitro* studies on the testing of AONs or other drugs on multinucleated myotubes to design treatments for human skeletal muscle disorders. We have developed a screen that identified cis-acting sequences whose blocking using AONs promoted exon inclusion by interfering with a natural pseudo exon. We also developed an *in vitro* system for the testing of AONs in patient-derived multinucleated myotubes. The AONs restored 66-99% of aberrant splicing caused by the IVS1 *GAA* variant via simultaneous inhibition of 3' and 5' splice sites of the natural pseudo exon. iPS-derived myogenic cells could be expanded to at least 10^{12} cells while maintaining the capacity to form multinucleated myotubes. These findings provide an unbiased and fast method for the identification of AONs that promote exon inclusion, and a method for *in vitro* testing of AONs and other drugs for skeletal muscle disorders. AONs that correct splicing from the IVS1 variant are candidate drugs for an alternative therapy for childhood/adult onset Pompe disease. We hypothesize that, besides splicing silencers, natural cryptic splice sites that may be part of natural pseudo exons play an unanticipated role in aberrant splicing in human disease.

RESULTS

Our purpose was to promote *GAA* exon 2 inclusion in cells from IVS1 patients to restore wild type *GAA* splicing. Primary fibroblasts from such patients show partial and complete skipping of exon 2 (Figure 1a), as reported previously (24-27). We aimed to block a splicing silencer sequence using AONs. However, no splicing silencer sequences have been described so far for *GAA*. To identify silencers of exon 2 splicing, *in silico* prediction analysis was performed using Human Splicing Finder (http://www.umd.be/HSF/) (Suppl. Figure S1a). This yielded many possible hits that failed to overlap between different prediction algorithms, and it was unclear which hits should be used to design and synthesize rather expensive chemically stable AONs. This indicated the need to screen the *GAA* pre-mRNA for possible splicing regulatory motifs (Figure 1b) in a functional and cost-effective assay.

We used modified *U7* snRNA to express AONs as shown previously (47,48). This enables the expression of AONs in the nucleus that are stabilized by a stem loop that is provided by the snRNA (Figure 1b). We aimed to test endogenous *GAA* splicing in primary cells, as these would be the closest to splicing regulation *in vivo*. Patient-derived primary fibroblasts, obtained via a skin biopsy, are routinely used for biochemical diagnosis of Pompe disease. GAA enzymatic activities of 1-20% of healthy control values indicate childhood/adult onset Pompe disease. Transfection of *U7* snRNA expression constructs in primary cells was inefficient, preventing efficient modulation of endogenous splicing (data not shown). We therefore cloned the *U7* snRNA cassette in a lentivirus and used lentiviral transduction, which resulted in ~100% transduction efficiency of primary fibroblasts. This vector was then modified by introduction of a NsiI site to allow 1-step cloning of AONs, introduced via a forward PCR primer, with a cloning success rate of >95% (Suppl. Figure S1b). We validated the lentiviral *U7* snRNA system by promoting exon skipping of a control gene, *cyclophilin A* (*CypA*) (47) in primary fibroblasts (Suppl. Figure S1c-e). We conclude that AONs expressed as *U7* snRNAs using a lentivirus provides a fast and cheap method to identify putative target sites for splice-switching AONs in primary cells.

A screen was then performed in primary fibroblasts from Pompe patient 1 (genotype IVS1, c.525delT; the second allele is not expressed), in which AONs targeted the *GAA* pre-mRNA surrounding the IVS1 variant in a non-overlapping tiling arrangement, from c.-32–319 to c.530 (Figure 1c). Three read outs were used: *GAA* mRNA expression by RT-qPCR and flanking exon PCR, and GAA enzyme activity (Figure 1d,e). This resulted in the identification of two regions in intron 1 (c.-32-219 and c.-32-179) that acted as splicing silencer sequences and whose repression by AONs promoted exon 2 inclusion and GAA enzyme activity. Lentiviral-mediated *U7* snRNA expression appeared to have a small window in which splicing modulation could be investigated, due to toxicity at high virus titers (Suppl. Figure S1f). We then performed a miniscreen around these targets using AONs that shifted 2 nt each, and this defined c.-32-219 and c.-32-179 as the peaks of the regions that acted as silencers of *GAA* exon 2 splicing (Suppl. Figure S1g-i).

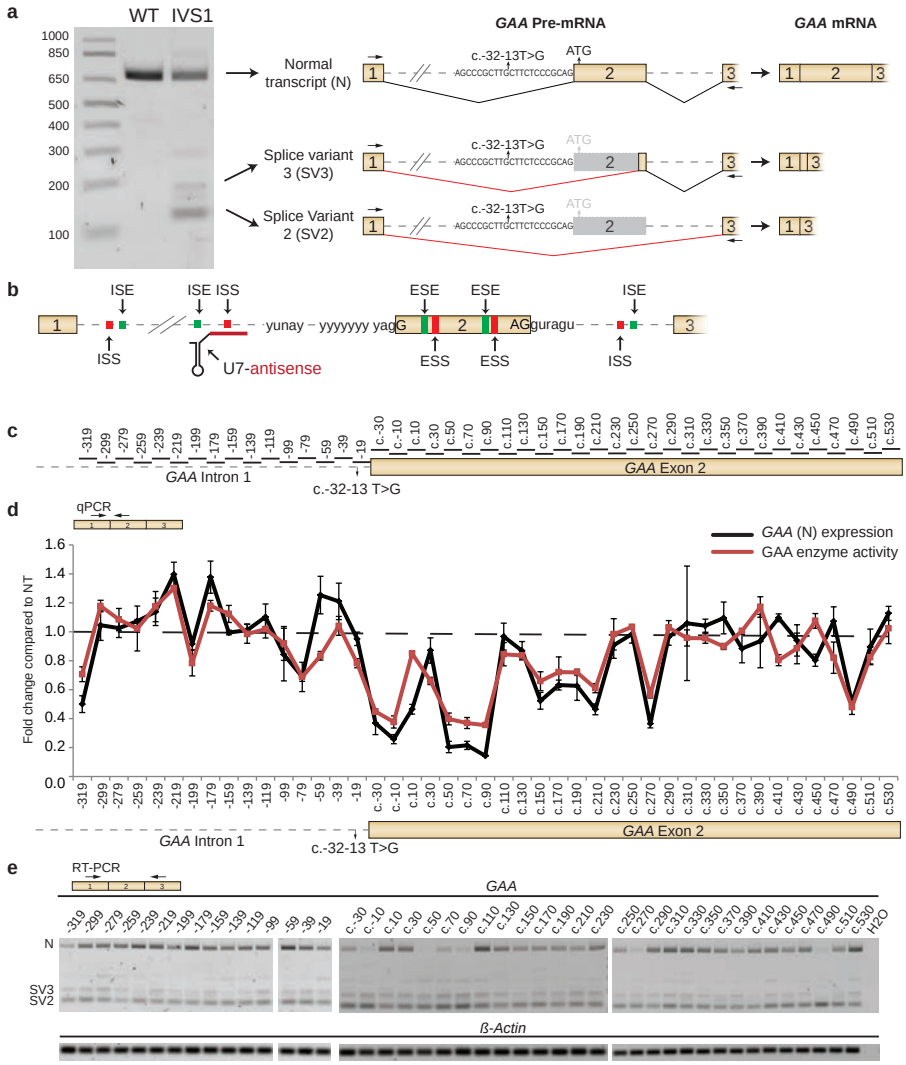


Figure 1. Screen to identify silencers of GAA exon 2 splicing. *(a) Outline of the three major splicing products of the GAA pre-mRNA caused by the IVS1 variant in patient-derived primary fibroblasts known to date. The gel illustrates the results of flanking exon RT-PCR analysis of exon 2 using primers that anneal to exon 1 and exon 3. WT: control fibroblasts; IVS1: fibroblasts from patient 1. Left lane: DNA size markers (in basepairs). Cartoons of pre-mRNAs illustrate splicing events as described (24-27). The location of the c.-32-13C>T (IVS1) variant in the pY tract is indicated. Spliced mRNA cartoons are shown on the far right. Sizes of introns and exons in the cartoon are to scale. (b) Cartoon showing hypothetical splicing regulatory elements that may be subject to modulation e.g. by a U7 snRNA. (c) Locations of U7 snRNA-based AONs used in the screen in (d). (d) Screen to identify splicing silencers of GAA exon 2. Primary fibroblasts from patient 1 (IVS1, c.525delT) were transduced with 200 ng U7 snRNA-expressing lentiviruses. The effects on GAA exon 2 expression were measured using RT-qPCR (black line; GAA (N) expression; primers are indicated in the upper left cartoon). Effects on GAA enzymatic activity are indicated by the red line. The cartoon of GAA pre-mRNA below the graph indicates the positions of the AONs tested. Data are expressed relative to non-transduced (NT) fibroblasts and represent means +/- SD of three biological replicates. Samples were normalized for β-Actin expression. (e) The experiment of (d) was also analyzed by flanking exon RT-PCR of GAA exon 2. β-Actin mRNA was used as loading control. Primers are indicated in the upper left cartoon. *P < 0.05 and **P < 0.01 (n = 3).*

To explore the possibility for the development of AONs that could be used in a clinical setting, we used phosphorodiamidate morpholino oligomer (PMO)-based AONs. These and related AONs are chemically stable and are currently tested in clinical trials for a number of diseases. In a validation experiment, exon 4 of *CypA* was efficiently skipped using AONs *CypA* 1 and *CypA* 2 that targeted the splice acceptor (Suppl. Figure S2a-d). No signs of toxicity were observed. This confirmed that PMO-based AONs are suitable for the modulation of splicing in primary fibroblasts, in agreement with previous reports (32,49,50).

Next, we designed PMO AONs based on the results of the *U7* snRNA screen, and tested these in fibroblasts derived from Pompe patient 1 for promoting *GAA* exon 2 inclusion (Figure 2a and Suppl. Figure S2a). The putative splicing silencer sequences at c.-32-219 and c.-32-179 were targeted using PMO-based AONs (Figure 2a). Blocking of c.-32-179 using AONs 3 or 4 resulted in promotion of exon 2 inclusion and enhancement of GAA enzymatic activity, while AONs 1 and 2 that targeted c.-32-219 were inactive (Figure 2b-d). It is likely that blocking of c.-32-219 may require further optimization of PMO-AON sequences. This was confirmed by flanking exon RT-PCR of exon 2 and in fibroblasts from a second patient (genotype IVS1, c.923A>C; the second allele is expressed) (Figure 2e and Suppl. Figure S2e,f). The maximal possible enhancement of GAA enzyme activity using this approach is ~3.5-5 fold: patients with the IVS1 allele have on average 10-15% leaky wild type splicing, and full restoration will amount to a maximum of 50% (derived from one allele) of the average value of healthy controls. AONs 3 and 4 promoted *GAA* exon 2 inclusion and GAA activity in fibroblasts with ~2.5 fold, indicating that these corrected 50-70% of exon 2 splicing.

To confirm that AONs acted by modulating splicing rather than total *GAA* mRNA expression, splicing product-specific RT-qPCR analysis was performed. This showed that AON 4 enhanced expression of wild type *GAA* mRNA while it repressed expression of aberrant splicing products SV2 and SV3 (Figure 2f). In addition, AON 4 was ineffective in fibroblasts from a healthy control (Figure 2f). Taken together, PMO AONs 3 and 4 were identified to promote exon 2 inclusion with 50-70% efficiency in fibroblasts from patients with the IVS1 *GAA* variant.

Splicing can occur in a tissue-specific manner, and it was unknown how the IVS1 variant and the putative splicing silencer would operate in differentiated skeletal muscle cells, which are affected in Pompe disease. To test this, we first used primary myoblasts derived from healthy controls and Pompe patients. However, these showed limited and heterogeneous capacity to proliferate and differentiate into multinucleated myotubes, which hindered the use of myoblasts for quantitative analysis of AONs (data not shown). A similar reduction of proliferation and differentiation capacity upon passaging of primary myoblasts has been reported previously (41).

We therefore developed an *in vitro* model for childhood/adult Pompe disease using iPS cells. Reprogramming of fibroblasts and characterization of iPS cells are described in Suppl.

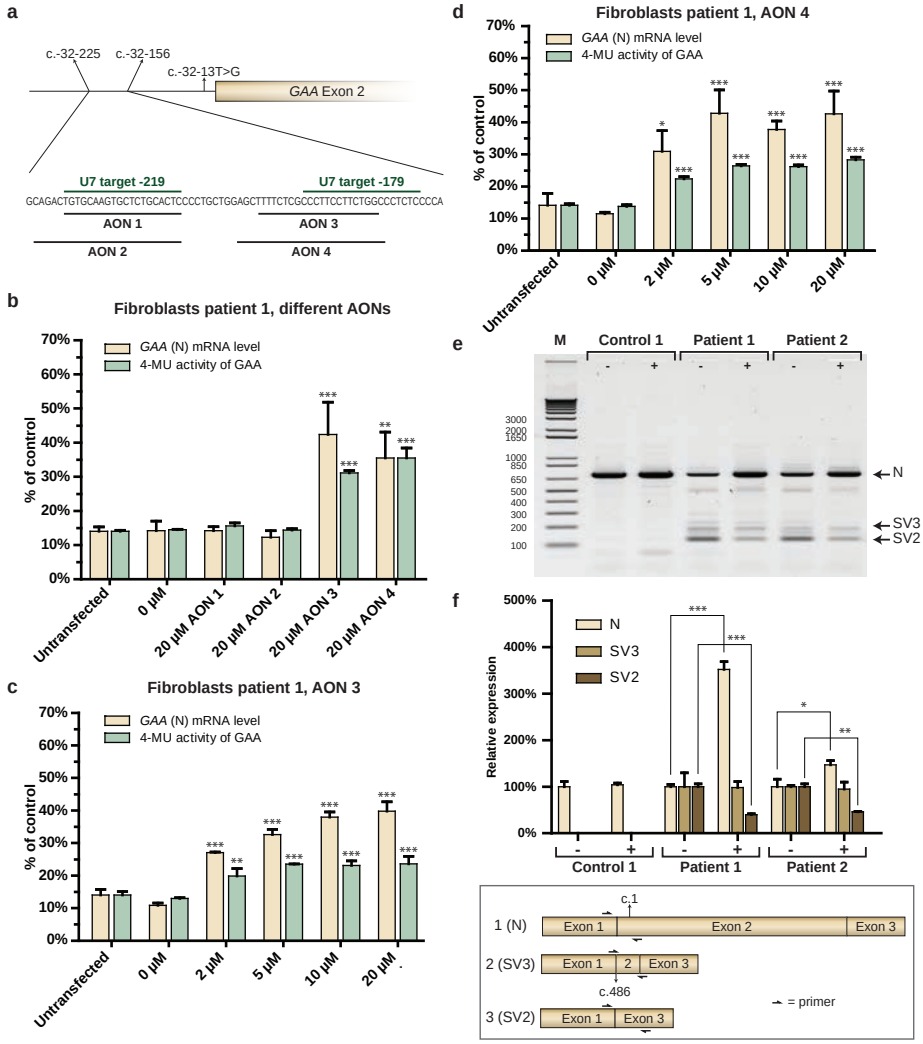


Figure 2. Splicing correction of GAA exon 2 in fibroblasts using PMO-based AONs. *(a) Positions in the GAA pre-mRNA to which PMO-based AONs1-4 anneal. (b) Effect of AONs1-4 in fibroblasts from patient 1. GAA exon 2 inclusion in the mRNA was measured using RT-qPCR analysis (GAA (N) mRNA level), and GAA enzymatic activity using 4-MU as substrate. Data are expressed relative to the average levels in healthy control fibroblasts and were corrected for β-Actin expression. (c) As in Fig. 3b, but now using a concentration range of AON 3. (d) As in Fig. 3b, but now using a concentration range of AON 4. (e) Flanking exon RT-PCR analysis (as in Fig. 2a) of the effect of AON 4 on GAA exon 2 inclusion in fibroblasts from patient 1 and 2. -: 0 µM AON, +: 20 µM AON. (f) RT-qPCR analysis of individual splicing products of GAA exon 2 splicing. The N, SV2, and SV3 products were quantified using primers as outlined in the cartoon, and the effect of AON 4 on GAA exon 2 splicing was determined in fibroblasts from patients 1 and 2 and control 1. Data are corrected for β-Actin expression and normalized per splicing variant for expression in untreated cells to visualize the effect per variant. Note that patient 2 carried a missense GAA variant on the second allele which shows mRNA expression (partially masking effects on the IVS1 allele), whereas patient 1 has no GAA mRNA expression from the second allele due to a frameshift and subsequent mRNA degradation. Data are means +/- SDs of three biological replicates. *p < 0.05, **p < 0.01, ***p < 0.001.*

Figure S3a-d. iPS cells from two patients and two healthy controls were differentiated into myogenic progenitors using a transgene-free protocol modified from Borchin et al. (42). While this method yielded purified Pax7+ myogenic progenitors after a 35-day protocol (Suppl. Figure S3e), the recovery after FACS sorting was low. Between 50,000 and 500,000 cells could be purified starting from a full 10 cm dish of iPS cells, yielding only a few wells in a tissue culture dish that could be used for testing AONs. In addition, the capacity to differentiate into multinucleated myotubes varied largely between individual purifications (Suppl. Figure S3f). It was therefore not possible to reproducibly test the effect of AONs on splicing in freshly isolated iPS-derived myogenic progenitors.

To address this, we tested cell culture conditions aiming to expand purified Pax7+ cells while maintaining proliferative and differentiation capacity. Out of 5 media tested (Suppl. Table S1), medium 5 supported prolonged proliferation of myogenic cells (Figure 3a). Critical components included DMEM as basal medium and FGF2, which supports proliferation. All 4 lines (from 2 Pompe patients and 2 healthy controls) could be expanded with nearly identical proliferation rates at an average of 29.4 ± 1.3 hrs/cell cycle, with at least 5 x 10^7 fold expansion starting with 40,000 cells to yield at least 1 x 10^{12} cells (Figure 3b). At several time points during the expansion phase, cells could be frozen in viable state and used for subsequent expansion. Proliferating myogenic progenitors were characterized by high expression of the myogenic markers *MyoD*, *Myogenin*, *Six1*, and *Six4*, moderately high expression of the myogenic differentiation marker α-*actinin* and of *FGF2*, while the neural crest marker *Sox1* was not expressed (Figure 3c and Suppl. Figure S3g,h). Upon expansion, the karyotype remained normal (Figure 3d). In addition, at any stage of expansion, cells could be differentiated into multinucleated myotubes with a high fusion index (tested in > 500 differentiations performed to date) (Figure 3e and Suppl. Figure S3i,j). Multinucleated myotubes showed high expression of the myogenic differentiation markers Myosin Heavy Chain (MHC) (Figure 3e and Suppl. Figure S3i,j) and α-*actinin* (Figure 3c). The lysosomal markers *LAMP1* and *LAMP2* were expressed at similar levels in myotubes from healthy controls and patients (Suppl. Figure S3g). This suggests that Pompe disease pathology, which includes enlarged lysosomes and elevated expression of LAMP1/2 in a subset of skeletal muscle fibers in patients (22,51,52), has not advanced to critical levels that affect lysosomal size and numbers *in vitro*, which is consistent with the late-onset phenotype of childhood/adult onset Pompe disease. We conclude that the expansion protocol reproducibly provided the amounts of multinucleated myotubes that were required for the quantitative analysis of AONs on splicing in differentiated skeletal muscle cells.

We expanded myogenic progenitors, differentiated these in a four-day protocol into multinucleated myotubes, and analyzed *GAA* splicing by flanking exon RT-PCR and quantitative RT-qPCR of splicing products. This showed leaky wild type splicing, and partial and complete skipping of exon 2 in patient-derived myotubes, but not in myotubes from

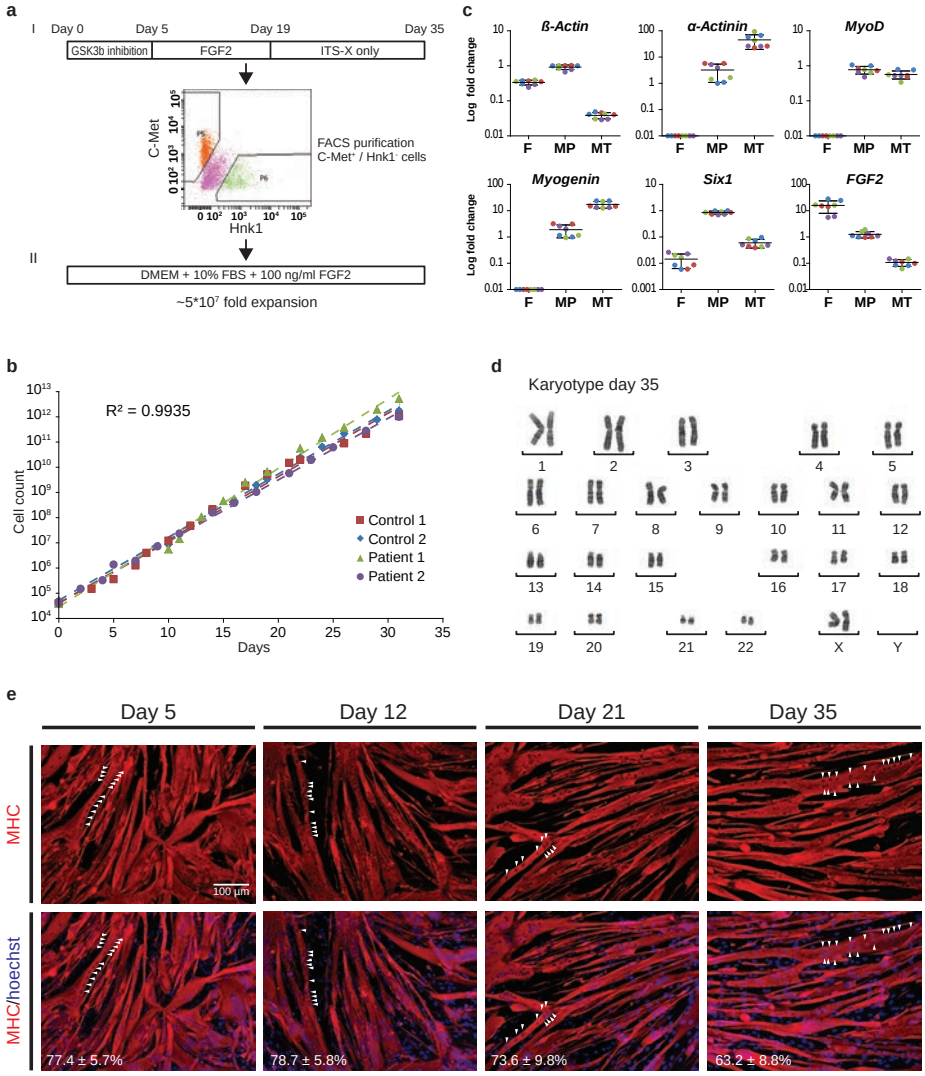


Figure 3. Expansion of purified iPS-derived myogenic progenitors and differentiation into multinucleated myotubes. *(a) I, Scheme for differentiation of iPS cells into myogenic progenitors and FACS purification; II, Expansion of purified myogenic progenitors. The expansion medium is indicated. The fold expansion is indicated. (b) Linear proliferation curves for all four iPS-derived myogenic progenitor lines during expansion. The average R2 shown was calculated for all datapoints of the 4 lines, and indicates high concordance between the four lines. (c) mRNA expression of iPS-derived myogenic progenitors (purified and expanded) and myotubes thereof. Equal amounts of total RNA were isolated from fibroblasts (F), myogenic progenitors (MP), and myotubes (MT), generated after 4 days of differentiation, and mRNA expression of the indicated genes was determined by RT-qPCR analysis. Log fold change was calculated relative to Control 1 sample 1. Lines represent means. Symbols are as in (b). Biological duplicates are shown. (d) Karyotype analysis after expansion of purified myogenic progenitors at day 35 (a representative example of 15 nuclei). (e) Myogenic progenitors retain their capacity to differentiate into multinucleated myotubes during expansion. Myogenic progenitors were expanded, and at several time points during expansion a subculture was differentiated for 4 days and stained for expression of the myogenic differentiation marker MHC (MF-20 antibody; red). Nuclei were stained with Hoechst (blue). The white arrowheads point to examples of aligned nuclei present in a single myotube. The fusion index of each differentiation is shown at the bottom. Data are means +/- SDs of five technical replicates.*

healthy controls, similar to primary fibroblasts (Figure 4a,b). Myotubes from both Pompe patients showed partial residual GAA enzyme activity consistent with residual leaky wild type splicing (Suppl. Figure S4a). This confirmed that the IVS1 variant caused aberrant splicing of exon 2 and GAA enzyme deficiency in skeletal muscle cells.

Next, we tested the effect of AONs 3 and 4 on exon 2 inclusion in myotubes. Treatment of patient-derived myotubes resulted in a concentration-dependent increase in wild type *GAA* splicing and a concomitant decrease in expression of aberrant splicing products SV2 and SV3, as shown by quantitative analysis of individual splicing products using RT-qPCR (Figure 4c,d and Suppl. Figure S4d,e,i). In myotubes from healthy controls, AONs 3 and 4 did not affect *GAA* exon 2 splicing (Figure 4e and Suppl. Figure S4f,i), indicating that these only restored normal splicing in the context of the IVS1 variant without promoting additional effects on *GAA* mRNA expression. This was confirmed by flanking exon RT-PCR analysis of exon 2 (Figure 4f). Importantly, treatment of patient-derived myotubes with AONs 3 or 4 resulted in elevation of GAA enzyme activity above the disease threshold of 20% of average healthy control levels (Figure 4g and Suppl. Figure S4g). Treatment of myotubes from healthy controls did not affect GAA enzyme activity, in agreement with the lack of effects on *GAA* mRNA expression in these cells (Figure 4h and Suppl. Figure S4h). AON treatment did not affect myogenic differentiation to myotubes (Figure 4i and Suppl. Figure S4b,c). We conclude that the splicing silencer sequence at c.-32-179 operates in skeletal muscle cells and that its inhibition by AONs can restore splicing in cells from Pompe patients carrying the IVS1 variant.

As it was unclear how AONs 3 and 4 restored exon 2 inclusion, we were interested to investigate their mechanism of action. We noted that the target sequence of these AONs showed similarity to a pY tract, which is usually present between 5-40 nucleotides upstream of a splice acceptor. We then performed *in silico* analysis of splice sites using Alamut®, and this predicted a strong natural cryptic splice acceptor site 12-13 nt downstream of the binding site for AONs 3 and 4 (Figure 5a). One hundred and two nt further downstream, a strong natural cryptic splice donor was predicted. These predictions were found for the wild type *GAA* gene and were unchanged by the IVS1 variant, suggesting that these cryptic splice sites defined a hypothetical natural pseudo exon. Usage of the natural pseudo exon was observed in a minigene construct harboring the IVS1 variant. Mutation of the natural cryptic splice sites of the natural pseudo exon abolished its inclusion in the context of the minigene harboring the IVS1 variant (Suppl. Figure S5a-c). This suggested the possibility that AONs 3 and 4 may act by inhibiting usage of a natural pseudo exon rather than by repressing a putative ISS.

To test this, we first analyzed whether splice products comprising the putative natural pseudo exon exist in cells from Pompe patients. To this end, mRNA isolated from patient-derived myotubes was analyzed by flanking exon RT-PCR of exon 2, and PCR products were cloned in a TOPO vector. Ninety six clones were analyzed by Sanger sequencing, and

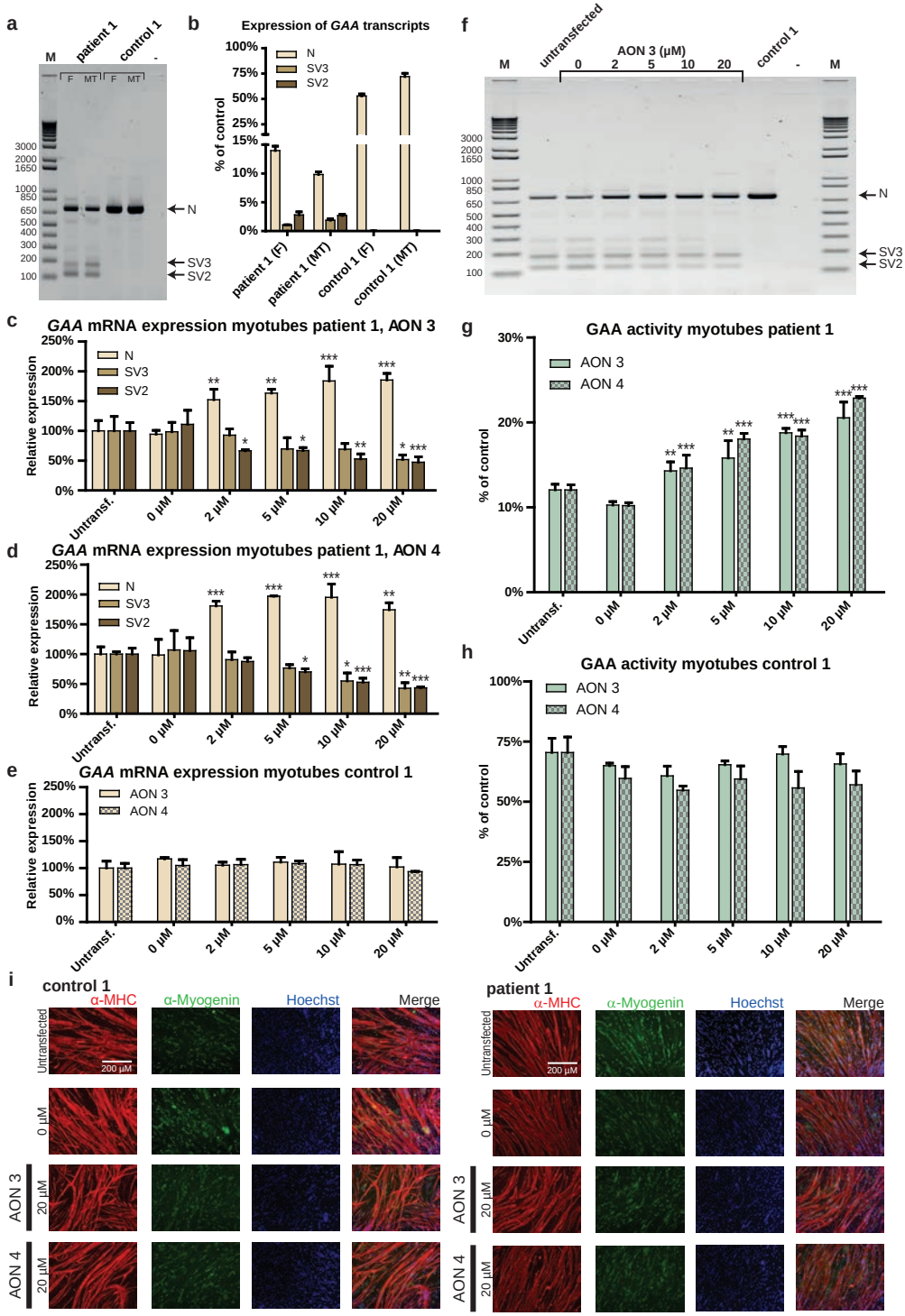

a
M
patient 1
control 1
F
MT
F
MT
-
3000
2000
1650
1000
850
650
500
400
300
200
100
N
SV3
SV2
b
Expression of GAA transcripts
100%
75%
50%
25%
15%
10%
5%
0%
% of control
N
SV3
SV2
patient 1 (F)
patient 1 (MT)
control 1 (F)
control 1 (MT)
f
M
untransfected
AON 3 (μM)
0
2
5
10
20
control 1
-
M
3000
2000
1650
1000
850
650
500
400
300
200
100
N
SV3
SV2
c
GAA mRNA expression myotubes patient 1, AON 3
250%
200%
150%
100%
50%
0%
Relative expression
N
SV3
SV2
Untransf.
0 μM
2 μM
5 μM
10 μM
20 μM
d
GAA mRNA expression myotubes patient 1, AON 4
250%
200%
150%
100%
50%
0%
Relative expression
N
SV3
SV2
Untransf.
0 μM
2 μM
5 μM
10 μM
20 μM
e
GAA mRNA expression myotubes control 1
250%
200%
150%
100%
50%
0%
Relative expression
AON 3
AON 4
Untransf.
0 μM
2 μM
5 μM
10 μM
20 μM
g
GAA activity myotubes patient 1
30%
20%
10%
0%
% of control
AON 3
AON 4
Untransf.
0 μM
2 μM
5 μM
10 μM
20 μM
h
GAA activity myotubes control 1
100%
75%
50%
25%
0%
% of control
AON 3
AON 4
Untransf.
0 μM
2 μM
5 μM
10 μM
20 μM
i
control 1
α-MHC
α-Myogenin
Hoechst
Merge
Untransfected
0 μM
AON 3
20 μM
AON 4
20 μM
200 μM
patient 1
α-MHC
α-Myogenin
Hoechst
Merge
Untransfected
0 μM
AON 3
20 μM
AON 4
20 μM
200 μM

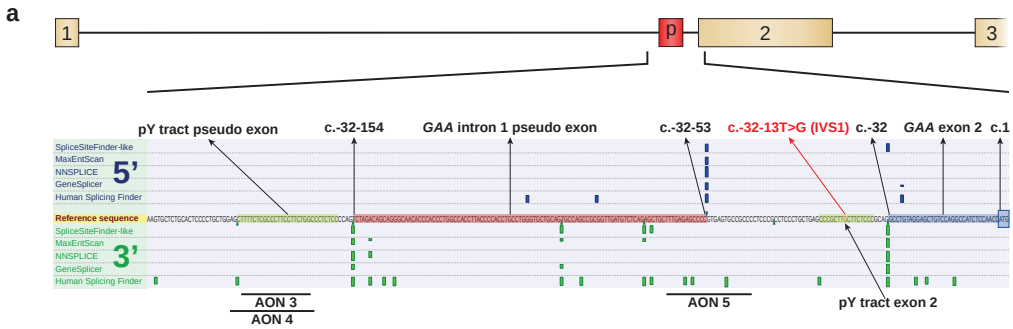
Figure 4. Quantitative analysis of GAA exon 2 splicing in expanded iPS-derived myotubes. *(a) Comparison of aberrant GAA splicing in fibroblasts and myotubes. Equal amounts of total RNA from primary fibroblasts (F) and their corresponding iPS-derived myotubes (MT), derived from patient 1 or a healthy control, were analyzed by flanking exon RT-PCR of exon 2 as described in Fig. 1a. (b), as (a) but now as analyzed by RT-qPCR of individual splicing products. To facilitate comparison between different cell types, no normalization was used, and all products were compared to the value of average control fibroblast product N levels using the delta-Ct method. (c-i) Quantitative analysis of splicing correction in iPS-derived myotubes. (c) Effect of AON 3 on GAA exon 2 splicing in myotubes from patient 1 as analyzed with RT-qPCR analysis of individual splicing products. Data were normalized against expression of four genes that showed no consistent changes in expression: MyoD, Myogenin, LAMP1, and LAMP2 (see Suppl. Figure S4i). (d) As (c), but now for AON 4. (e) Effect of AONs 3 and 4 on GAA exon 2 splicing in myotubes from control 1 as analyzed with RT-qPCR analysis of splice product N. Control cells have undetectable levels of aberrant splice products SV2 and SV3. Normalization was as in (c). (f) Flanking exon RT-PCR analysis of the effect of AON 3 on GAA exon 2 splicing in myotubes from patient 1 and control 1. (g) Effects of AON 3 and 4 on GAA enzymatic activity in myotubes from patient 1. (h) As (g), but now in myotubes from control 1. (i) AON treatment does not affect myogenic differentiation. Immunofluorescent stainings of myotubes after treatment with AONs 3 and 4. Red: MHC (anti-MF-20); green: Myogenin; blue: nuclei (Hoechst). 0 μM: mock transfection. Representative pictures are shown. Quantitative data are means +/- SDs of three biological replicates. *p < 0.05, **p < 0.01, ***p < 0.001.*

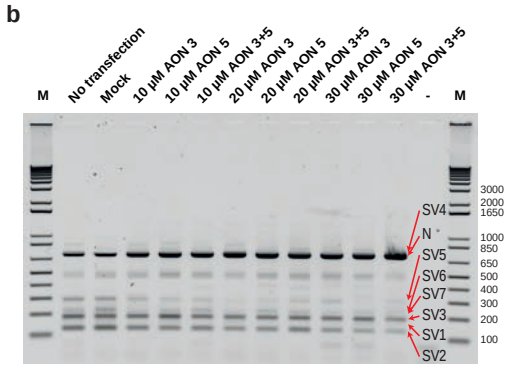
this resulted in the identification of 8 splice variants (Figure 5b,c, Table 1 (column: colony count mock transfection) and Suppl. Figure S5d). The predicted natural pseudo exon was indeed detected in two splice products, in which exon 2 was fully (SV6) or partially (SV5) skipped. Both products were likely subject to mRNA degradation due to the lack of the translation start codon, explaining their low abundance. Nevertheless, these could also be identified on agarose gels following flanking exon PCR of exon 2 (Figure 5b). Other low abundant splice products (SV1, SV4, and SV7) utilized a previously described cryptic splice donor nearby exon 1 (25,26). However, these never contained the natural pseudo exon. We conclude that the predicted natural pseudo exon indeed exists *in vivo* and that it is preferentially included in splice products in which exon 2 is partially or fully skipped due to the IVS1 variant.

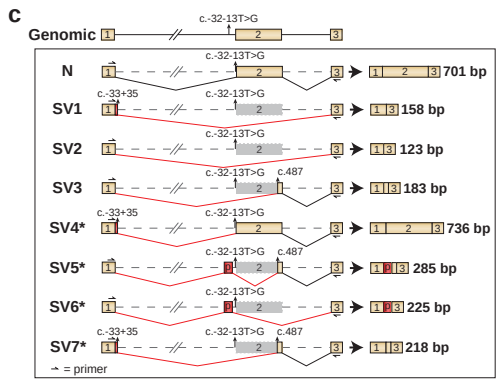
Table 1 *Variants detected before and after treatment with AONs 3 and 5.*

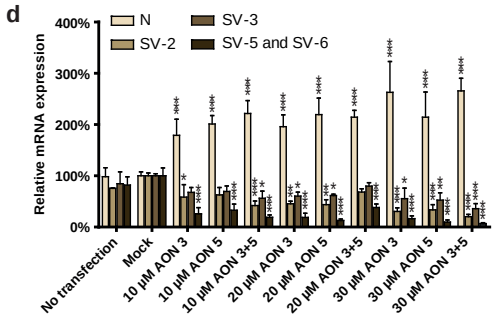
Splice variant	colony count mock transfection	colony count transfection of 15 μM AON 3 and 5
total	96	96
Readable	93	90
N	14	45
SV1	3	3
SV2	44	16
SV3	24	23
SV4	0	2
SV5	4	1
SV6	2	0
SV7	2	0

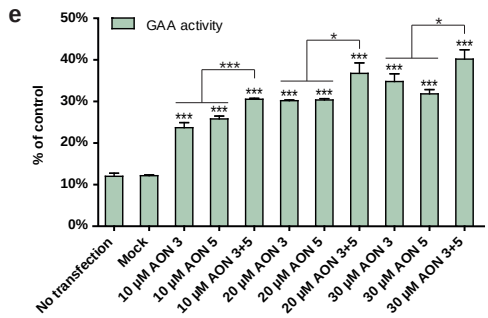
Short introns are unfavorable for successful splicing and have a typical minimum length of 70-80 nt (53). The length of the intron between the natural pseudo exon and exon 2 is 52 nt, which violates this rule. This suggested the possibility that inclusion of the natural pseudo exon competes with exon 2 inclusion, which is in agreement with the mutually


a
1
p
2
3
pY tract pseudo exon
c.-32-154
GAA intron 1 pseudo exon
c.-32-53
c.-32-13T>G (IVS1)
c.-32
GAA exon 2
c.1
SpliceSiteFinder-like
MaxEntScan
NNSPLICE
GeneSplicer
Human Splicing Finder
5'
Reference sequence
3'
AON 3
AON 4
AON 5
pY tract exon 2


b
M
No transfection
Mock
10 µM AON 3
10 µM AON 5
10 µM AON 3+5
20 µM AON 3
20 µM AON 5
20 µM AON 3+5
30 µM AON 3
30 µM AON 5
30 µM AON 3+5
3000
2000
1650
1000
850
650
500
400
300
200
100
SV4
N
SV5
SV6
SV7
SV3
SV1
SV2


c
Genomic
c.-32-13T>G
N
SV1
SV2
SV3
SV4*
SV5*
SV6*
SV7*
c.-33+35
c.487
701 bp
158 bp
123 bp
183 bp
736 bp
285 bp
225 bp
218 bp
= primer


d
N
SV-3
SV-2
SV-5 and SV-6
Relative mRNA expression
400%
300%
200%
100%
0%
No transfection
Mock
10 µM AON 3
10 µM AON 5
10 µM AON 3+5
20 µM AON 3
20 µM AON 5
20 µM AON 3+5
30 µM AON 3
30 µM AON 5
30 µM AON 3+5


e
GAA activity
% of control
50%
40%
30%
20%
10%
0%
No transfection
Mock
10 µM AON 3
10 µM AON 5
10 µM AON 3+5
20 µM AON 3
20 µM AON 5
20 µM AON 3+5
30 µM AON 3
30 µM AON 5
30 µM AON 3+5

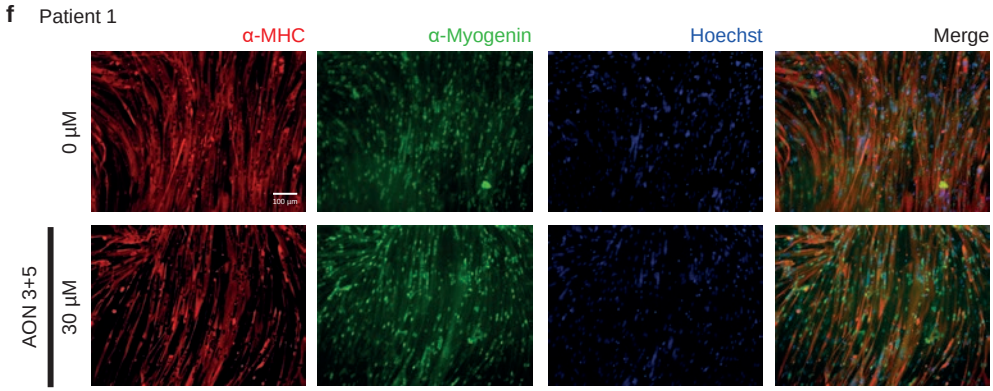

f
Patient 1
α-MHC
α-Myogenin
Hoechst
Merge
0 µM
AON 3+5
30 µM
100 µm

Figure 5. Blocking of a natural pseudo exon restores GAA exon 2 splicing. *(a) The splicing silencer in intron 1 is predicted to be the pY tract of a natural pseudo exon. Alamut® was used to predict splice sites around the splicing silencer identified in Fig. 1. Note that predictions were independent of the IVS1 variant. A strong 3' splice site was predicted at c.-32-154, and a strong 5' splice site at c.-32-53, which suggested the presence of a natural pseudo exon, indicated by 'p' in the cartoon. The canonical 3' splice site of exon 2 at c.-32 showed strong prediction and is also indicated. (b) Blocking of natural pseudo exon splicing restores GAA exon 2 splicing. AON 5 was designed to block the predicted 5' splice site, and AONs 3 and 5 were tested alone or in combination in myotubes from patient 1. Flanking RT-PCR analysis of GAA exon 2 was performed. Splicing products were identified by TOPO cloning (see Table 1) and are indicated in the gel and in the cartoons in (c). (d). Analysis of the experiment in (b) by RT-qPCR of individual splicing products. Splicing to the natural pseudo exon is represented by SV5 and SV6 and these products were quantified using unique PCR primers. Data were normalized against expression of four genes that showed no consistent changes in expression: MyoD, Myogenin, LAMP1, and LAMP2 (see Suppl. Figure S5e). (e) Analysis of the experiment in (c) on GAA enzyme activity. (f) Combined treatment with AONs 3 and 5 does not interfere with myogenic differentiation to myotubes. Immunofluorescent staining results are shown for treatment of iPS-derived myotubes obtained from patient 1. Red: MHC (anti-MF-20); green: Myogenin; blue: nuclei (Hoechst). 0 μM: mock transfection. Representative pictures are shown. Quantitative data are means +/- SDs of three biological replicates. *p < 0.05, **p < 0.01, ***p < 0.001.*

exclusive inclusion of the natural pseudo exon or exon 2 in splice products. Such scenario explains why AONs 3 and 4 promote exon 2 inclusion, namely by repression of inclusion of the natural pseudo exon via interfering with the pY tract of the natural cryptic splice acceptor site. We hypothesized that repression of the natural cryptic splice donor would likewise promote exon 2 inclusion. To test this, AON 5 was designed to target the natural cryptic splice donor site of the natural pseudo exon (Figure 5a and Suppl. Figure S2a). In patient-derived myotubes, AON 5 promoted exon 2 inclusion (product N) and repressed inclusion of the natural pseudo exon (products SV5 and SV6), as shown by flanking exon RT-PCR and splicing product-specific RT-qPCR (Figure 5b,d, and Suppl. Figure S5e). AON 5 was equally effective in splicing correction compared to AON 3, in agreement with the idea that both AONs prevent utilization of the natural pseudo exon. GAA enzyme activity was enhanced by AON 5 to similar levels compared to AON 3 (Figure 5e) and myotube differentiation was not altered by the AON treatment (Figure 5f). These results suggest that the natural pseudo exon competes with exon 2 splicing and that natural pseudo exon skipping by AONs promotes exon 2 inclusion.

The identification of the natural pseudo exon offered an additional option for splicing correction, namely by the simultaneous targeting of the cryptic splice acceptor and donor sites. To test this, a combination of AON 3 plus AON 5 was tested in patient-derived myotubes. At the same total AON concentrations, combinations of AON 3 plus AON 5 were very efficient in promoting exon 2 inclusion and repressing aberrant exon 2 splicing (Figure 5b,d). We used TOPO cloning as above to analyze all products that arise from treatment with AON 3 plus AON 5 (Table 1). No additional products besides the 8 known splicing products were identified. Compared to mock treated cells, cells treated with AON 3 plus 5 showed an increase in the number of clones with a wild type exon 2 insert from 14 to 45 (3.2 fold), while the number of clones that contained the natural pseudo exon was reduced 6 fold from 6 to 1 (Table 1). When tested on GAA enzyme activity, combinations of AONs 3 plus 5 were more efficient compared to single AONs, at an enhancement of

up to 3.3 fold (Figure 5e). Following the calculation outlined above, this amounts to a highly efficient splicing correction of the IVS1 allele of 66-99%, calculated based on the average GAA enzyme levels in healthy controls (see Suppl. Figure S4a). We conclude that the simultaneous inhibition of the cryptic splice donor and acceptor sites of the natural pseudo exon is an efficient way to promote exon 2 inclusion and is able to restore the majority of GAA enzymatic activity in patient-derived skeletal muscle cells towards healthy control levels.

DISCUSSION

Here, we have developed methods for promotion of exon inclusion, and for drug testing and screening in purified patient-derived multinucleated myotubes, and used these to develop AONs that restore aberrant splicing of the common IVS1 *GAA* variant in Pompe disease. The *U7* snRNA-based approach using lentiviral expression via a modified construct provides a platform for screening of AON sequences in a fast and cheap manner. Purified iPS-derived myogenic progenitors divided fast at 29.4 ± 1.3 hrs per cell cycle in optimized cell culture medium. These could be expanded at least 5 x 10^7 fold without progressive cell cycle delay, and showed highly reproducible *in vitro* differentiation indicated by morphology, marker gene expression, and a high fusion index. The *U7* snRNA-based screen was used to identify a silencer sequence of *GAA* exon 2 splicing, which was used to design PMO-based AONs that restored exon inclusion in cells carrying the IVS1 allele. This led to the identification of a natural pseudo exon that is utilized in IVS1 cells to repress exon 2 splicing, providing a mechanistic basis for the promotion of exon inclusion, and options for improved targeting via the simultaneous blocking of 3' and 5' cryptic splice sites of the natural pseudo exon.

Many disorders exist that are caused by exon skipping and these would profit from enhancement of exon inclusion (54-56). Mechanisms that regulate exon inclusion in human disease are however still poorly understood. Current insight focusses on the identification of ISS and ESS sequences, which can be cumbersome. Our results suggest that it may be possible to use natural cryptic splice site usage as criterion to design AONs for exon inclusion for other human splicing disorders. Many drugs are developed in heterologous cells using artificial reporters, resulting in a low percentage of drugs that successfully proceed into the clinic. Human patient-derived cells are preferred and iPS technology has great promise for this (57-59). Transgene-free differentiation of human iPS cells into myogenic cells has been achieved by a number of laboratories (42-46), but expansion of (purified) cells and subsequent differentiation into multinucleated myotubes has been limited. Shelton et al reported expansion of unpurified cells of 3-fold during a two-week culturing period (45) using a pre-plating protocol. Choi and coworkers reported expansion of purified iPS-derived myogenic progenitors of at least 1 x 10^4 fold. These cells divided every 47.9 ± 17.0 hrs, but showed limited capacity to differentiate into multinucleated

myotubes, with a fusion index of maximal 10% (46). Our protocol yielded at least 1 x 10^{12} myogenic cells with full potential to differentiate into multinucleated myotubes with fusion indexes of up to 80%, which should allow drug screens in skeletal muscle cells at a more advanced differentiation state. The results of the current study identify AONs that promote endogenous production of wild type GAA protein in patient cells. This reflects a fundamental different mechanism compared to ERT, which relies on utilization of an inefficient route for cellular uptake via M6PRs expressed at the plasma membrane (21). Future studies are required to develop AONs into an alternative treatment option for Pompe disease, which would help patients that respond poorly to ERT.

MATERIALS & METHODS

Generation of induced pluripotent stem cells

Dermal fibroblasts from control 1 and two patients (1 and 2) with Pompe disease were obtained via skin biopsy with informed consent. The Institutional Review Board approved the study protocol. All patient and control primary cell lines were negative for HIV, hepatitis B, hepatitis C as tested by quantitative PCR analysis at the diagnostic department of Virology of the Erasmus MC Rotterdam, The Netherlands. Both patient cell lines contain the IVS1 mutation on one allele. The second allele was c.525delT for patient 1, and c.923A>C (his>pro) for patient 2, which both are established pathogenic *GAA* variants (http://www.pompecenter.nl). Primary fibroblasts were reprogrammed into iPS cells using a polycistronic lentiviral vector of *Oct4*, *Sox2*, *Klf4*, and *c-Myc* as described (60). iPS control 2 cell line was a gift from Christian Freund and Christine Mummery and has been characterized previously (61). iPS cells were cultured on γ-irradiated mouse embryonic feeder (MEF) cells. The iPS culture medium consisted of DMEM/F12 medium (Invitrogen), 20% knock-out serum replacement (Invitrogen), 1% non-essential amino acids (Gibco), 1% penicillin/streptomycin/L-glutamine (100x, Gibco), 2 mM β-mercaptoethanol (Invitrogen) and 20 ng/ml basic fibroblast growth factor (Peprotech).

Immunofluorescence

Cells were cultured on normal tissue culture dishes and fixed with 4% paraformaldehyde (Merck) in PBS for 10 minutes at room temperature, washed with PBS and permeabilized for 5 minutes with 0.1% Triton X-100 (AppliChem) in PBS. Blocking was performed for 30 minutes at room temperature with blocking solution containing PBS-T (0.1% Tween, Sigma) with 3% BSA (Sigma). Primary antibodies (Suppl. Table S2) were diluted into 0.1% BSA in PBS-T and incubated either 1 hour at room temperature. After incubation cells were washed three times for 5 minutes with PBS-T and incubated with the secondary antibodies (1:500, Alexa-Fluor-594-α-goat, Alexa-Fluor-488-α-mouse, Alexa-Fluor-594-α-rabbit, Alexa-Fluor-488-α-rabbit, Invitrogen or horse anti-mouse biotin (Vector Laboratories)) in

PBS-T for 30 minutes at room temperature. For incubations with secondary biotinylated antibodies cells were washed three times for 5 minutes with PBS-T and incubated with Streptavidine 594 (Invitrogen, 1:500). The cells were subsequently washed two times for 5 minutes with PBS and incubated for 15 minutes with Hoechst (1:15000, Thermo Scientific). Cells were imaged in PBS.

Microarray analysis

RNA samples to be analyzed by microarrays were prepared using RNeasy columns with on-column DNA digestion (Qiagen). 300 ng of total RNA per sample was used as input into a linear amplification protocol (Ambion), which involved synthesis of T7-linked double-stranded cDNA and 12 hours of *in vitro* transcription incorporating biotin-labelled nucleotides. Purified and labeled cRNA was then hybridized for 18h onto HumanHT-12 v4 expression BeadChips (Illumina) following the manufacturer's instructions. After recommended washing, chips were stained with streptavidin-Cy3 (GE Healthcare) and scanned using the iScan reader (Illumina) and accompanying software. Samples were exclusively hybridized as biological replicates. The bead intensities were mapped to gene information using BeadStudio 3.2 (Illumina). Background correction was performed using the Affymetrix Robust Multi-array Analysis (RMA) background correction model (62). Variance stabilization was performed using the log2 scaling and gene expression normalization was calculated with the method implemented in the lumi package of R-Bioconductor. Data post-processing and graphics was performed with in-house developed functions in Matlab. Hierarchical clustering of genes and samples was performed with one minus correlation metric and the unweighted average distance (UPGMA) (also known as group average) linkage method. The microarray data have been deposited at Gene Expression Omnibus (GEO) with accession number GSE75713.

In vitro differentiation

iPS colonies were washed once with PBS and treated for 45 minutes with 1 mg/ml collagenases IV (Invitrogen) at 37 °C, scraped and centrifuged for 15 seconds at 800 rpm. The pellet was slowly resuspended in EB medium (iPS culture medium without FGF2) containing 10 μM Y-27632 dihydrochloride (Ascent Scientific), and plated on low binding plates (Cyto one). For endodermal differentiation, 10 μM SB 431542 (Ascent Scientific) was added to the EB medium. Six days later EBs were plated in 12 wells coated with 0.1% gelatin (Sigma) for endodermal and mesodermal differentiation or with matrigel-coated plates for ectodermal differentiation in endoderm/mesoderm/ectoderm medium (Suppl.Table S3). Cells were fixed after 14 days of differentiation with 4% paraformaldehyde (Merck) in PBS for 5 minutes at room temperature and processed for immunofluorescence.

Karyotype analysis

iPS or myogenic progenitors were detached with TrypLe (Gibco) for 5 minutes at 37 °C. The pellet was incubated with 10 µg/ml colcemid (Gibco) for 30 minutes at room temperature. Cells were then centrifuged for 10 minutes at 1100 rpm and resuspended into prewarmed 0.075 M KCL and incubated for 10 minutes at 37 °C. After incubation cells were five times washed with fixation solution (3:1 methanol:acetic acid) and spread onto glass slides. Hoechst staining was performed as described above. Fifteen slides were analyzed per cell line.

Differentiation of iPS cells to myogenic progenitor cells

Differentiation of iPS cells to myogenic progenitors cells was modified from Borchin et al. (42). Briefly, 0.6 mm large iPS colonies cultured in 10 cm dishes on MEF feeders were treated for 5 days with 3.5 µM CHIR99021 (Axon Medchem) in myogenic differentiation medium (DMEM/F12, 1x ITS-X and 1x Penicillin/Streptomycin/Glutamine, all Gibco). CHIR99021 was removed and cells were cultured in myogenic differentiation medium containing 20 ng/ml FGF2 (Prepotech) for 14 days, and were then cultured for an additional 16 days in myogenic differentiation medium only. Medium was refreshed daily.

Purification of myogenic progenitors using FACS

Following the 35-day protocol for differentiating iPS cells into a mixture of cells including myogenic progenitors as outlined above, cells were harvested and purified by FACS. To this end, cells were washed once with PBS, incubated for 5 minutes with TrypLe (Gibco) at 37 °C and gently detached with a pipetboy. The cell suspension was filtered through a 0.45 µM FACS strainer (Falcon) to remove cell aggregates. Cells were centrifuged for 4 minutes at 1000 rpm and incubated with anti-HNK-1-FITC (1:100, Aviva Systems Biology) and anti-C-MET-APC (1:50, R&D Systems) antibodies for 30 minutes on ice in myogenic differentiation medium. Cells were washed three times with ice-cold 1% BSA in PBS before FACS sorting. Hoechst (33258, Life Technology) was used as viability marker. Hoechst/C-MET-positive cells were sorted with a 100 µm nozzle and collected in ice cold iPS-myogenic progenitor proliferation medium (iPS-MP-pro medium) containing DMEM high glucose (Gibco) supplemented with 100 U/ml Penicillin/Streptomycin/Glutamine (Life Technology), 10% Fetal bovine serum (Hyclone, Thermo Scientific),100 ng/ml FGF2 (Prepotech). To reduce cell death, medium was supplemented with 1x RevitaCell™ Supplement (Gibco) during collection and the first 24 hrs of cell culture. Sorting time was limited to 20 minutes per well. Plates/wells were coated for 30 minutes at room temperature with ECM (Sigma-Aldrich, E6909-5ml, 1:200 in iPS-MP-pro medium).Sorted cells were plated either at 40,000 cells in one well of a 48 well plate or at 80,000 cells in one well of a 24 well plate, depending on the amount of cells.

Expansion of myogenic progenitor cells

One day after plating FACS sorted myogenic progenitors, the medium was refreshed with iPS-MP-pro medium. When cells reached 90% confluence, cells were passaged using 2x diluted TrypLe in PBS and plated on ECM-coated plastic. Myogenic progenitors could be frozen at any time during expansion without loss of proliferation and differentiation capacity, using iPS-MP-pro medium supplemented with 10% DMSO.

Differentiation of myogenic progenitors into multinucleated myotubes

For differentiation to multinucleated myotubes, myogenic progenitors were grown to 90% confluence and the medium was then replaced with myogenic differentiation medium (DMEM/F12, 1x ITS-X and Penicillin/Streptomycin/Glutamine, all Gibco). After 4 days, myotubes were harvested.

Modification of the U7 snRNA vector for efficient one-step cloning of AON sequences

The *U7* snRNA gene and promoter was amplified by PCR from female mouse genomic DNA using Fw-ms-*U7*snRNA-Pst1 and Rv-ms-*U7*snRNA-SalI primers, which included PstI and SalI overhang restriction sites. The PCR fragment (425 bp) was cloned into a pCRII-TOPO vector according to the manufacturers manual (Invitrogen). SMopt and NsiI sites were generated by site-directed mutagenesis according to an inner and outer primer design with Fw- and Rv-*U7*snRNA-SMopt or Fw- and Rv-U7snRNA-NsiI as inner primers and with Fw-M13 and Rv-M13 as outer primers (Suppl. Table S4), and subcloned using the PstI and SalI sites in front of the polypurine tract fragment of the lentiviral vector used for reprogramming, from which OSKM and the SF promoter were removed.

Cloning of AONs into the U7 snRNA vector

AONs were inserted via PCR amplification using a forward primer that contained the desired antisense sequence and the unique NsiI restriction site and the reverse primer Rv-ms-*U7*snRNA-SalI. The amplified PCR product was purified by agarose gel electrophorese, extracted (gel extraction kit, Qiagen), digested with NsiI and SalI, purified (PCR purification kit, Qiagen), and cloned into the NsiI and SalI sites of the *U7* snRNA vector. Clones were verified by sequencing with the Fw-ms-*U7*snRNA-PstI (Suppl. Table S4) and restriction enzyme digestion.

Cell culture

HEK293T cells or human primary fibroblasts were cultured in Dulbecco's Modified Eagle's Medium (DMEM) high glucose (Gibco) supplemented with 100 U/ml Penicillin/ Streptomycin/Glutamine (Gibco) and 10% Fetal bovine serum (Hyclone, Thermo Scientific). Cells were passaged after reaching 80/90% confluence with TrypLE (Gibco). Human ES lines H1 and H9 were obtained from Wicell Research Institute, Madison, WI,

USA. The identity of cell lines used in this study was confirmed by DNA sequence and microarray analyses. All cell lines were routinely tested for mycoplasma infection using the MycoAlert™ Mycoplasma Detection Kit (Lonza) and were found negative.

Virus production

Lentiviruses were produced by co-transfecting HEK293T cells at 80% confluency in a 10 cm culture dish with the lentivirus transfer vector (3 µg SF-OSKM or SF-*U7*snRNA vectors) and packaging plasmids (2 µg psPAX2 and 1 µg pVSV vectors) using Fugene 6 transfection according to manufacturer's protocol (Promega). Lentiviruses were harvested from the medium after 72 hours of transfection and filtered using a 0.45 µm PDFV filter (Milipore). After filtering lentiviruses were concentrated by high speed centrifugation for 2 hours at 20000 rpm in a Beckman Coulter Ultracentrifuge with SW32 Ti rotor at 4 °C. The supernatant was removed and the pellet was dissolved in 25 µl Dulbecco's Modified Eagle's medium Low Glucose (Invitrogen) per plate and stored in aliquots at -80 °C.

P24 ELISA

Viral titers were determined with the HIV-1 p24 antigen ELISA kit (Retrotek) according to manufacturer's manual. Each virus was diluted 1:40000 and 1:100000 and the OD450 nm was measured with a varioskan (Thermos Scientific) reader.

Transduction of U7 snRNA vectors

One day before infection $6x10^4$ cells per single well of a 12 wells plate of patient 1-derived primary fibroblasts were seeded. One day later the cells were infected with 200 ng virus containing the SF-*U7*snRNA constructs, and after 24 hours cells were washed three times with PBS before adding fresh medium. After 4 days cells were washed with PBS and harvested with RLT buffer of the RNAeasy kit for RNA isolation (Qiagen). For GAA enzyme activity assay cells were harvested after 12 days.

Morpholino transfections

Fibroblasts or myogenic progenitors (day -1 or 0 of differentiation) were transfected with morpholino AONs using Endoporter reagent (Gene-Tools, LLC). Cells were grown to 90% confluency before transfection. Endoporter was used at a concentration of 4.5 µl per ml of medium. Morpholino was dissolved in sterile water to a concentration of 1 mM and the appropriate volume was added to each culture well. Cells were harvested 3-5 days after AON addition.

RNA isolation and cDNA synthesis

RNA was extracted with the RNeasy mini kit with Dnase treatment (Qiagen) and was stored at -80 °C in RNase-free water. cDNA was synthesized from 500 ng RNA using iScript cDNA synthesis kit (Bio-Rad).

qPCR

cDNA was diluted five, ten or twenty times and used with 7.5 µl iTaq Universersal SYBR Green Supermix (Bio-Rad) and 10 pmol/µl forward and reverse primers (Suppl. Table S4) in a CFX96 real-time system (Bio-Rad). Ct values were related to amounts using standard curves of 4-6 dilutions. Quantification of expression was calculated relative to β-*Actin* expression in experiments where primary fibroblasts used, to expression of four markers (*Myog*, *MyoD*, *LAMP1* and *LAMP2*) in experiments where myotubes were used, and to RNA input in experiments were multiple tissues (fibroblasts, myogenic progenitors and myotubes) were compared.

Flanking exon RT-PCR

Ten times diluted cDNA with GC *GAA* Exon1-3 fw and GC *GAA* Exon1-3 rv primers were used for RT-PCR with the Advantage GC 2 PCR kit (Clontech) and a GC-melt concentration of 0.5 M according to manufacturer's protocol. The whole GC-PCR reaction was analyzed on a 1.5% agarose gel containing 0.5 µg/ml ethidium bromide (Sigma).

GAA enzyme activity assay

Cells were harvested with ice cold lysis buffer (50 mM Tris (pH 7.5), 100 mM NaCl, 50 mM NaF, 1% Triton X-100 and one tablet Protease Inhibitor Cocktail (cOmplete, with EDTA, Roche) and incubated for 10 minutes on ice. Samples were centrifuged at 14000 rpm for 10 minutes at 4 °C. GAA enzyme activity was measured using 4-methylumbelliferyl α-D-glucopyranoside (Sigma) as substrate as described (27). Total protein concentration was determined using a BCA protein assay kit (Pierce, Thermo Scientific).

Statistical analysis

All data represent mean +/-SD, and p-values refer to two-sided t-tests. Bonferroni multiple testing correction was applied where necessary. A p-value < 0.05 was considered to be significant. Data showed normal variance. There was no power calculation in any of the experiments. No randomization method was used. No samples were excluded from the analyses. Experiments on expansion of iPS-derived muscle progenitors, differentiation into myotubes, and AON treatment have been performed at least two times. Investigators were not blinded to the identity of the samples.

ACKNOWLEDGEMENTS

We thank Martina Radstaak and Hannie Douben for technical assistance, Dr. Schambach for the OSKM-lentiviral vector, Drs. Christian Freund and Christine

Mummery for providing a control fibroblast and iPS cell line, Michelle Minneboo for brain cDNA, Philip Lijnzaad for advice on statistical analysis, and Dr. Arnold Reuser for discussion. This work was funded by the Sophia Children's Hospital Foundation (SSWO) (grant S-687) and the Prinses Beatrix Spierfonds/Stichting Spieren voor Spieren (grant W.OR13-21).

AUTHOR CONTRIBUTIONS

E.W., A.B., J.P., T.G. and S.G. conducted the experiments; H.Z., A.P., H.R. and W.P. supervised the experiments; E.W., A.B. and W.P. conceived and designed the experiments; M.A. performed bioinformatics analyses; E.W., A.B. and W.P. wrote the paper.

COMPETING FINANCIAL INTERESTS

Parts of this work are filed in a patent application by the Erasmus MC. A.P. has provided consulting services for various industries in the field of Pompe disease under an agreement between these industries and Erasmus MC, Rotterdam, The Netherlands. The other authors declare that they have no conflict of interest.

REFERENCES

1. Kaplan, J.C. and Hamroun, D. (**2014**) The 2015 version of the gene table of monogenic neuromuscular disorders (nuclear genome). *Neuromuscul Disord,* 24, 1123-1153.
2. Thompson, R. and Straub, V. (**2016**) Limb-girdle muscular dystrophies - international collaborations for translational research. *Nat Rev Neurol,* 12, 294-309.
3. van der Ploeg, A.T. and Reuser, A.J. (**2008**) Pompe's disease. Lancet, 372, 1342-1353.
4. Kishnani, P.S. and Beckemeyer, A.A. (2014) New therapeutic approaches for Pompe disease: enzyme replacement therapy and beyond. *Pediatr Endocrinol Rev,* 12 Suppl 1, 114-124.
5. Schoser, B., Stewart, A., Kanters, S., Hamed, A., Jansen, J., Chan, K., Karamouzian, M. and Toscano, A. (**2016**) Survival and long-term outcomes in late-onset Pompe disease following alglucosidase alfa treatment: a systematic review and meta-analysis. *J Neurol.*
6. Van den Hout, H., Reuser, A.J., Vulto, A.G., Loonen, M.C., Cromme-Dijkhuis, A. and Van der Ploeg, A.T. (**2000**) Recombinant human alpha-glucosidase from rabbit milk in Pompe patients. *Lancet,* 356, 397-398.
7. Kishnani, P.S., Corzo, D., Nicolino, M., Byrne, B., Mandel, H., Hwu, W.L., Leslie, N., Levine, J., Spencer, C., McDonald, M. et al. (**2007**) Recombinant human acid [alpha]-glucosidase: major clinical benefits in infantile-onset Pompe disease. *Neurology,* 68, 99-109.
8. Bembi, B., Pisa, F.E., Confalonieri, M., Ciana, G., Fiumara, A., Parini, R., Rigoldi, M., Moglia, A., Costa, A., Carlucci, A. et al. (**2010**) Long-term observational, non-randomized study of enzyme replacement therapy in late-onset glycogenosis type II. *J Inherit Metab Dis,* 33, 727-735.
9. van der Ploeg, A.T., Clemens, P.R., Corzo, D., Escolar, D.M., Florence, J., Groeneveld, G.J., Herson, S., Kishnani, P.S., Laforet, P., Lake, S.L. et al. (**2010**) A randomized study of alglucosidase alfa in late-onset Pompe's disease. *N Engl J Med,* 362, 1396-1406.
10. Strothotte, S., Strigl-Pill, N., Grunert, B., Kornblum, C., Eger, K., Wessig, C., Deschauer, M., Breunig, F., Glocker, F.X., Vielhaber, S. et al. (**2010**) Enzyme replacement therapy with alglucosidase alfa in 44 patients with late-onset glycogen storage disease type 2: 12-month results of an observational clinical trial. *J Neurol,* 257, 91-97.
11. Orlikowski, D., Pellegrini, N., Prigent, H., Laforet, P., Carlier, R., Carlier, P., Eymard, B., Lofaso, F. and 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.
12. Angelini, C., Semplicini, C., Ravaglia, S., Bembi, B., Servidei, S., Pegoraro, E., Moggio, M., Filosto, M., Sette, E., Crescimanno, G. et al. (**2012**) Observational clinical study in juvenile-adult glycogenosis type 2 patients undergoing enzyme replacement therapy for up to 4 years. *J Neurol,* 259, 952-958.

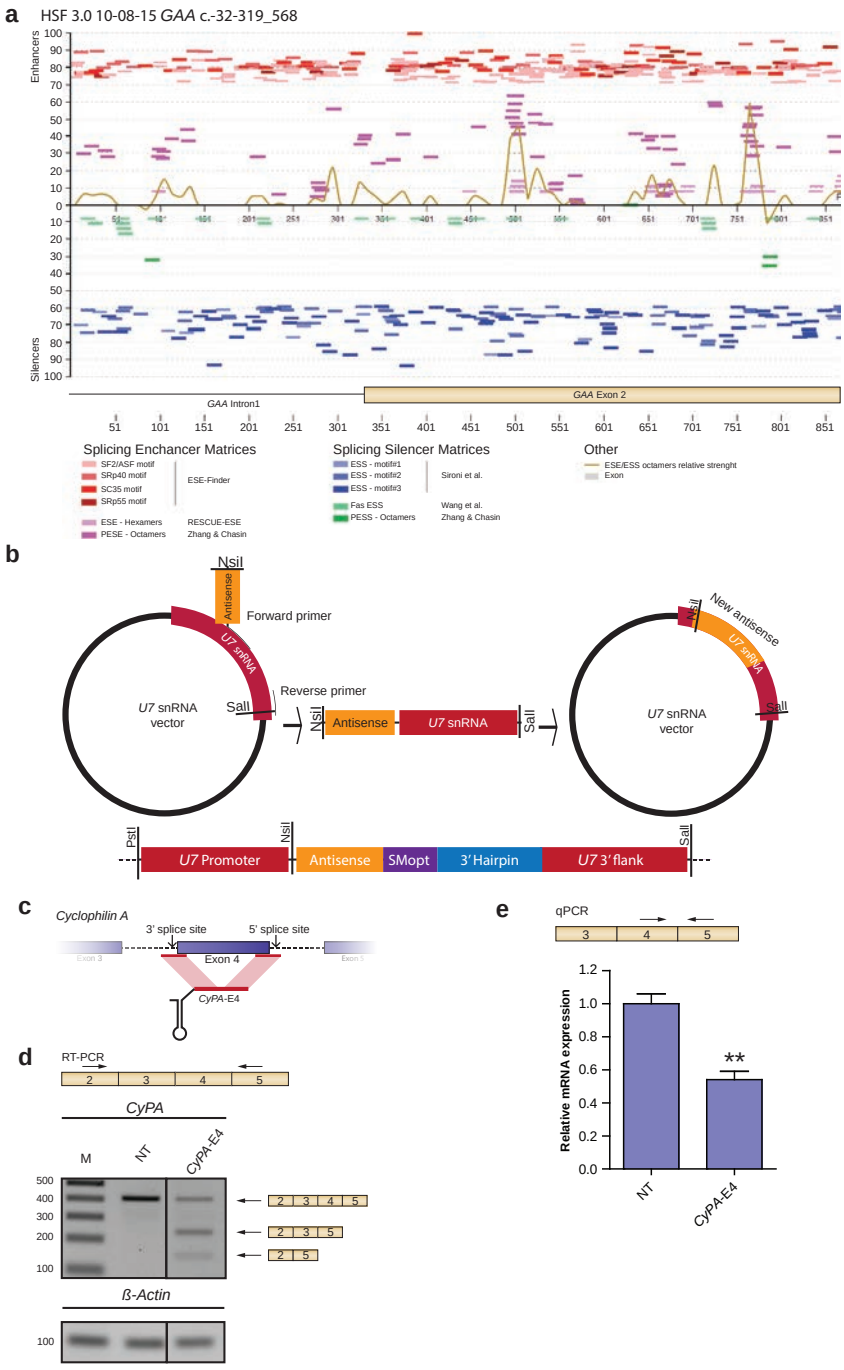
13. van der Ploeg, A.T., Barohn, R., Carlson, L., Charrow, J., Clemens, P.R., Hopkin, R.J., Kishnani, P.S., Laforet, P., Morgan, C., Nations, S. et al. (**2012**) Open-label extension study following the Late-Onset Treatment Study (LOTS) of alglucosidase alfa. *Mol Genet Metab,* 107, 456-461.
14. Gungor, D., Kruijshaar, M.E., Plug, I., D'Agostino, R.B., Hagemans, M.L., van Doorn, P.A., Reuser, A.J. and van der Ploeg, A.T. (**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.
15. Regnery, C., Kornblum, C., Hanisch, F., Vielhaber, S., Strigl-Pill, N., Grunert, B., Muller-Felber, W., Glocker, F.X., Spranger, M., Deschauer, M. et al. (**2012**) 36 months observational clinical study of 38 adult Pompe disease patients under alglucosidase alfa enzyme replacement therapy. *J Inherit Metab Dis,* 35, 837-845.
16. Anderson, L.J., Henley, W., Wyatt, K.M., Nikolaou, V., Waldek, S., Hughes, D.A., Lachmann, R.H. and Logan, S. (**2014**) Effectiveness of enzyme replacement therapy in adults with late-onset Pompe disease: results from the NCS-LSD cohort study. *J Inherit Metab Dis,* 37, 945-952.
17. de Vries, J.M., van der Beek, N.A., Hop, W.C., Karstens, F.P., Wokke, J.H., de Visser, M., van Engelen, B.G., Kuks, J.B., van der Kooi, A.J., Notermans, N.C. et al. (**2012**) Effect of enzyme therapy and prognostic factors in 69 adults with Pompe disease: an open-label single-center study. *Orphanet J Rare Dis,* 7, 73.
18. de Vries, J.M., van der Beek, N.A., Kroos, M.A., Ozkan, L., van Doorn, P.A., Richards, S.M., Sung, C.C., Brugma, J.D., Zandbergen, A.A., van der Ploeg, A.T. et al. (**2010**) High antibody titer in an adult with Pompe disease affects treatment with alglucosidase alfa. *Mol Genet Metab,* 101, 338-345.
19. Patel, T.T., Banugaria, S.G., Case, L.E., Wenninger, S., Schoser, B. and Kishnani, P.S. (**2012**) The impact of antibodies in late-onset Pompe disease: a case series and literature review. *Mol Genet Metab,* 106, 301-309.
20. de Vries, J.M., Kuperus, E., Hoogeveen-Westerveld, M., Kroos, M.A., Wens, S.C., Stok, M., van der Beek, N.A., Kruijshaar, M.E., Rizopoulos, D., van Doorn, P.A. et al. (**2016**) Pompe disease in adulthood: effects of antibody formation on enzyme replacement therapy. *Genet Med.*
21. Cardone, M., Porto, C., Tarallo, A., Vicinanza, M., Rossi, B., Polishchuk, E., Donaudy, F., Andria, G., De Matteis, M.A. and Parenti, G. (**2008**) Abnormal mannose-6-phosphate receptor trafficking impairs recombinant alpha-glucosidase uptake in Pompe disease fibroblasts. *Pathogenetics,* 1, 6.
22. Raben, N., Takikita, S., Pittis, M.G., Bembi, B., Marie, S.K., Roberts, A., Page, L., Kishnani, P.S., Schoser, B.G., Chien, Y.H. et al. (**2007**) Deconstructing Pompe disease by analyzing single muscle fibers: to see a world in a grain of sand. *Autophagy,* 3, 546-552.
23. Settembre, C., Fraldi, A., Rubinsztein, D.C. and Ballabio, A. (**2008**) Lysosomal storage diseases as disorders of autophagy. Autophagy, 4, 113-114.
24. Huie, M.L., Chen, A.S., Tsujino, S., Shanske, S., DiMauro, S., Engel, A.G. and Hirschhorn, R. (**1994**) 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.
25. Boerkoel, C.F., Exelbert, R., Nicastri, C., Nichols, R.C., Miller, F.W., Plotz, P.H. and 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.
26. Dardis, A., Zanin, I., Zampieri, S., Stuani, C., Pianta, A., Romanello, M., Baralle, F.E., Bembi, B. and 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.
27. Bergsma, A.J., Kroos, M., Hoogeveen-Westerveld, M., Halley, D., van der Ploeg, A.T. and Pijnappel, W.W. (**2015**) Identification and characterization of aberrant GAA pre-mRNA splicing in pompe disease using a generic approach. *Hum Mutat,* 36, 57-68.
28. Rigo, F., Seth, P.P. and Bennett, C.F. (**2014**) Antisense oligonucleotide-based therapies for diseases caused by pre-mRNA processing defects. Adv Exp Med Biol, 825, 303-352.
29. McClorey, G. and Wood, M.J. (**2015**) An overview of the clinical application of antisense oligonucleotides for RNA-targeting therapies. *Curr Opin Pharmacol,* 24, 52-58.
30. Jirka, S. and Aartsma-Rus, A. (**2015**) An update on RNA-targeting therapies for neuromuscular disorders. *Curr Opin Neurol,* 28, 515-521.

31. Havens, M.A. and Hastings, M.L. (**2016**) Splice-switching antisense oligonucleotides as therapeutic drugs. *Nucleic Acids Res.*
32. Bergsma, A.J., In 't Groen, S.L.M., Verheijen, F.W., van der Ploeg, A.T. and Pijnappel, W.W. (**In press**) From cryptic towards canonical pre-mRNA splicing in Pompe disease: a pipeline for the development of antisense oligonucleotides. *Mol Ther Nucleic Acids.*
33. Lorson, C.L. and Androphy, E.J. (**2000**) An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene SMN. *Hum Mol Genet,* 9, 259-265.
34. Cartegni, L. and Krainer, A.R. (2002) Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat Genet,* 30, 377-384.
35. Kashima, T. and Manley, J.L. (**2003**) A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nat Genet,* 34, 460-463.
36. Singh, N.K., Singh, N.N., Androphy, E.J. and Singh, R.N. (**2006**) Splicing of a critical exon of human Survival Motor Neuron is regulated by a unique silencer element located in the last intron. *Mol Cell Biol,* 26, 1333-1346.
37. Hua, Y., Vickers, T.A., Okunola, H.L., Bennett, C.F. and Krainer, A.R. (**2008**) Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. *Am J Hum Genet,* 82, 834-848.
38. Singh, N.N., Shishimorova, M., Cao, L.C., Gangwani, L. and Singh, R.N. (**2009**) A short antisense oligonucleotide masking a unique intronic motif prevents skipping of a critical exon in spinal muscular atrophy. *RNA Biol,* 6, 341-350.
39. Merkin, J., Russell, C., Chen, P. and Burge, C.B. (**2012**) Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science,* 338, 1593-1599.
40. Lee, Y. and Rio, D.C. (**2015**) Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem,* 84, 291-323.
41. Nehlin, J.O., Just, M., Rustan, A.C. and Gaster, M. (**2011**) Human myotubes from myoblast cultures undergoing senescence exhibit defects in glucose and lipid metabolism. *Biogerontology,* 12, 349-365.
42. Borchin, B., Chen, J. and Barberi, T. (**2013**) Derivation and FACS-mediated purification of PAX3+/PAX7+ skeletal muscle precursors from human pluripotent stem cells. *Stem Cell Reports,* 1, 620-631.
43. Xu, C., Tabebordbar, M., Iovino, S., Ciarlo, C., Liu, J., Castiglioni, A., Price, E., Liu, M., Barton, E.R., Kahn, C.R. et al. (**2013**) A zebrafish embryo culture system defines factors that promote vertebrate myogenesis across species. *Cell,* 155, 909-921.
44. Chal, J., Oginuma, M., Al Tanoury, Z., Gobert, B., Sumara, O., Hick, A., Bousson, F., Zidouni, Y., Mursch, C., Moncuquet, P. et al. (**2015**) Differentiation of pluripotent stem cells to muscle fiber to model Duchenne muscular dystrophy. *Nat Biotechnol,* 33, 962-969.
45. Shelton, M., Kocharyan, A., Liu, J., Skerjanc, I.S. and Stanford, W.L. (**2016**) Robust generation and expansion of skeletal muscle progenitors and myocytes from human pluripotent stem cells. *Methods,* 101, 73-84.
46. Choi, I.Y., Lim, H., Estrellas, K., Mula, J., Cohen, T.V., Zhang, Y., Donnelly, C.J., Richard, J.P., Kim, Y.J., Kim, H. et al. (**2016**) Concordant but Varied Phenotypes among Duchenne Muscular Dystrophy Patient-Specific Myoblasts Derived using a Human iPSC-Based Model. *Cell Rep,* 15, 2301-2312.
47. Liu, S., Asparuhova, M., Brondani, V., Ziekau, I., Klimkait, T. and Schumperli, D. (**2004**) Inhibition of HIV-1 multiplication by antisense U7 snRNAs and siRNAs targeting cyclophilin A. *Nucleic Acids Res,* 32, 3752-3759.
48. Nlend, R.N. and Schumperli, D. (**2012**) Antisense genes to induce exon inclusion. *Methods Mol Biol,* 867, 325-347.
49. Kollberg, G. and Holme, E. (**2009**) Antisense oligonucleotide therapeutics for iron-sulphur cluster deficiency myopathy. *Neuromuscul Disord,* 19, 833-836.
50. Sun, X., Marque, L.O., Cordner, Z., Pruitt, J.L., Bhat, M., Li, P.P., Kannan, G., Ladenheim, E.E., Moran, T.H., Margolis, R.L. et al. (**2014**) Phosphorodiamidate morpholino oligomers suppress mutant huntingtin expression and attenuate neurotoxicity. *Hum Mol Genet,* 23, 6302-6317.
51. Drost, M.R., Schaart, G., van Dijk, P., van Capelle, C.I., van der Vusse, G.J., Delhaas, T., van der Ploeg, A.T. and Reuser, A.J. (**2008**) Both type 1 and type 2a muscle fibers can respond to enzyme therapy in Pompe disease. *Muscle Nerve,* 37, 251-255.
52. Schaaf, G.J., van Gestel, T.J., Brusse, E., Verdijk, R.M., de Coo, I.F., van Doorn, P.A., van der Ploeg, A.T. and Pijnappel, W.W. (**2015**) Lack of robust satellite cell activation and muscle regeneration during the progression of Pompe disease. *Acta Neuropathol Commun,* 3, 65.

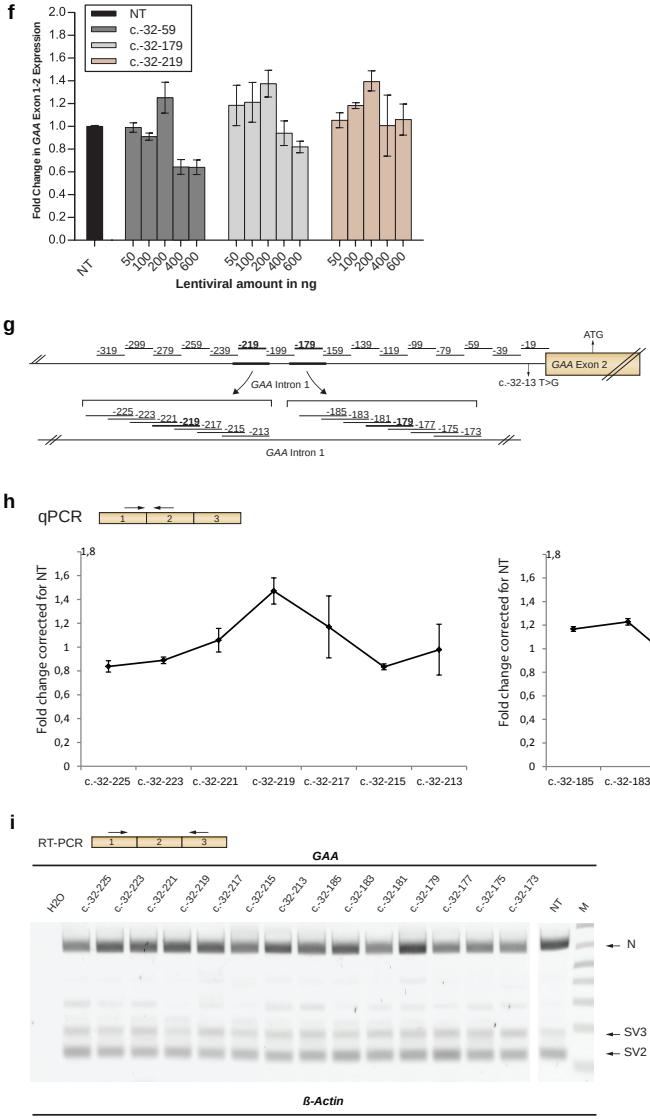
53. Hube, F. and Francastel, C. (**2015**) Mammalian introns: when the junk generates molecular diversity. *Int J Mol Sci,* 16, 4429-4452.
54. Droge, C., Schaal, H., Engelmann, G., Wenning, D., Haussinger, D. and Kubitz, R. (**2016**) Exon-skipping and mRNA decay in human liver tissue: molecular consequences of pathogenic bile salt export pump mutations. *Sci Rep,* 6, 24827.
55. Xue, Y., Schoser, B., Rao, A.R., Quadrelli, R., Vaglio, A., Rupp, V., Beichler, C., Nelson, S.F., Schapacher-Tilp, G., Windpassinger, C. et al. (**2016**) Exome Sequencing Identified a Splice Site Mutation in FHL1 that Causes Uruguay Syndrome, an X-Linked Disorder With Skeletal Muscle Hypertrophy and Premature Cardiac Death. *Circ Cardiovasc Genet,* 9, 130-135.
56. Iida, K., Ohkuma, Y., Hayashi, T., Katagiri, S., Fujita, T., Tsunoda, K., Yamada, H. and Tsuneoka, H. (**2016**) A novel heterozygous splice site OPA1 mutation causes exon 10 skipping in Japanese patients with dominant optic atrophy. *Ophthalmic Genet,* 37, 354-356.
57. Ebert, A.D., Liang, P. and Wu, J.C. (**2012**) Induced pluripotent stem cells as a disease modeling and drug screening platform. *J Cardiovasc Pharmacol,* 60, 408-416.
58. Heilker, R., Traub, S., Reinhardt, P., Scholer, H.R. and Sterneckert, J. (**2014**) iPS cell derived neuronal cells for drug discovery. *Trends Pharmacol Sci,* 35, 510-519.
59. Takahashi, K. and Yamanaka, S. (**2016**) A decade of transcription factor-mediated reprogramming to pluripotency. *Nat Rev Mol Cell Biol,* 17, 183-193.
60. Warlich, E., Kuehle, J., Cantz, T., Brugman, M.H., Maetzig, T., Galla, M., Filipczyk, A.A., Halle, S., Klump, H., Scholer, H.R. et al. (**2011**) Lentiviral vector design and imaging approaches to visualize the early stages of cellular reprogramming. *Mol Ther,* 19, 782-789.
61. Dambrot, C., van de Pas, S., van Zijl, L., Brandl, B., Wang, J.W., Schalij, M.J., Hoeben, R.C., Atsma, D.E., Mikkers, H.M., Mummery, C.L. et al. (**2013**) Polycistronic lentivirus induced pluripotent stem cells from skin biopsies after long term storage, blood outgrowth endothelial cells and cells from milk teeth. *Differentiation,* 85, 101-109.
62. Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B. and Speed, T.P. (**2003**) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res,* 31, e15.

SUPPLEMENTARY FIGURES AND TABLES

Supplementary Figure 1a-e



Supplementary Figure 1f-i



Supplementary Figure 1 A U7 snRNA screen to identify splicing repressors. *(a) In silico prediction of exonic and intronic splicing silencers around the GAA IVS1 variant. Algorithms from Human Splicing Finder 2.4.1 are indicated below the graph. (b) One-step cloning strategy for rapid cloning of AONs in the lentiviral U7 snRNA expression vector. A unique NsiI site was introduced in the U7 snRNA. AON sequences and the NsiI site were part of a forward primer in PCR, and a unique SalI site was included in the reverse PCR primer. (c) Cartoon of the region of the Cyclophilin A (CypA) gene that was targeted using a U7 snRNA-expressed AON (CyPA-E4) as described previously by Liu et al (47). (d) RT-PCR analysis of patient 1 fibroblasts in which the CypA pre-mRNA was targeted using CyPA-E4. As control, non-transduced cells were used (NT). The PCR strategy is shown above the gel.* β*-actin was used as loading control. (e) RT-qPCR analysis of the samples of (d). The PCR strategy is shown above the figure. (f) Testing of the optimal viral amount for detection of splicing modulation sequences. Patient 1 fibroblasts were infected with various lentiviruses at the amounts indicated. The optimum amount was determined to be 200 ng lentivirus per ml of medium. Data are means +/- SD of two biological replicates. Data points from 200 ng were taken from Fig. 2d (N = 3). NT: non-transduced. (g) Two hits from the screen shown in Fig. 2d were further tested in a microwalk using the U7 snRNA system. Primer locations are shown here. (h) Results of the microwalk, as analyzed by RT-qPCR. (i) As (h), using RT-PCR analysis. Results are expressed relative to NT fibroblasts and represent means +/- SD of three biological replicates. **P< 0.01.*

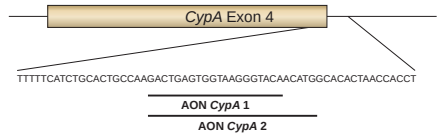
Supplementary Figure 2

a

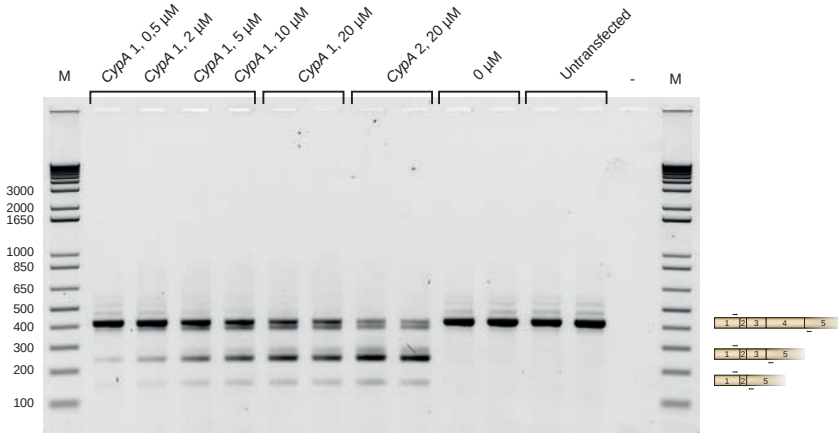
Name:	cDNA location	sequence 5' to 3'	nucleotides
CypA 1	*CypA* c.165_173+11	TGTACCCTTACCACTCAGTC	20
CypA 2	*CypA* c.165_173+16	CATGTTGTACCCTTACCACTCAGTC	25
AON 1	*GAA* c.-32-219_-200	GAGTGCAGAGCACTTGCACA	20
AON 2	*GAA* c.-32-224_-200	GAGTGCAGAGCACTTGCACAGTCTG	25
AON 3	*GAA* c.-32-187_-167	CCAGAAGGAAGGGCGAGAAAA	21
AON 4	*GAA* c.-32-190_-166	GCCAGAAGGAAGGGCGAGAAAAGCT	25
AON 5	*GAA* c.-32-64_-40	AAAGCAGCTCTGAGACATCAACCGC	25

reference sequence for cDNA annotation is NM_000152.3

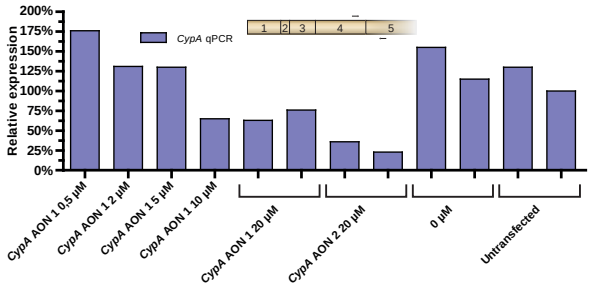
b



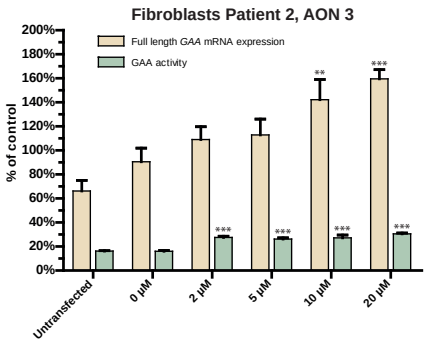
c



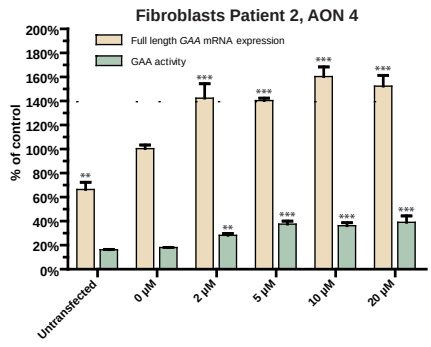
d



e

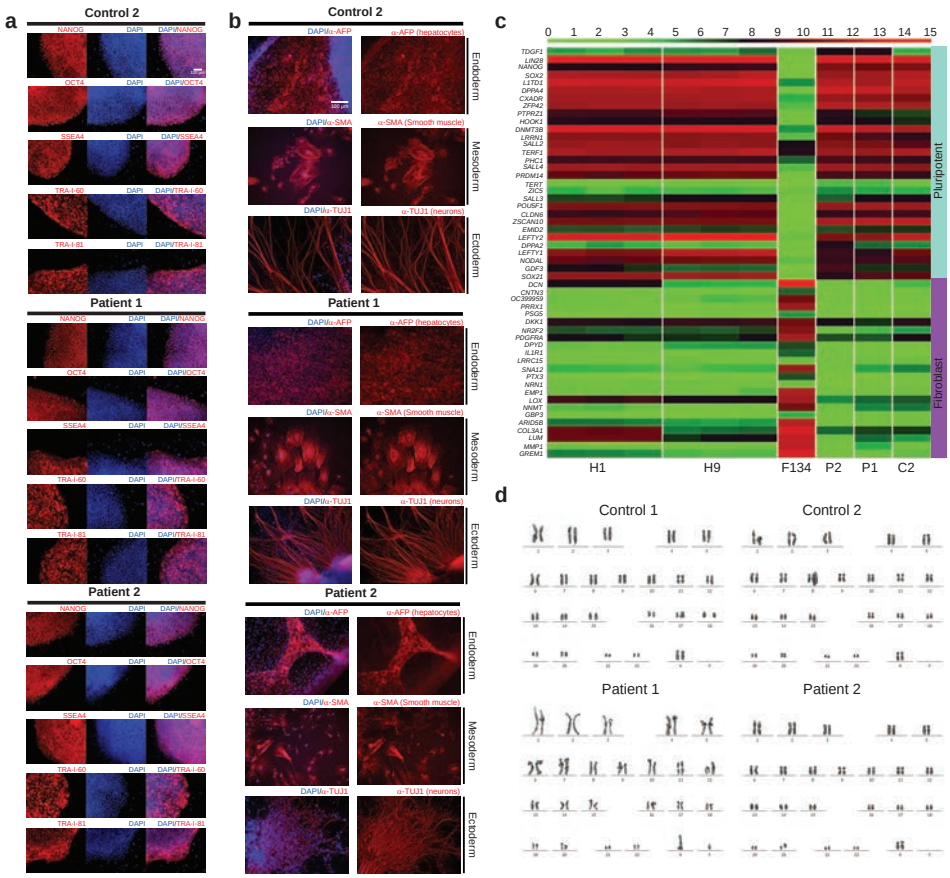


f

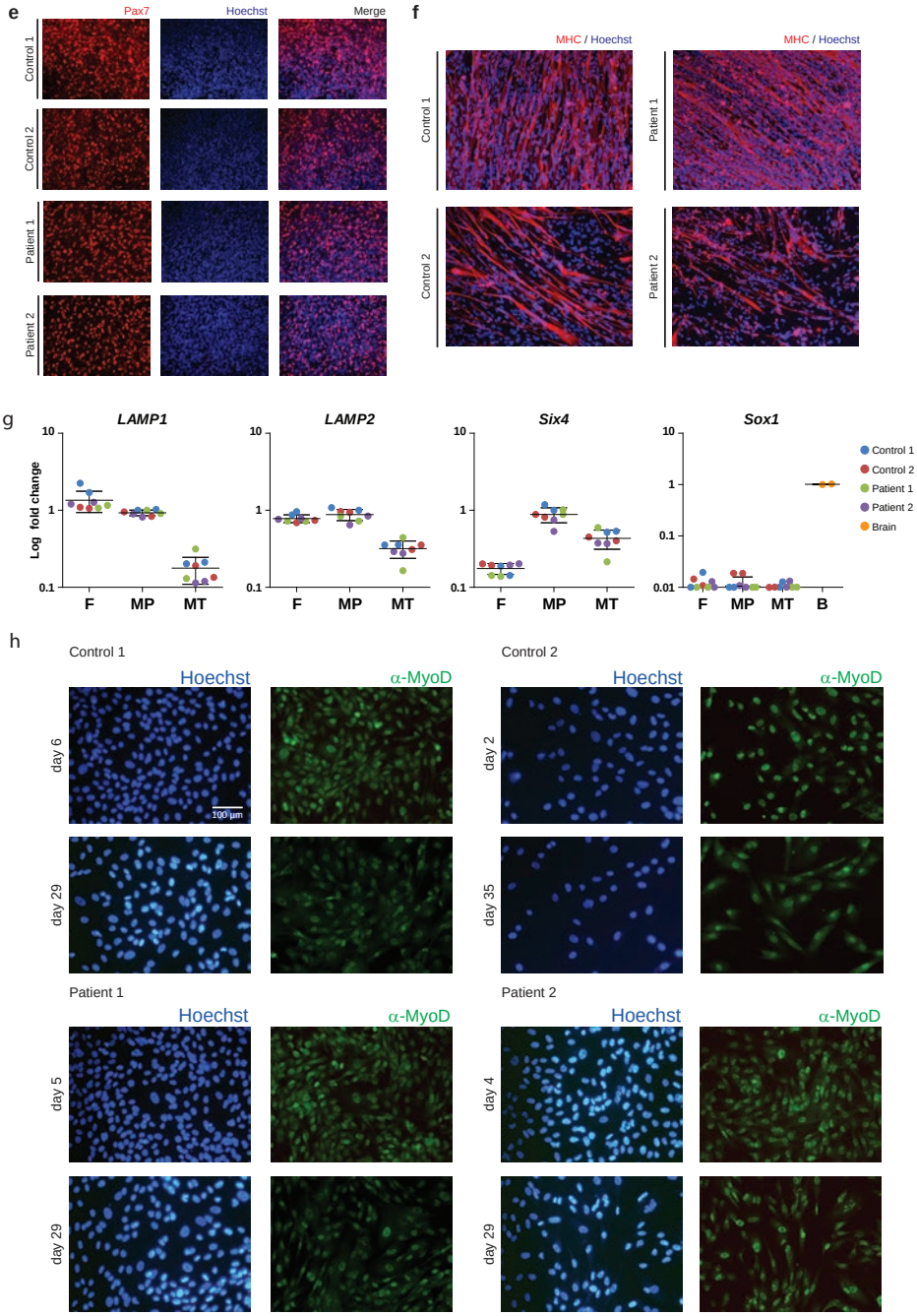


Supplementary Figure 2. PMO-based AONs promote exon inclusion in primary fibroblasts from Pompe patients. *(a) Sequences of PMO-AONs used. (b-d) Test of PMO-based AONs on the positive control gene CypA. (b) Location of AONs designed to block the splice donor of CypA exon 4. (c) Fibroblasts from patient 1 were transfected with AONs at various concentrations as indicated, and CyPA mRNAs were analyzed by RT-PCR. Cartoons at the right side of the gel indicate spliced products. (d) RT-qPCR analysis of exon 4 skipping of the experiment in (c). The cartoon highlights the primer locations. Data represent means of 3 technical replicates. (e-f) Promotion of GAA exon 2 inclusion. (e) Effect of AON 3 on GAA exon 2 inclusion measured using RT-qPCR analysis and on GAA enzymatic activity in fibroblasts from patient 2. Note that this patient has genotype IVS1, c.923A>C, and that the c.923A>C allele causes background expression of the N form of GAA mRNA. Data are expressed relative to average levels in healthy control fibroblasts and are means +/- SD from three biological replicates. (f) As (e) but with AON 4. Data for Supplementary Fig. 2e,f are means +/- SD from three biological replicates. *p < 0.05, **p < 0.01, ***p < 0.001.*

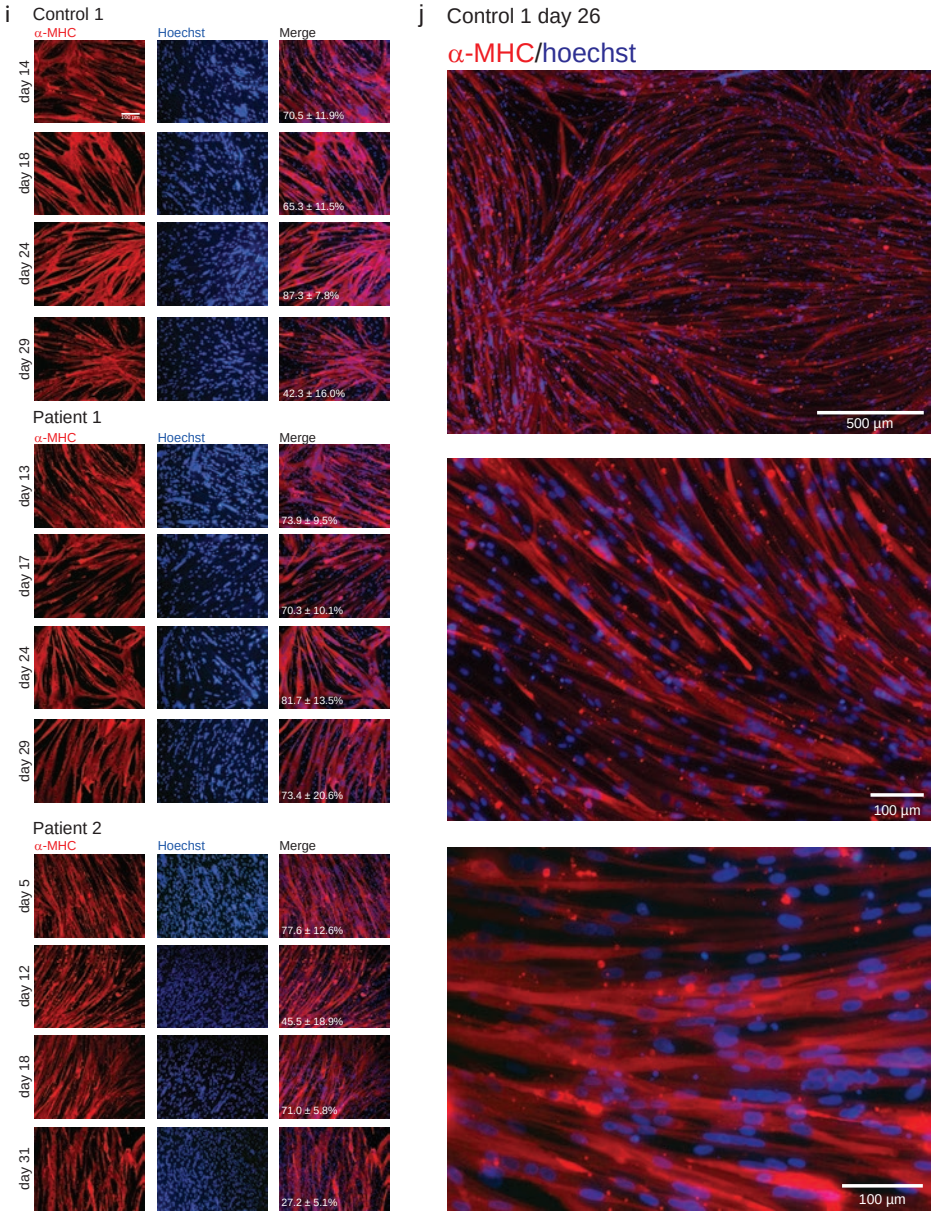
Supplementary Figure 3a-d



Supplementary Figure 3e-h



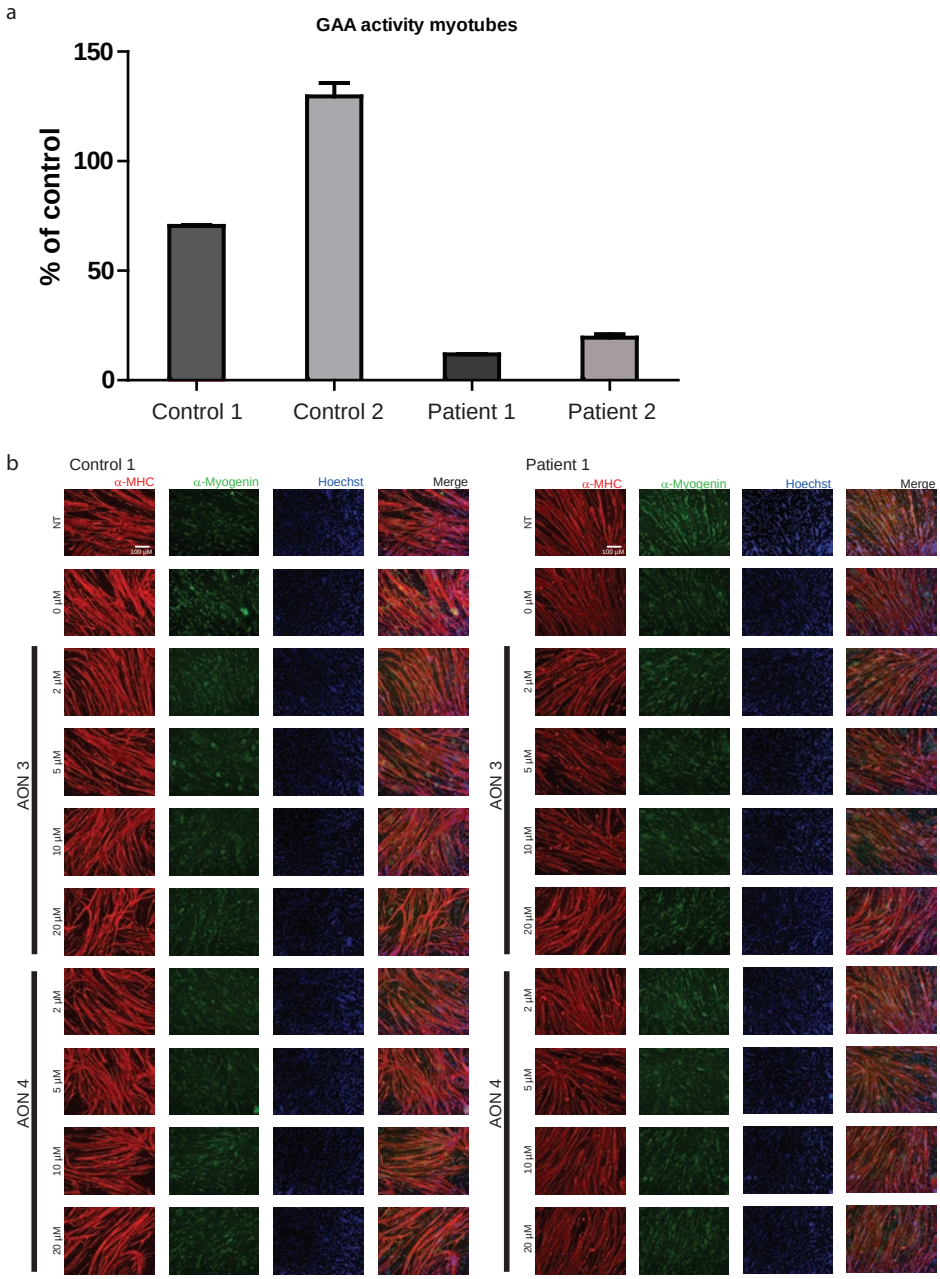
Supplementary Figure 3i-j



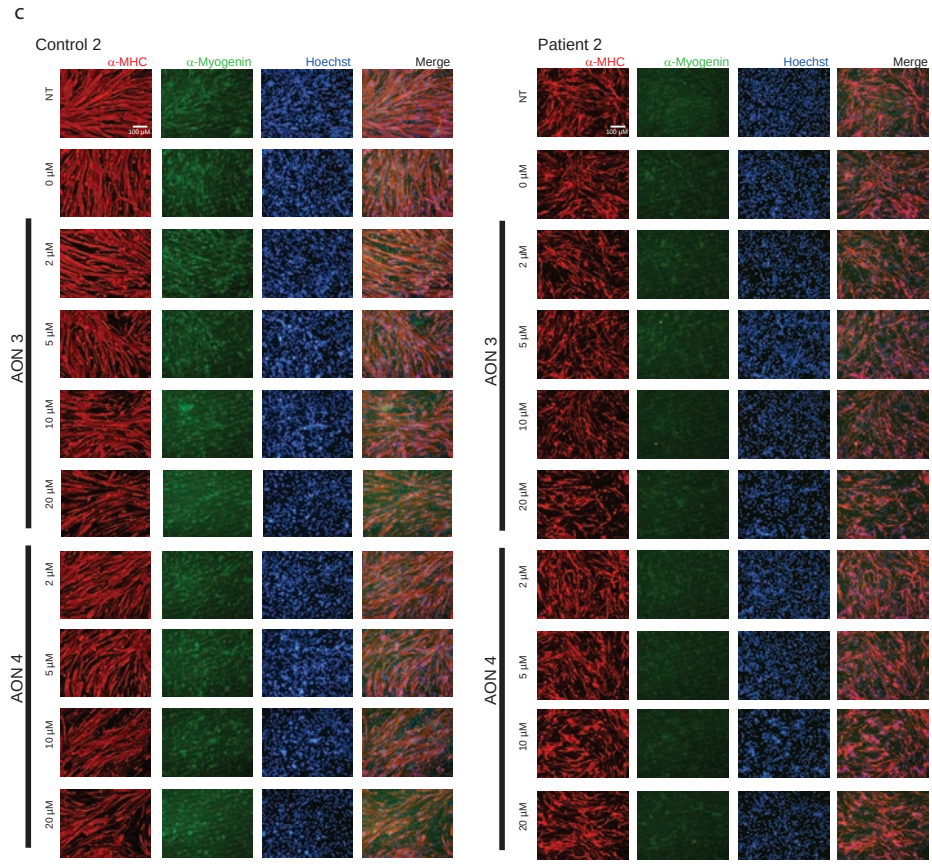
Supplementary Figure 3. Purification and expansion of iPS-derived myogenic progenitors. *(a-d) Generation and characterization of iPS cells. (a) Immunofluorescent analysis of iPS cells from control 2 and patient 1 and 2 with antibodies to Nanog, Oct4, SSEA4, TRA-I-60 an TRA-I-81 (red). DAPI was used to stain nuclei (blue). Control 1 iPS cells were published previously (61). (b) In vitro differentiation of iPS lines from (a) into derivatives of the three germ layers. Stainings for α-Fetoprotein (AFP) show hepatocytes (endoderm; red), stainings for smooth muscle actin (SMA) show smooth muscle cells (mesoderm, red), and neuron-specific class III α-tubulin (TUJ1) stainings show neurons (ectoderm, red). DAPI staining shows nuclei in blue. (c) Microarray analysis of mRNA expression of pluripotency and fibroblast genes. iPS cell are marked as P2, P1 and C2 (patients 2 and 1, and control 2, respectively). For comparison, human embryonic stem cell lines H1 and H9 and fibroblast line F134 were also analyzed. (d) Karyotype analysis of the four iPS lines used in this study. All lines have normal karyotypes. Representative karyotypes of 10 nuclei per cell line are shown. (e-j) Expansion and differentiation of purified iPS-derived myogenic progenitors. (e) Immunofluorescent staining for Pax7 (in red) in selected areas from a culture dish derived from non-purified myogenic progenitors following the 35-day differentiation protocol outlined in Fig. 3A I. Nuclei were stained with Hoechst (blue). (f) Myogenic progenitors from (e) were purified by FACS for HNK-1-/C-MET+ cells, and differentiated into myotubes, which were stained with an MF-20 antibody to MHC (red). Nuclei were stained with Hoechst (blue). Attempts to differentiate freshly purified myogenic progenitors using standard differentiation medium (DMEM/F12 1% ITS-X, 1% P/S/G) failed as cells died. Therefore, we first applied a 2-day period in DMEM HG, 2% horse serum, 1% P/S/G, followed by a 6 day period in standard differentiation medium. These results are shown here. (g-j) Characterization of expanded myogenic progenitors. (g) Equal amounts of total RNA from fibroblasts (F), purified and expanded myogenic progenitors (MP) and after differentiation into myotubes (MT) were analyzed by RT-qPCR analysis. Biological duplicates are shown. Lines represent means. A human brain autopsy sample was used as a positive control for Sox1 expression. (h) Immunofluorescent analysis of MyoD in expanded myogenic progenitors. Myogenic progenitors were expanded in proliferation medium and stained at the start of expansion and after expansion to ~1012 cells. Representative pictures are shown. (i). Capacity to differentiate into multinucleated myotubes during expansion. Myogenic progenitors were expanded and at several time points during expansion, a subculture from the expansion was differentiated for 4 days and stained for MHC expression (anti-MF20, red). Nuclei were stained with Hoechst (blue). The fusion index of each differentiation is shown at the bottom. Data are means +/- SDs of five technical replicates. We noted that, upon differentiation, myogenic progenitors from patient 2 (generated from iPS clone 1 and used in the experiments shown in Fig. 3, 4 and Supplementary Fig. 4) showed shorter myotubes and in some cases abnormal fusion into large round MHC-positive multinucleated structures. To investigate this further, we generated independent myogenic progenitors from this iPS clone, but these also showed similar properties. We then generated myogenic progenitors from an independent iPS clone 2 from patient 2 (which expressed pluripotency markers and showed in vitro differentiation into derivatives of the three germ layers, data not shown). These showed formation of multinucleated myotubes that were similar to those formed from patient 1 and the healthy controls, and we used these cells in Supplementary Fig. 3i. Myogenic progenitors from patient 2 clones 1 (Fig. 3b) and 2 (data not shown) showed highly similar proliferation curves upon expansion of at least 1012 in 31 days. (j) Examples of myogenic differentiation after expansion of myogenic progenitors. Staining was as in (i). Multiple aligned myonuclei were seen in extended myotubes.*

Supplementary Figure 4 Promotion of exon inclusion in patient-derived myotubes. *(a) GAA enzyme activitiy in iPS-derived multinucleated myotubes. Myogenic progenitors from the cells indicated were differentiated for 4 days, and GAA enzyme activity was determined. The average activity present in the two healthy controls paralleled those present in fibroblasts (data not shown), and was used to calculate the % residual activity in myotubes from the two patients. (b) Morphology of differentiated myotubes, obtained from purified myogenic progenitors from control 1 and patient 1, with and without AON treatment. Cells were stained with antibodies against Myosin Heavy Chain (MHC) (red), and Myogenin (green). Nuclei were visualized with Hoechst (blue). (c) Same as (b), but for control 2 and patient 2. (d-h) AONs promote exon 2 inclusion and GAA enzyme activity in patient-derived myotubes but not in myotubes from a healthy control. (d) Effect of AON 3 on GAA pre-mRNA splicing in myotubes from patient 2, measured with RT-qPCR analysis of individual splicing products. Data were normalized using a panel of reference genes (MyoD, Myog, LAMP1, LAMP2) that did not change during treatment (see (i)). (e) As (d), but using AON 4. (f) Effects of AON 3 and 4 on expression of the N form of GAA mRNA in myotubes from control 2. Normalization was as in (d). (g) Effects of AON 3 and 4 on GAA enzymatic activity in myotubes from patient 2. (h) Effects of AON 3 and 4 on GAA enzymatic activity in myotubes from control 2. (i) Effects of AON 3 and 4 on expression of reference genes (MyoD, Myog, LAMP1, LAMP2) in myotubes from patients and controls. These genes were used for normalization of all RT-qPCR data from myotubes, unless otherwise stated. In all experiments, data represent means +/- SD of three biological replicates. *p < 0.05, **p < 0.01, ***p < 0.001.*

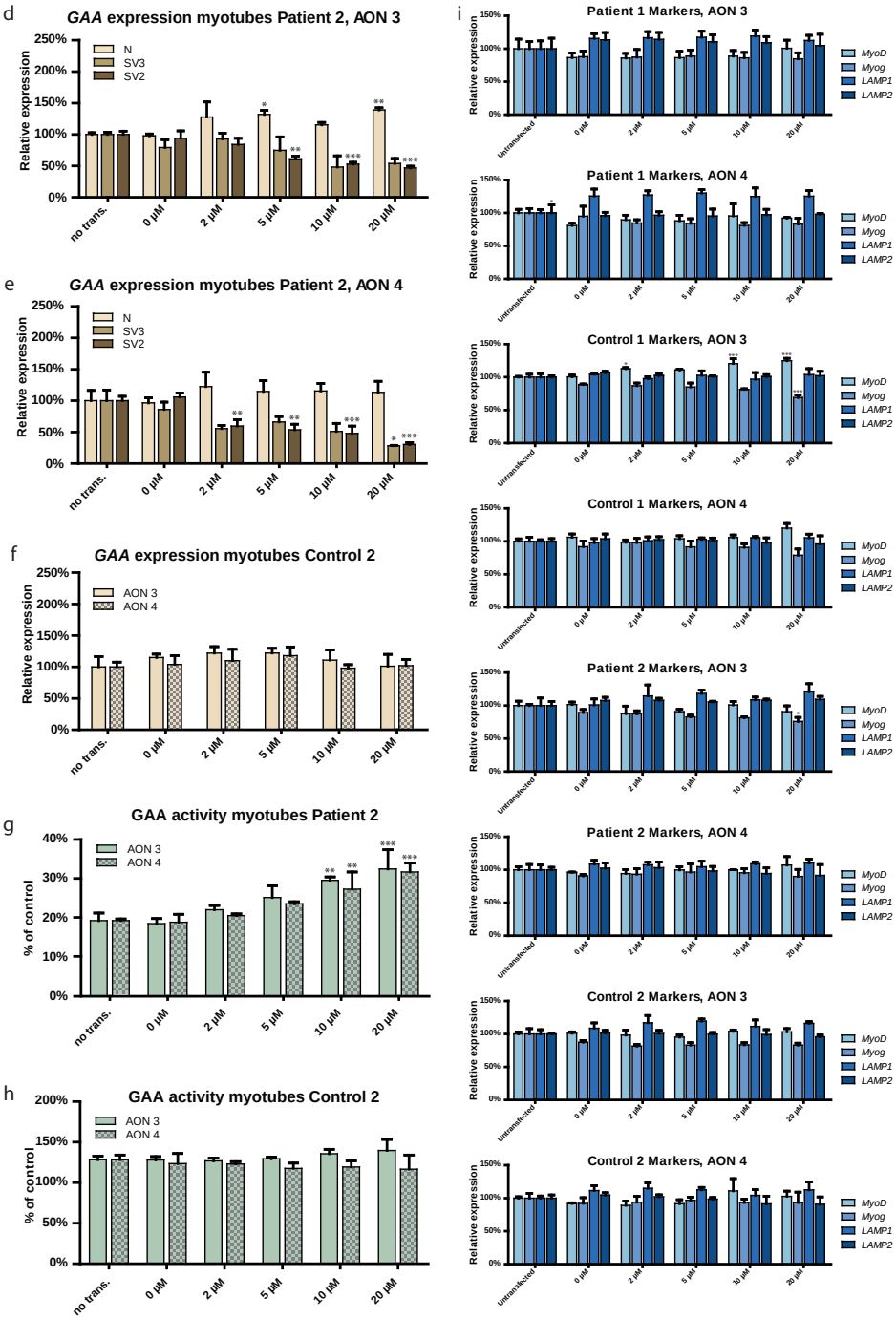
Supplementary Figure 4a-b



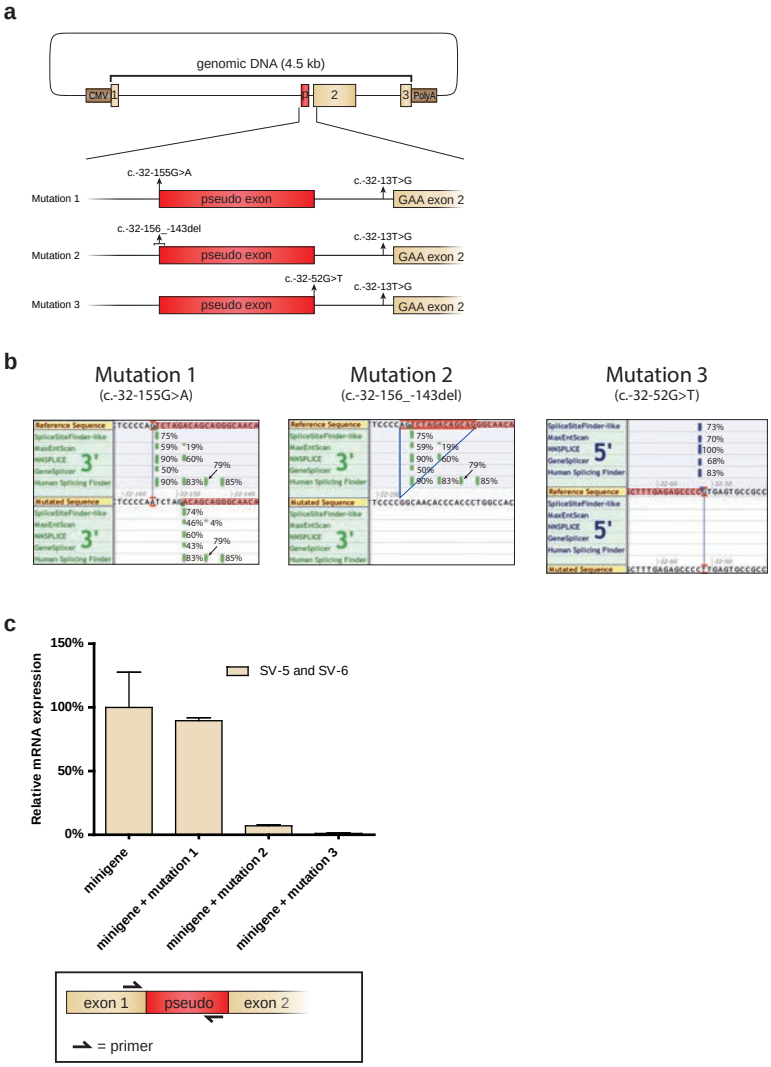
Supplementary Figure 4c



Supplementary Figure 4d-i

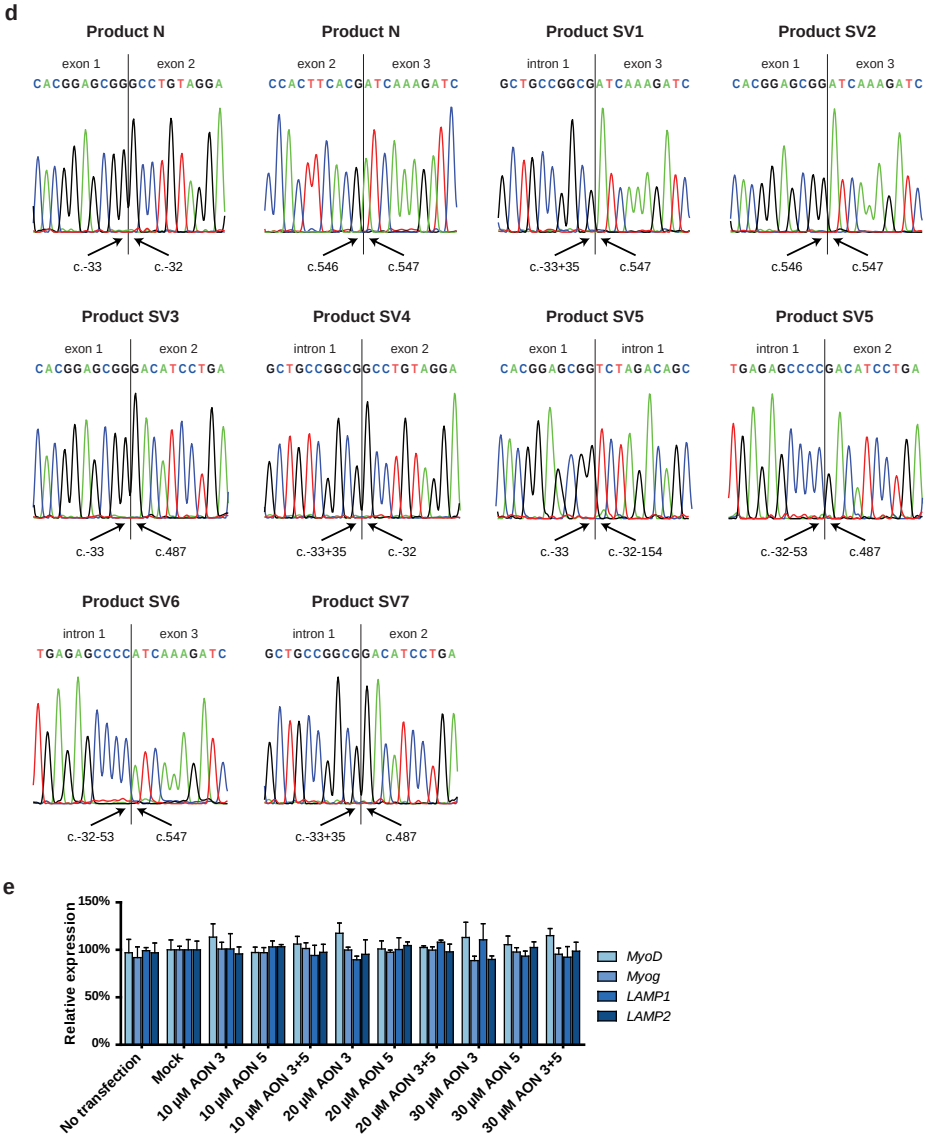


Supplementary Figure 5a-c



Supplementary Figure 5 Identification of a natural pseudo exon that competes with GAA exon inclusion. *(a-c) Mutations in splice sites of the natural pseudo exon abolish pseudo exon inclusion. (a) Cartoon of the minigene comprising a 4.5 kb genomic GAA sequence including exons 1-3. This sequence was obtained by PCR and cloned into pcDNA3.1. The natural pseudo exon is indicated along with the natural cryptic splice sites that were mutated by site directed mutagenesis. (b) Splicing prediction of the effect of the mutations shown in (a). Mutation 1 generated a new predicted 3'splice site 5 nt downstream, whereas mutations 2 and 3 completely abolished predicted 3'and 5'splice site, respectively. (c) Wild type and mutated minigenes were transfected in HEK293T cells, and expression of GAA splice variants containing the natural pseudo exon was quantified by RT-qPCR analysis using the primers indicated. (d) Sequence analysis of splicing products from Table 1. (e) AON treatment does not change expression of reference genes in myotubes. The experiment of Fig. 5b-d was analyzed by RT-qPCR for expression of the reference genes shown. Equal amounts of total RNA were used. Data represent means +/- SD of three biological replicates. *p < 0.05, **p < 0.01, ***p < 0.001.*

Supplementary Figure 5a-b



Supplementary Table 1. *Cell culture media tested for sustaining expansion of myogenic progenitors*

Medium no	Composition
1	DMEM/F12, 1% ITS-X and 1% P/S/G
2	HAM F10, 20% FBS and 1% P/S/G
3	HAM F10, 20% FBS, 1% P/S/G and 100 ng/ml FGF2
4	DMEM HG, 10% FBS and 1% P/S/G
5	DMEM HG, 10% FBS, 1% P/S/G and 100 ng/ml FGF2

Supplementary Table 2. *Antibodies used in experiments*

Name	Dilution	Company
Goat-α-NANOG	1:50	R&D systems (AF1997)
Goat-α-OCT4	1:100	Santa Cruz (sc-8629)
Mouse-α-SSEA4	1:100	Millipore (SCR001)
Mouse-α-Tra-I-60	1:100	Millipore (SCR001)
Mouse-α-Tra-I-80	1:100	Millipore (SCR001)
Mouse-α-SMA	1:50	Dako (M0851)
Mouse-α-AFP	1:200	Sigma (A8452)
Mouse-α-TUJ1	1:1000	Sigma (T8660)
Mouse-α-MF20	1:50	DSHB
Rabbit-α-Myogenin	1:100	Santa Cruz (sc-576)
Rabbit-α-MyoD	1:100	Santa Cruz (sc-304)
Mouse-α-Pax7	1:100	DSHB
Mouse-α-C-MET-APC	1:50	R&D systems (FAB3582A)
Mouse-α-HNK1-FITC	1:100	AVIVA SYSTEMS BIOLOGY (OASA02271)

Supplementary Table 3. *Composition of in vitro differentiation medium into three germ layer derivatives*

Medium	Component	Supplier
Endoderm/Mesoderm	DMEM High Glucose	Gibco
	20% FBS	Thermo scientific
	1% PSG	Gibco
	α-Thioglycerol (4µl/100ml)	Sigma
	1x NEAA	PAA
	0.1% ß-mercaptoethanol	Life technologies
Ectoderm	50% Neurobasal medium	Gibco
	50% DMEM/F12	Gibco
	1% PSG	Gibco
	0.1% ß-mercaptoethanol	Life technologies
	1:500 7.5% BSA fraction V	Gibco
	1:200 N2	Gibco
	1:100 B27 w/o VitA	Gibco

Supplementary Table 4. *Primers used for RT-qPCR, RT-PCR, cloning and sequencing*

Primer target	Sequence (5'-3')	Used for
β-Actin fw	AACCGCGAGAAGATGACCC	qPCR/RT-PCR
β-Actin rv	GCCAGAGGCGTACAGGGATAG	qPCR/RT-PCR
GAA Exon 1-2 fw	AAACTGAGGCACGGAGCG	qPCR
GAA Exon 1-2 rv	GAGTGCAGCGGTTGCCAA	qPCR
GAA Cryptic Exon 2 fw	GGCACGGAGCGGGACA	qPCR
GAA Cryptic Exon 2 rv	CTGTTAGCTGGATCTTTGATCGTG	qPCR
GAA Full Skip Exon 2 fw	AGGCACGGAGCGGATCA	qPCR
GAA Full Skip Exon 2 rv	TCGGAGAACTCCACGCTGTA	qPCR
GAA Pseudo Exon fw	AAACTGAGGCACGGAGCG	qPCR
GAA Pseudo Exon rv	GCAGCTCTGAGACATCAACCG	qPCR
α-Actinin fw	GAGACAGCGGCTAACAGGAT	qPCR
α-Actinin fw	ATTCCAAAAGCTCACTCGCT	qPCR
Six1 fw	GTCCAGAACCTCCCCTACTCC	qPCR
Six1 rv	CGAAAACCGGAGTCGGAACTT	qPCR
Six4 fw	CCATGCTGCTGGCTGTGGGAT	qPCR
Six4 rv	AGCAGTACAACACAGGTGCTCTTGC	qPCR
FGF2 fw	CAAAAACGGGGGCTTCTTCC	qPCR
FGF2 rv	GCCAGGTAACGGTTAGCACA	qPCR
Sox1 fw	GAGCTGCAACTTGGCCACGAC	qPCR
Sox1 rv	GAGACGGAGAGGAATTCAGAC	qPCR
MyoD fw	CACTCCGGTCCCAAATGTAG	qPCR
MyoD rv	TTCCCTGTAGCACCACACAC	qPCR
Myog fw	CACTCCCTCACCTCCATCGT	qPCR
Myog rv	CATCTGGGAAGGCCACAGA	qPCR
LAMP1 fw	GTGTTAGTGGCACCCAGGTC	qPCR
LAMP1 rv	GGAAGGCCTGTCTTGTTCAC	qPCR
LAMP2 fw	CCTGGATTGCGAATTTTACC	qPCR
LAMP2 rv	ATGGAATTCTGATGGCCAAA	qPCR
Fw-*U7*snRNA-smOPT	GCTCTTTTAGAATTTTTGGAGCAGGTTTTCTGACTTCG	Cloning
Rv-*U7*snRNA-smOPT	CGAAGTCAGAAAACCTGCTCCAAAAATTCTAAAAGAGC	Cloning
Fw-*U7*snRNA-NsiI	CCTGGCTCGCTACAGATGCATAGGAGGACGGAGGACG	Cloning
Rv-*U7*snRNA-NsiI	CGTCCTCCGTCCTCCTATGCATCTGTAGCGAGCCAGG	Cloning
M13 fw	GTAAAACGACGGGCCAG	Sequencing
M13 rv	CAGGAAACAGCTATGAC	Sequencing
GAA Exon1-3 fw	AGGTTCTCCTCGTCCGCCCGTTGTTCA	RT-PCR
GAA Exon1-3 rv	TCCAAGGGCACCTCGTAGCGCCTGTTA	RT-PCR
Fw-ms-*U7*snRNA-PstI	GCGCCTGCAGTAACAACATAGGAGCTGTG	Cloning
Rv-ms-*U7*snRNA-SalI	GCGCGTCGACCAGATACGCGTTTCCTAGGA	Cloning

Chapter 6

General Discussion

DISCUSSION

The current standard for genetic diagnosis of monogenic disorders is sequence analysis of the coding and intron flanking regions of the gene of interest from genomic DNA. This allows the identification of variants present in the gene. However, the functional consequences that these variants have cannot be assessed by this type of analysis. Functional diagnostic assays are widely available to evaluate function at the protein level, i.e. Western-blot, enzymatic activity measurements and cDNA expression assays, but functional assays are not well implemented for assessment at the level of pre-mRNA. This thesis presents a method for systematic unbiased screening for aberrant splice events and mRNA expression levels for single genes, using standard PCR analysis and Sanger sequencing. The assay was utilized for the screening of aberrant splicing in patients with Pompe disease, a recessive monogenic disorder caused by variants in the *GAA* gene. More than 20 previously unidentified aberrant splicing events were detected, some of which were caused by mutations that have not been related to splicing before, or were even never identified before. Further investigation of the identified aberrant splicing events led to the design and testing of antisense oligonucleotides to block aberrant and restore canonical splicing. These findings highlight the importance of unbiased functional investigation of the effect of variants on splicing. Furthermore, the results obtained from the splicing assay can be used to specify targets for antisense-based therapeutics as a new type of treatment for Pompe disease. This thesis further highlights the potential of AON mediated therapy for this disease by the identification of AONs that are able to promote exon 2 inclusion to close to wild-type levels in myotube cultures generated from patient-derived iPS cells with an IVS1/null background. The IVS1 variant occurs in 75% of Caucasian Pompe patients, highlighting that a novel therapy for this variant would be relevant for a large group of Pompe patients. Furthermore, application of the splicing assay for the generation of AONs represents a new pipeline for the development of new antisense therapy.

CURRENT PRACTICE IN DIAGNOSTICS OF MONOGENIC DISORDER

Sanger sequencing of genomic DNA is a technique that is well suited for the identification of potentially pathogenic variants present in regions that have the highest probability of harboring pathogenic variants that cause the monogenic disease. Genomic DNA can be obtained from blood, which presents an easy and harmless form of sample acquisition. However, one of the disadvantages of genomic DNA screening is that the effect of identified variants on protein activity are difficult to interpret. These effects can be exerted at the level of gene expression, pre-mRNA processing and/or protein translation. There are several ways to investigate these effects, either using prediction techniques or functional assays. Some of these methods are discussed below.

In silico prediction

The use of prediction algorithms is commonly used to estimate the pathogenicity of a variant. Algorithms exist for predicting the effect both on the protein level, by testing if the amino acid change has a noteworthy effect on protein folding/stability, or on the pre-mRNA level, where algorithms test the effect of variants on the process of splicing. Various algorithms have been developed for predicting the impact of variants on splicing. These algorithms are used to predict the presence of not only 3' and 5' splice sites (reviewed in (1)), but also splice enhancer and splice silencer elements (2,3). Online tools like Human Splicing Finder and Sroogle have been generated that integrate many of these algorithms in one interface. However, interpretation of the output in these programs can be challenging. Recently, the Automated Splice Site and Exon Definition Analyses (ASSEDA) tool has become available online that predicts the outcome of splicing in the context of a tested variant that makes an assumption of the splice sites that are utilized based on predicted splice site strengths (4). Even though all algorithms contribute to a better understanding of the effect of splicing mutations, most algorithms have been generated based on different datasets and do not always predict each splice site with the same accuracy. In this thesis, we used the Alamut® platform that uses five different algorithms to predict splice site strength of both 3' and 5' splice sites. The program has a user-friendly interface, making interpretation of the data straightforward. We performed both functional analysis and in silico prediction and were able to compare these results. Indeed, all splice sites either generated or destroyed by specific variants could be faithfully predicted, with the exception of the effect exerted by the IVS1 variant. However, various splice sites were predicted at locations that we could not identify with the functional assay, indicating that these algorithms generate a lot of false positive hits. Furthermore, predicted splice site strength did not always correlate with utilization of these splice sites, highlighting the difficulty of interpretation of these predictions. Even so, the use of splice prediction algorithms did help us to identify the mechanism behind pathogenicity of the IVS1 variant. Both 3' and 5' splice sites utilized by pseudo exon present in *GAA* intron 1 were predicted with relatively high prediction strength. More and more splice prediction programs are generated that are based on functional data and are able to more accurately predict splicing outcome (5,6). Combining in silico predictions with findings generated from functional investigation will lead to a better understanding of splicing in general. Furthermore, unbiased functional data produced by utilization of the splicing assay can aid in the generation of new algorithms, leading to more accurate splicing predictions in the future. However, it should be mentioned that algorithms will always generate predictions, and that these predictions should be verified using a functional assay.

cDNA expression and minigene assays

If the pathogenic effect on protein activity of an identified missense variant is unknown,

the consequences can be functionally investigated by introducing the variant in a cDNA expression vector. By comparing the activity of the protein from this vector with a wildtype counterpart the effect can be assessed. However, missense variants can also have an effect on pre-mRNA splicing. A widely used approach to functionally look at the effects that variants have on splicing is to introduce the variant including the relevant genomic region in an artificial minigene system. This technique is often used to study the process of splicing in genes that have extremely low or absent expression in tissue obtained from patients or if patient material cannot be obtained. The system is utilized to perform splicing analysis for well described disease genes, such as *DMD*, *CFTR* and *BRCA1* and 2 (7-10). Studies using a minigene have also been performed for the IVS1 variant in the *GAA* gene (11). These assays utilize a dedicated exon trapping vector in which only the exon of interest and short (30-150 bp) flanking intronic regions are included. The relatively small size of these constructs allows for the use of site directed mutagenesis to introduce any variant of interest. The system represents a quick and easy way to analyze the effect of selected variants on pre-mRNA splicing. However, the assay is particularly biased, because deep intronic sequences and natural splice site sequences of neighboring exons are missing, placing the exon out of its in vivo context. A more accurate way of testing the effects of splicing would be to insert the exon of interest including neighboring introns and exons in an expression vector, as long as the final vector size does not become too large for using site directed mutagenesis (>12 kb). In this thesis, we implemented two minigenes that fit these criteria. One of these minigenes contains the genomic region of *GAA* exon 1 to exon 3 and is used to study the effects of, among others, the IVS1 variant. We show that residual expression of the IVS1 minigene is about 20% of wildtype levels. Furthermore, expression of the aberrant transcripts identified from this minigene replicate the *in vivo* situation, with high expression of aberrantly spliced products from the IVS1 minigene compared to the wildtype minigene. An important finding is that this minigene harbors full length *GAA* intron 1 including the region that we identified to be important for the mechanism of IVS pathogenicity. The minigene used in a previous study to look at this mechanism (11) contained 50 bp of intronic sequence of intron 1, missing the pseudo-exon identified in this thesis, which could have implications for the recapitulation of the endogenous splicing process. These results highlight the importance of maintaining as much of the in vivo intron/exon context as possible in these type of experiments.

Testing the effect of splicing variants in their native context

Although the minigene system can be used for functional investigation, it should be noted that the minigene system does not always reflect what happens in vivo. Several factors play a role in this, including lack of the complete genomic context and the major overexpression in such a system. Another important factor is that pre-mRNA splicing can be tissue-specific, and thus have a different outcome in different cell types used for testing

splicing. One of the most accurate ways to test splicing in its native environment is by analysis of a tissue biopsy from the patient itself. However, these biopsies cannot always be readily obtained, or can be a significant burden for the patient. In the case of Pompe disease muscle biopsies should be taken from patients who already have weak muscles, highlighting the importance of alterative means to model disease. A good way to model Pompe disease progression as it seen in patients is the use of animal models. A GAA knock-out mouse model exists that is able to accurately replicate infantile onset Pompe disease phenotype (12). Although an animal model would be beneficial for the study of juvenile and adult Pompe disease, no models are currently available that replicate the IVS1 phenotype. This model would be a valuable addition to current models and can be used to study the in vivo situation more accurately. It should however be noted that the process of splicing is differentially regulated between different organisms, which makes the choice of animal models a complex task. Another model to study splicing in the tissue of interest can be generated by using human induced pluripotent stem cells (hiPS) for the generation of cells from a specific lineage. Protocols for differentiation into almost any tissue types of interest have been developed in recent years. These protocols allow for the generation of cell cultures that can be used for modeling monogenic disorders, including β-thalassaemia (hematopoietic lineage), Niemann–pick disease type C1 (neuronal lineage) and cystic fibrosis (mature airway epithelial cells) ((13-15) reviewed in (16)). These examples show the enormous potential of these systems. For the modeling of non-classic Pompe disease, a disorder in which skeletal muscle tissue is mainly affected, we used hiPS cells generated from Pompe patients carrying the IVS1 variant to generate skeletal muscle progenitors. These progenitors were further differentiated into muscle tubes that can be used for modelling skeletal muscle tissue. With these cells we were able to model the splicing defects caused by IVS1 variant in a cell culture system that correctly replicates splicing patterns in vivo, indicating this is a powerful method for modeling juvenile and adult onset Pompe disease.

Other functional assays to in patient derived RNA

Whether samples are obtained from muscle biopsies, skin biopsies or from myotube cultures generated from hiPS, RNA from patient material can yield valuable information about the effect that splicing has on disease. Although the splicing assay is well suited for detection of aberrant splicing, there are various other methods that can be used to investigate the process of pre-mRNA splicing. One of these methods is the exon-junction array (17). This is a micro-array containing probes for splice junctions of all known human mRNA/EST transcripts. With this micro-array, the whole transcriptome can be screened for alternative splicing events, and can thus also be used in screening of diseases with unknown etiology. A drawback of this assay is that it cannot screen for aberrant splicing events that have not been detected before, hampering identification of unknown aberrant products. Still,

cryptic splice site choice most often leads to less canonically spliced products. These can be assessed with such an array by looking at expression levels of genes and even individual exons within those genes, which can give valuable information on the process of splicing. Another technique that is becoming increasingly important in genetic diagnostics is RNA-sequencing (RNA-seq). With this technique, a plethora of aberrant splicing events can be unbiasedly detected (reviewed in (18)). However, to obtain full coverage of the gene of interest, deep-sequencing is often required, which makes the assay expensive for diagnostics of single gene disorders. Although at present RNA-seq might not be suitable for implementation in diagnostics of monogenic disease, the assay will become cheaper in the future. Combined with whole genome sequencing, these techniques will most likely eventually replace currently screening methods used in genetic diagnostics altogether.

POSSIBILITIES FOR OPTIMIZATION OF THE SPLICING ASSAY

The splicing assay is based on existing PCR-based technology and can be performed fairly quickly. However, the assay is not as quick as the standard screening of genomic DNA in a diagnostic setting. For this reason, optimization of the protocol is preferred before implementation in diagnostics. There are several parts of the assay that are amendable for optimization. One of these is the sequencing method used to analyze the PCR products. Sanger sequencing allows for accurate identification of (splice) variants in a sample containing one product, and is considered to be the gold standard in genetic diagnostics of monogenic disorders. However, data generated in this type of sequencing becomes more challenging to analyze if two different products are present in the sample, and almost impossible to read if three of more products are present. For this reason, PCR samples that contain multiple products first need to be separated, either via TOPO-cloning or amplification of specific products. To circumvent this problem, next-generation sequencing technology can be used to discriminate between multiple products within one sample due to massive parallel sequencing. Although this technique is still relatively expensive, sequencing platforms like the MiSeq™ (Illumina) and the Ion-Torrent™(Thermo Fisher Scientific) could allow for cheaper sequencing. Furthermore, due to the limited amount of unique reads, in-depth sequencing is not necessary, which further reduces costs. It should be noted that these techniques need a certain amount of input DNA, so multiple samples need to be run at the same time. This leads to the second step of optimization of the splicing assay. Multiplex PCR allows for the amplification of more than one PCR product at the same time in one reaction. This strategy is currently under investigation in combination with the MiSeq™ sequencing system for forensic analysis of STR markers (19) and could potentially be utilized for easy detection of aberrant splicing events in samples from patients with Pompe disease as well. Another possibility is the usage of the PacBio system for single molecule real-time (SMRT) sequencing (20). With this technique, unique reads of up to 10 kilobases can be detected. SMRT can be used for targeted sequencing and can

thus be used for sequencing of transcripts from individual genes, which will dramatically reduce costs compared to whole genome sequencing techniques like RNA-seq.

GAINING MECHANISTIC INSIGHT OF SPLICING USING THE SPLICING ASSAY

The unbiased nature of the splicing assay can be used as a platform to study the functional consequences that variants have on splicing and thus to study the mechanism of splicing. We investigated many aberrant splicing events caused by variants located close to a canonical splice site. The pathogenic effect of all of these variants was caused by weakening of the canonical splice site, leading to skipping of the exon that harbored the variant close to one of its splice sites, or by usage of a nearby cryptic splice site. However, several variants were not located near canonical splice sites, e.g. missense variants c.1256A>T and c.1927G>A, and the deep intronic variant c.2190-345A>G. Although these variants were not close to normal splice sites, we identified that these variants had an effect on splicing. The effects on splicing of variants which are not located close to canonical splice sites are often attributed to the disruption of splicing enhancer or silencer elements. Many reviews highlight the importance of splice enhancer and silencer elements in relation to disease (figure 1a, modified from (21) and (22)). However, in our study we mainly noticed that pathogenic splicing variants act as a modulator of splice site strength, either by weakening a canonical splice site, or strengthening a cryptic splice site. A good example our efforts to investigate the mechanism of pathogenicity of the IVS1 variant. Our primary focus was to identify a splice silencer that acted upon the *GAA* exon 2 splice site. Blockage of this site with an AON would then be tested to obtain more *GAA* exon 2 inclusion in IVS1 patient cells. However, blockage of predicted splice silencer sites did not lead to higher exon 2 inclusion (data not shown in thesis). Instead, blockage of cryptic splice sites residing within *GAA* intron 1 did result in upregulation of canonical exon two splicing in the background of the IVS1 variant, highlighting the importance of splice site strength. Although mutation of splice silencer or enhancer elements plays an important role in disease, we propose an additional mechanism in which splice site strength plays a significant role (figure 1b). In any respect, whether a variant affects splicing by mutation of a splice enhancer or element or a cryptic or canonical splice site, antisense modulation of splicing remains a potential therapeutic option.

THERAPY FOR POMPE DISEASE

Pompe disease is one of the few muscle disorders for which treatment is currently available in the form of enzyme replacement therapy (ERT). The therapy is based on intravenous injections of Myozyme®, a human recombinant form of GAA. Treatment of infantile Pompe patients with Myozyme® increased survival in these patients showing that the treatment is effective. Without treatment, these patients rarely survive the first year of life. Currently

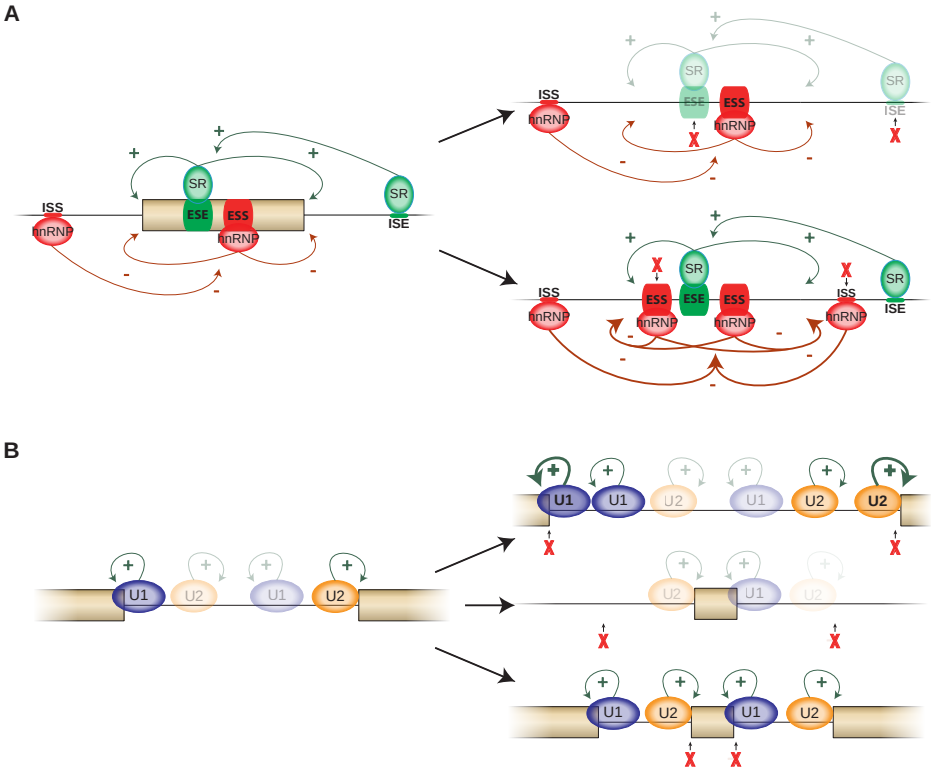


Figure 1. Mechanisms of how splicing can be influenced by (pathogenic) variants. *(A) Cartoon depicting an example of how exon recognition is regulated by intronic and exonic splicing enhancers (ISE and ESE) and silencers (ISS and ESS). Variants present in ISE and ESE elements can weaken the effect of these elements (indicated by lower opacity of these elements), causing a negative shift in the balance of exon recognition, leading towards exon skipping. Variants that generate a new ISS of ESS element have the same effect (indicated by thickened red lines). (B) Cartoon depicting an example of how exon recognition is regulated by changes in canonical or cryptic splice site strength due to the presence of (pathogenic) variants. Here we highlight three possibilities for alternative splice site usage. First, exonic variants that generate strong cryptic splice sites can cause a shift in splice site usage from the canonical site towards the newly generated cryptic splice site (exemplified by patient 8 in Chapter 2). Second, the presence of a variant close to a canonical splice site can cause loss of recognition of this splice site. This can lead to alternative splicing, for instance by utilization of a weak natural cryptic splice site (exemplified by patient 4 in Chapter 3). A third option is the generation of new or strengthening of an existing intronic cryptic splice site due to the presence of an intronic variant (exemplified by patient 1 and 2 in Chapter 4).*

the oldest patient diagnosed with infantile onset Pompe disease has reached 18 years of age, indicating the potential of ERT (23). However, not all Pompe patients benefit from ERT. Pompe patients with juvenile or adult onset of the disease show heterogeneous responses to the medication. One possible cause for the variable response is the generation of antibodies against human recombinant GAA (24). However, a recent study from our center suggested little impact of antibody formation in the majority of adult Pompe patients (De Vries et al., GIM, in press). This also applies to Furthermore, the costs of the medication are high, with an average cost of 400,000 euro annually. This warrants the research for additional

treatment strategies for Pompe disease. We focused our research on the identification of specific antisense oligonucleotides that can alter aberrant pre-mRNA splicing in patients that have variants that affect this process.

AON therapy as treatment for disease:

Since 20 years antisense oligonucleotides have been of particular interest to the scientific community for their therapeutic potential. A simple Pubmed search with the words “antisense oligonucleotide therapy” results in almost 7000 hits. However, only two antisense therapeutics (Fomivirsen and Mipomersen) have currently been approved by the American Food and Drug Association (FDA). Efficacy of AON therapy has been demonstrated for a range of diseases in cell culture systems, and may have been verified in vivo in mouse models, but translation towards clinical application has proven to be difficult. One of the main causes has been the fact that delivery of AONs towards the target tissue is problematic. This was particularly evident in a phase III clinical trial, in which an AON targeting dystrophin in patients with DMD failed to meet primary endpoints. Several factors were indicated to play a role in this outcome. One of these was the setup of the clinical trial itself, from which lessons can be learned for future trials (25). Furthermore, this particular trial was performed with a type of AON that was designed with an older 2’O-methyl phosphorothioate (2’O-Me-PS) type of backbone chemistry. Currently, newer backbone chemistries are available which have a higher binding affinities compared to the 2’O-Me-PS backbone, including morpholino (PMO) and peptide nucleic acid (PNA) chemistries (26,27). However, these chemistries do not carry a charge and are therefore more difficult to deliver in vivo. To enhance uptake of these newer type of backbones, many conjugates have been applied to AONs improving delivery to the target tissue (reviewed in (28)). Furthermore, recent advances in nanoparticle delivery provide another way for improving uptake (29). Techniques to improve uptake through optimizing infusion solutions are also currently investigated (30). Application of these improved AON delivery methods could provide a solid base for testing the AONs identified in this thesis.

CONCLUSION

Current genetic diagnostic screening of variants gives limited insight in the mechanism of pathogenicity of variants. These insights are important, because they can provide a basis for new therapeutic strategies. Use of the splicing assay can result in a better understanding of the pathogenic effect of the variants identified. Implementation of such assay in diagnostics of monogenic disorders will lead to a better understanding of the process of splicing in general, and may aid in the development of new antisense therapeutics. Further investigation towards the efficacy and potential of AONs identified in this thesis might lead to a new type treatment for Pompe disease.

REFERENCES

1. Jian, X., Boerwinkle, E. and Liu, X. (**2014**) In silico tools for splicing defect prediction: a survey from the viewpoint of end users. *Genet Med*, 16, 497-503.
2. Sironi, M., Menozzi, G., Riva, L., Cagliani, R., Comi, G.P., Bresolin, N., Giorda, R. and Pozzoli, U. (**2004**) Silencer elements as possible inhibitors of pseudoexon splicing. *Nucleic Acids Res*, 32, 1783-1791.
3. Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q. and Krainer, A.R. (**2003**) ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res*, 31, 3568-3571.
4. Mucaki, E.J., Shirley, B.C. and Rogan, P.K. (**2013**) Prediction of mutant mRNA splice isoforms by information theory-based exon definition. *Hum Mutat*, 34, 557-565.
5. Soukarieh, O., Gaildrat, P., Hamieh, M., Drouet, A., Baert-Desurmont, S., Frebourg, T., Tosi, M. and Martins, A. (**2016**) Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. *PLoS Genet*, 12, e1005756.
6. Jian, X., Boerwinkle, E. and Liu, X. (**2014**) In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res*, 42, 13534-13544.
7. Giorgi, G., Casarin, A., Trevisson, E., Dona, M., Cassina, M., Graziano, C., Picci, L., Clementi, M. and Salviati, L. (**2015**) Validation of CFTR intronic variants identified during cystic fibrosis population screening by a minigene splicing assay. *Clin Chem Lab Med*, 53, 1719-1723.
8. Sanz, D.J., Acedo, A., Infante, M., Duran, M., Perez-Cabornero, L., Esteban-Cardenosa, E., Lastra, E., Pagani, F., Miner, C. and Velasco, E.A. (**2010**) A high proportion of DNA variants of BRCA1 and BRCA2 is associated with aberrant splicing in breast/ovarian cancer patients. *Clin Cancer Res*, 16, 1957-1967.
9. Hendriks, G., Morolli, B., Calleja, F.M., Plomp, A., Mesman, R.L., Meijers, M., Sharan, S.K., Vreeswijk, M.P. and Vrieling, H. (**2014**) An efficient pipeline for the generation and functional analysis of human BRCA2 variants of uncertain significance. *Hum Mutat*, 35, 1382-1391.
10. Malueka, R.G., Yagi, M., Awano, H., Lee, T., Dwianingsih, E.K., Nishida, A., Takeshima, Y. and Matsuo, M. (**2011**) Antisense oligonucleotide induced dystrophin exon 45 skipping at a low half-maximal effective concentration in a cell-free splicing system. *Nucleic Acid Ther*, 21, 347-353.
11. Dardis, A., Zanin, I., Zampieri, S., Stuani, C., Pianta, A., Romanello, M., Baralle, F.E., Bembi, B. and 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.
12. Bijvoet, A.G., van de Kamp, E.H., Kroos, M.A., Ding, J.H., Yang, B.Z., Visser, P., Bakker, C.E., Verbeet, M.P., Oostra, B.A., Reuser, A.J. et al. (**1998**) Generalized glycogen storage and cardiomegaly in a knockout mouse model of Pompe disease. *Hum Mol Genet*, 7, 53-62.
13. Fan, Y., Luo, Y., Chen, X., Li, Q. and Sun, X. (**2012**) Generation of human beta-thalassemia induced pluripotent stem cells from amniotic fluid cells using a single excisable lentiviral stem cell cassette. *J Reprod Dev*, 58, 404-409.
14. Efthymiou, A.G., Steiner, J., Pavan, W.J., Wincovitch, S., Larson, D.M., Porter, F.D., Rao, M.S. and Malik, N. (**2015**) Rescue of an in vitro neuron phenotype identified in Niemann-Pick disease, type C1 induced pluripotent stem cell-derived neurons by modulating the WNT pathway and calcium signaling. *Stem Cells Transl Med*, 4, 230-238.
15. Wong, A.P., Bear, C.E., Chin, S., Pasceri, P., Thompson, T.O., Huan, L.J., Ratjen, F., Ellis, J. and Rossant, J. (**2012**) Directed differentiation of human pluripotent stem cells into mature airway epithelia expressing functional CFTR protein. *Nat Biotechnol*, 30, 876-882.
16. Spitalieri, P., Talarico, V.R., Murdocca, M., Novelli, G. and Sangiuolo, F. (**2016**) Human induced pluripotent stem cells for monogenic disease modelling and therapy. *World J Stem Cells*, 8, 118-135.
17. Shen, S., Warzecha, C.C., Carstens, R.P. and Xing, Y. (**2010**) MADS+: discovery of differential splicing events from Affymetrix exon junction array data. *Bioinformatics*, 26, 268-269.
18. Byron, S.A., Van Keuren-Jensen, K.R., Engelthaler, D.M., Carpten, J.D. and Craig, D.W. (**2016**) Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet*.
19. Kim, E.H., Lee, H.Y., Yang, I.S., Jung, S.E., Yang, W.I. and Shin, K.J. (**2016**) Massively parallel sequencing of 17 commonly used forensic autosomal STRs and amelogenin with small amplicons. *Forensic Sci Int Genet*, 22, 1-7.

20. Roberts, R.J., Carneiro, M.O. and Schatz, M.C. (**2013**) The advantages of SMRT sequencing. *Genome Biol*, 14, 405.
21. Lee, Y. and Rio, D.C. (**2015**) Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem*, 84, 291-323.
22. Scotti, M.M. and Swanson, M.S. (**2016**) RNA mis-splicing in disease. *Nat Rev Genet*, 17, 19-32.
23. van der Ploeg, A.T. and Reuser, A.J. (**2008**) Pompe's disease. *Lancet*, 372, 1342-1353.
24. van Gelder, C.M., Hoogeveen-Westerveld, M., Kroos, M.A., Plug, I., van der Ploeg, A.T. and Reuser, A.J. (**2015**) Enzyme therapy and immune response in relation to CRIM status: the Dutch experience in classic infantile Pompe disease. *J Inherit Metab Dis*, 38, 305-314.
25. Merlini, L. and Sabatelli, P. (**2015**) Improving clinical trial design for Duchenne muscular dystrophy. *BMC Neurol*, 15, 153.
26. Larsen, H.J., Bentin, T. and Nielsen, P.E. (**1999**) Antisense properties of peptide nucleic acid. *Biochim Biophys Acta*, 1489, 159-166.
27. Taylor, M.F., Paulauskis, J.D., Weller, D.D. and Kobzik, L. (**1996**) In vitro efficacy of morpholino-modified antisense oligomers directed against tumor necrosis factor-alpha mRNA. *J Biol Chem*, 271, 17445-17452.
28. Boisguerin, P., Deshayes, S., Gait, M.J., O'Donovan, L., Godfrey, C., Betts, C.A., Wood, M.J. and Lebleu, B. (**2015**) Delivery of therapeutic oligonucleotides with cell penetrating peptides. *Adv Drug Deliv Rev*, 87, 52-67.
29. Petrilli, R., Eloy, J.O., Marchetti, J.M., Lopez, R.F. and Lee, R.J. (**2014**) Targeted lipid nanoparticles for antisense oligonucleotide delivery. *Curr Pharm Biotechnol*, 15, 847-855.
30. Han, G., Gu, B., Cao, L., Gao, X., Wang, Q., Seow, Y., Zhang, N., Wood, M.J. and Yin, H. (**2016**) Hexose enhances oligonucleotide delivery and exon skipping in dystrophin-deficient mdx mice. *Nat Commun*, 7, 10981.

Chapter 7

Appendix

Summary

Confirmation of diagnosis for monogenic disorders with known etiology through genetic screening has been standard practice for more than a decade. Techniques involved are based on sequence analysis of genomic DNA, directed towards exonic and short flanking intronic regions within the gene of interest. Identified variants are considered to be pathogenic if they have been implicated in disease before, or by the use of prediction software. Algorithms have been generated that can predict the severity of missense mutations and the impact of the mutation on splicing. These have aided in the determination of causative pathogenic variants in numerous monogenic disorders. However, predictions do not necessarily reflect the in vivo situation. Functional assays aid in the confirmation of these predictions. Many functional assays are available at the protein level, i.e. enzymatic and cDNA expression assays, but at the RNA level, standardized assays are not yet implemented. **Chapter 2 and 3** in this thesis present a method for systematic unbiased screening for aberrant splice events and mRNA expression levels for single genes, using standard PCR analysis of all exons. With this method, both in frame and out of frame mRNA species can be detected due to blockage of the nonsense-mediated mRNA decay pathway. Application of this method to primary fibroblasts from one healthy control and twelve patients with Pompe disease, a recessive monogenic disorder in which the *GAA* gene is affected, resulted in the identification of more than twenty previously unidentified aberrant splicing events. These events were caused by either known or previously unknown variants in the *GAA* gene, including missense variants that were not implicated in splicing before. Furthermore, multiple aberrant splicing events were identified that occur in all individuals, which could point to a general mechanism in which splicing regulates overall *GAA* expression. Interestingly, nearly all aberrant splicing events identified in this thesis were caused by changes in splice site strength of canonical and cryptic splice sites due to a pathogenic variant, leading to cryptic splice site utilization. We hypothesized that antisense oligonucleotide (AON) mediated modulation of these aberrant splicing events could rescue canonical *GAA* pre-mRNA expression. The use of AONs for the modulation of splicing has been under investigation in the last two decades. This has resulted in the generation of potential treatments for diseases like spinal muscular atrophy and Duchenne muscular dystrophy that are currently under clinical investigation. In **Chapter 4**, we tested this hypothesis by blockage of cryptic splice sites using AONs in three primary fibroblast lines from Pompe patients in which these sites were identified using the splicing assay. The cryptic splice site utilized in one of these patients was generated by a deep intronic variant that could not be identified with conventional diagnostic gDNA Sanger sequencing. The cryptic splice site utilized in a second patient was a naturally occurring intronic cryptic splice site that

was utilized due to weakening of a canonical splice site, and the third patient utilized a exonic cryptic splice site generated by a missense variant. Blockage of these cryptic splice sites resulted in correction of canonical splicing in all three patients. This led to a concomitant increase in GAA enzymatic activity. These findings suggest that the splicing assay can be used to screen for aberrant splice sites that can be targeted for AON mediated splicing correction. Utilization of this new pipeline for identification and correction of aberrant splicing could lead to a novel treatment strategy for not only for Pompe disease, but for other monogenic disorders as well.

In chapter four, proof of principle was demonstrated for the use of AONs to rescue aberrant splicing in Pompe disease. However, the investigated variants only occur sporadically in Pompe patients. The most common variant in Pompe disease is the c.-32-13T>G (IVS1) splicing variant, which leading to partial or complete skipping of *GAA* exon 2. In **Chapter 5**, we set out to correct aberrant splicing caused by the IVS1 splicing variant with the use of AONs. A potential target site was identified using a *U7* antisense screening assay. Further investigation of the target site revealed the presence of a pseudo-exon in *GAA* intron 1 that was recognized in the context of the IVS1 variant as a result of weakening of the canonical *GAA* exon 2 splice donor site. To test whether blockage of splice acceptor and/or donor site of this pseudo-exon led to increased *GAA* exon 2 inclusion, we designed two AONs targeting these sites. Transfection of these AONs in skeletal muscle myotubes generated from Pompe patient derived induced pluripotent stem cells resulted in increased *GAA* expression and GAA enzymatic activity up to healthy control levels from the IVS1 allele. With further in vivo testing these findings could potentially lead to the generation of a new therapeutic treatment option for juvenile/adult onset Pompe patients carrying the IVS1 variant.

In conclusion, the findings described in this thesis underlie the importance of functional investigation of variants for their effect on pre-mRNA splicing, even if these variants are not predicted to effect splicing. Furthermore, the assay can be used to identify variants that could not be identified using conventional genetic diagnostics. Results obtained with the help of the splicing assay can be used for development of new treatment strategies based on RNA targeting. We identified eight unique antisense oligonucleotides that together could potentially be used in the treatment of more than half of all Pompe patients. Importantly, the concept of the splicing assay can be easily used for investigation of splicing in other monogenic disorders, leading to a better understanding of the mechanism of splicing and for the identification of potential antisense drug targets in other diseases.

Samenvatting

Bevestiging van diagnose door middel van genetische tests voor monogenetische aandoeningen met een bekende oorzaak wordt al meer dan tien jaar stadaard uitgevoerd in diagnostische laboratoria. De technieken die daarbij worden gebruikt zijn gebaseerd op het analyseren van genomisch DNA, waarbij in het gen van interesse wordt gefocust op exonische regio's en korte intronische regio's die daaraan grenzen. Geïdentificeerde varianten worden als pathogeen beschouwd als ze eerder zijn in verband zijn gebracht met de ziekte of als ze een potentieel pathogene indicatie krijgen wanneer ze in silico worden geanalyseerd. Verschillende algoritmes zijn beschikbaar die de ernst van missense mutaties en de gevolgen van mutaties op het splicing proces helpen te voorspellen. Deze algoritmes helpen bij de determinatie van pathogene varianten in veel monogenetische aandoeningen. Echter, in silico voorspellingen komen niet altijd overeen met de in vivo situatie. Functionele assays kunnen worden gebruikt om de in silico voorspellingen te testen. Tal van functionele analysemethoden zijn beschikbaar om mutaties op eiwitniveau te testen, zoals enzymatische en cDNA expressie assays, maar op het RNA niveau worden functionele assays nog niet vaak gebruikt. **Hoofstuk 2 en 3** in deze thesis beschrijven een methode voor het systematisch en onbevooroordeeld screenen op afwijkende splicing events en mRNA expressie van een specifiek gen, waarbij gebruik wordt gemaakt van standaard PCR analyse van alle exonen. Met deze methode kunnen zowel in frame als out of frame mRNA vormen worden gedetecteerd door middel van het blokkeren van nonsense-mediated decay pathway. Toepassing van deze methode op primaire fibroblasten van één gezonde controle en twaalf patiënten met de ziekte van Pompe, een recessieve monogenetische aandoening waarbij het *GAA* gen is aangedaan, heeft geresulteerd in de identificatie van meer dan twintig nieuwe aberrante splicevormen. Deze splicevormen zijn ontstaan door toedoen van bekende of voorheen onbekende varianten in het *GAA* gen, inclusief missense varianten die niet eerder in verband zijn gebracht met splicing. Bovendien konden er aberrante splicevormen worden aangetoond die voorkomen in elk individu, wat kan wijzen op een algemener mechanisme waarin splicing wordt gebruikt voor de regulatie van de expressie van het *GAA* gen. Een interessant detail is dat bijna alle aberrante splicevormen die zijn geïdentificeerd in deze thesis zijn ontstaan als gevolg van het veranderen van de sterkte van normale of cryptische splice site door een pathogene variant, wat leidt tot het gebruik van cryptische splice sites. Wij stelde de hypothese dat deze aberrante splice vormen kunnen worden hersteld naar de normale splicevorm met behulp van antisense oligonucleotiden (AON). Het gebruik van AONs voor het manipuleren van splicing wordt al geruime tijd onderzocht. Dit heeft geresulteerd in potentiële medicatie voor ziektes als spinale musculaire atrofie en Duchenne musculaire dystrofie, welke beiden momenteel in klinische trials worden getest. In **Hoofstuk 4** testen we de hypothese door het blokkeren

van cryptische splice sites met behulp van AONs in drie primaire fibroblastlijnen afkomstig van Pompe patienten. De cryptische splice site werden eerder geïdentificeerd met behulp van de splicing assay. De splice site die werd geïdentificeerd in één van deze patienten werd gegenereerd door een diep intronische mutatie. De gebruikte cryptische splice site in een andere patient was op zichzelf niet veranderd, maar werd gebruikt omdat een natuurlijke splice site in de buurt was verzwakt door een mutatie. Een derde patient had een exonische cryptische splice site die was gegenereerd door een missense mutatie. Het blokkeren van deze sites resulteerde in correctie van aberrante splicing in alle drie de patiënten. Dit leidde in alle gevallen in een hogere enzymatische activiteit van GAA. Deze bevindingen suggereren dat de splicing assay gebruikt kan worden om te screenen voor cryptische splice sites die met behulp van een AON kunnen worden geblokkeerd en gewone splicing kunnen herstellen. Het gebruik van deze nieuwe pipeline voor identificatie en correctie van aberrante splicing kan tot een nieuwe behandelmethode leiden voor de ziekte van Pompe, maar kan ook voor andere aandoeningen van belang zijn.

In hoofdstuk vier werd de mogelijkheid van het concept voor het gebruik van AONs voor het hertstellen van aberrante splicing voor de ziekte van Pompe aangetoond. Echter, de varianten die zijn onderzocht komen zelden voor in Pompe patiënten. De meest voorkomende variant in de ziekte van Pompe is de c.-32-13T>G (IVS1) splicing variant. Deze variant zorgt ervoor dat *GAA* exon 2 grotendeels of geheel wordt overgeslagen tijdens splicing. In **Hoofstuk 5** kijken we naar de mogelijkheid om AONs te gebruiken voor de correctie van aberrante splicing als gevolg van de IVS1 variant. Een potentiele target site werd geïdentificeerd met behulp van een U7 antisense screening methode. Dieper onderzoek naar deze site leidde tot de identificatie van een pseudo-exon in *GAA* intron 1. Dit pseudo-exon wordt herkent in de context van de IVS1 variant doordat deze variant de normale *GAA* exon 2 splice site verzwakt. Twee AONs werden ontworpen om te testen of het blokkeren van de splice acceptor en/of splice donor site van het pseudo-exon resulteert in meer *GAA* exon 2 inclusie. Behandeling van myotubes, verkregen vanuit differentiatie van geïnduceerde pluripotente stamcellen gegenereerd uit primaire fibroblasten van Pompe patiënten met de IVS1 variant, leidde tot een verhoogde *GAA* expressie en GAA enzymatische activiteit welke richting het niveau van een gezonde controle op het IVS1 allel. Verdere in vivo tests kunnen leiden tot de generatie van een nieuw type medicijn voor de behandeling van juveniele en adulte patiënten die de IVS1 variant dragen.

De bevindingen beschreven in deze thesis onderstrepen het beland van functioneel onderzoek om het effect wat varianten hebben op pre-mRNA splicing te bepalen, zelfs als deze varianten geen voorspeld effect hebben op splicing. Verder kan deze methode worden gebruikt voor de identificatie van varianten die met conventionele genetische diagnostiek niet worden gevonden. Resultaten die zijn gegenereerd met

behulp van de splicing assay kunnen worden gebruikt voor de ontwikkeling van nieuwe behandelstrategieën die zijn gebaseerd op behandeling met RNA als doelwit. Wij hebben acht unieke AONs geïdentificeerd die mogelijk kunnen worden gebruikt in de behandeling van meer dan de helft van alle Pompe patienten. Een belangrijk punt is dat het concept van de splicing assay met kleine aanpassingen kan worden gebruikt voor het onderzoeken van splicing in andere monogenetische aandoeningen, wat kan leiden tot een beter begrip van het mechanisme van splicing en voor identificatie van potentiële targets voor antisense medicatie van andere aandoeningen.

List of Publications

Bergsma A.J., in 't Groen S.L.M., Yamanaka M., Takahashi S., OkumiyaT., van der Ploeg A.T., Pijnappel W.W.M., (2016) 'Inhibition of mRNA Decay for Unbiased Detection of Aberrant Splicing in Pompe Disease', in preparation.

van der Wal E.*, **Bergsma A.J.*,** Pijnenburg J.M., van Gestel T.J.M., Zaehres H., Araúzo-Bravo M.J., Schöler H.R., van der Ploeg A.T., Pijnappel W.W.M., (2016) '*GAA* splicing in Pompe disease is restored by exon inclusion in expanded iPS cell-derived myotubes', submitted.

Bergsma A.J., in 't Groen S.L.M., Verheijen F.W., van der Ploeg A.T., Pijnappel W.W.M., (2016) 'From cryptic towards canonical pre-mRNA splicing in Pompe disease: a pipeline for the development of antisense oligonucleotides', Molecular Therapy Nucleic Acids, in press.

Wijchers P.J., Geeven G., Eyres M., **Bergsma A.J.,** Janssen M., Verstegen M., Zhu Y., Schell Y., Vermeulen C., de Wit E., de Laat W., 'Characterization and dynamics of pericentromere-associated domains in mice', Genome Research. 2015 Jul;25(7):958-69. doi: 10.1101/gr.186643.114.

Bergsma A.J., Kroos M., Hoogeveen-Westerveld M., Halley D., van der Ploeg A.T., Pijnappel W.W.M., 'Identification and characterization of aberrant GAA pre-mRNA splicing in Pompe disease using a generic approach', Hum Mutation. 2015 Jan;36(1):57-68. doi: 10.1002/humu.22705.

Pijnappel W.W.M, Esch D., Baltissen M.P., Wu G., Mischerikow N., **Bergsma A.J.**, van der Wal E., Han D.W., Bruch H.v, Moritz S., Lijnzaad P., Altelaar A.F., Sameith K., Zaehres H., Heck A.J., Holstege F.C., Schöler H.R., Timmers H.T., 'A central role for TFIID in the pluripotent transcription circuitry', Nature. 2013 Mar 28;495(7442):516-9. doi: 10.1038/nature11970.

Curriculum Vitae

Name: Atze Jacobus Bergsma
Address: Livingstonelaan 808
Zip code: 3526 JK
City: Utrecht
Tel.: +31644013113
E-mail: a.bergsma@erasmusmc.nl
D.O.B.: 27/06/1984

PROFESSIONAL EXPERIENCE

Erasmus MC, Rotterdam, The Netherlands

PhD Student in the group of Dr. W.W.M. Pim Pijnappel, Dept. of Pediatrics / Clinical Genetics

July 2012 – September 2016

We focus on the genetics behind Pompe disease, a rare lysosomal storage disorder which mainly effects the muscles. We are particularly interested in genetic variants that change canonical splicing, and are conducting translational research to find new treatment strategies for this disease.

EDUCATION

Utrecht University, Utrecht, The Netherlands

Master in Molecular and Cellular Life Sciences (2012)

Thesis: “Generation of cell lines with randomy targeted LacO-repeats to direct genic loci to chromocenters”

Hogeschool Utrecht, Utrecht, The Netherlands

Bachelor of Applied Sciences, molecular biology (2010)

Thesis: “Identification of specific protein interactions in chemotaxis signal transduction of Campylobacter jejuni”

Regionaal Opleidings Centrum (ROC) Utrecht, Utrecht, The Netherlands

intermediate vocational education, Chemistry (2005)

INTERNSHIPS

Hubrecht Institute, Utrecht, The Netherlands

Group of Prof. Dr. Wouter de Laat, supervision of Dr. Patrick Wijchers

October 2011 – June 2012

Organisation of chromatin in the nucleus is an important factor in the regulation of gene

expression. The nucleus of the cell contains specific chromatic sub-compartment each containing their own respective marks. Pericentric heterochromatin (PCH) is one of these compartments, and is generally regarded as a silencing compartment. During my intership we generated cell lines in which we could study the effect of recruitment of genes towards PCH had on transcriptional activity. Further research led to the paper described below (Wijchers et al., 2015).

UMC Utrecht, Utrecht, The Netherlands
Group of Prof. Dr. Marc Timmers, supervision of Dr. W.W.M. Pim Pijnappel
January 2011 – June 2012
Reprogramming of human fibroblast towards pluripotency has paved the way for disease modeling using patient own cells. However, the exact mechanism of this reprogramming is not yet completely understood. We looked at the role of transcription factor II D in maintenance of pluripotency in mouse ES cells. Further research led to the paper described below (Pijnappel et al., 2013).

Griffith University Institute for Glycomics, Gold Coast, Australia
At the Institute for Glycomics, the group of Dr. Victoria Korolik is interested in the underlying pathways of chemotaxis in Campylobacter jejuni. Using a yeast-two hybrid technique, we searched for specific protein interaction of a library of proteins with TLP1, a chemosensory transmembrane protein.

Hindustan Unilever Research Center, Bangalore, India
Group of Dr. Gautam Banerjee, Supervision of Jyoti Bhat
February 2009 – October 2009
Toll-like receptors (TLRs) play an important role in innate immunity. During my internship at Hindustan Unilever we investigated wether a particular herb used in preperation of tea could alter the response to ligand binding of different combinations of TLRs.

Certificate

Article 9 (Dutch laboratory animal law), Animal handling in IVC facilities, Statistics, Working with radiation

PhD Portfolio

	Year	ECTS
Courses:		
Biochemistry and Biophysics	2012	3
Laboratory animal science	2013	3
Cell and Developmental Biology course	2013	3
Genetics course	2013	3
Research integrity	2014	0,5
Literature Course	2014	2
Statistics	2015	2
Safely working in the laboratory	2015	0,5
Biomedical English Writing and Communication	2015	2
Workshops:		
20th MGC PhD Workshop, Luxembourg	2013	1
21th MGC PhD Workshop, Münster	2014	1
22th MGC PhD Workshop, Maastricht	2015	1
23th MGC PhD Workshop, Dortmund	2016	1
(Inter)National Meetings:		
Netherlands Institute for Regenerative Medicine Annual Meeting 2012, Amsterdam	2012	1
Sophia Research Day, Rotterdam	2013	0,5
Sophia Research Day, Rotterdam	2014	0,5
Steps Forward in Pompe Disease meeting , Turin IT	2014	1
Erfelijke stofwisselingsziekten Nederland (ESN) meeting, Rotterdam	2015	0,5
Dutch Society for Stem Cell research (DSSCR) annual meeting, Utrecht	2015	0,5
Sophia Research Day, Rotterdam	2015	0,5
American Society of Human Genetics (ASHG), Annual Meeting, Baltimore USA	2015	2
	Year	ECTS
Presentations:		
Poster presentation Sophia Research day	2013	0,25
Poster presentation MGC PhD Workshop	2014	0,25
Oral presentation Sophia Research day	2014	0,5
Poster presentation Steps Forward in Pompe disease meeting	2014	0,25

Oral presentation ESN meeting	2015	0,5
Poster presentation (ASHG), Annual Meeting	2015	0,25
Oral presentation MGC PhD Workshop	2015	0,5
Oral presentation MGC PhD Workshop	2016	0,5
Teaching Activities:		
"Rolemodel" for students CSG Calvijn College	2013	0,25
Supervisor of master student, 1/9/13 - 31/6/14	2013/2014	2
Presenter at International Pompe day	2015	1
Supervisor of master student, 1/9/14 - 31/8/15	2014/2015	2
Other		
Organizing committee 22th MGC PhD Workshop, Maastricht	2014/2015	2
Biomedical Sciences party committee member	2013/2014	1
Total ECTS		**40,75**

Dankwoord

Een jaar of 16 geleden liep ik over de studiebeurs in Utrecht te dwalen en had ik nog geen idee welke richting ik op zou gaan. Aan het einde van de dag liep ik nog langs een kraampje waar ik een boekje zag liggen waarop iemand in een laboratoriumjas iets stond te pipetteren. Ik dacht, is dat niet iets voor mij? Door de jaren heen ben ik het leren nooit zat geworden en heb ik doorgestudeerd, nooit verwachtend dat ik ooit zou promoveren… Tot het nu dan toch eindelijk zo ver is! Dit moment had ik nooit kunnen bereiken zonder de hulp van tal van mensen, die ik hiervoor graag wil bedanken.

Beste Pim, “het kan toch niet waar zijn”, ik ga promoveren! Ontzettend bedankt dat je mij de kans hebt gegeven om in jou groep mijn promotieonderzoek uit te voeren. Al direct na mijn stage vroeg je of ik het onderzoek beschreven in deze thesis wilde uitvoeren in Rotterdam. Ik heb er geen moment spijt van gehad dat ik daar ja op heb gezegd. Ik wist ten slotte wel wat ik aan je had als begeleider. Ook tijdens mijn promotieonderzoek heb je me echt ontzettend veel geleerd en geholpen een betere wetenschapper te worden. Ik weet zeker dat het komende jaar gaat zorgen voor nog meer nieuwe inzichten in ons onderzoek.
Beste Ans, ook jij ontzettend bedankt voor het vertrouwen dat je mij hebt gegeven. Eén van de mooie dingen aan dit onderzoek is toch wel het contact tussen het lab en de kliniek, iets waar jij veel tijd instopt om dat voor elkaar te krijgen. We leren enorm van elkaar tijdens de meetings/journal clubs die we samen hebben.
Leden van de kleine commissie, prof. R.M.W. Hofstra, prof. S.M. van der Maarel en prof. E.H.H.M. Rings. Bedankt dat jullie de tijd hebben genomen om mijn proefschrift door te lezen en te beoordelen en voor de wil om plaats te nemen als opponenten bij mijn verdediging. Ook hartelijke dank aan prof. J. Gribnau, prof. R. Willemsen en dr. Arnold Reuser, ook jullie bedankt voor het versterken van mijn opponentengroep.

Graag wil ik ook mijn paranimfen bedanken. Erik, we kennen elkaar al een tijdje. Dezelfde HBO en allebei stage gelopen bij het UMC Utrecht. Het echte samenwerken begon hier in Rotterdam, en ik heb een hoop van je kunnen leren. Ik ben erg trots op het hoofdstuk dat we samen hebben geschreven, en ook bedankt voor het vele lachen wat we hebben gedaan in kamer 928. Ronald, Ik heb echt enorm veel respect voor hoe jij je PhD met succes hebt afgerond. Je bent altijd zo relaxed! Tot vlak voor je promotie misschien, maar dat is niets om je voor te schamen, die kriebels beginnen bij mij nu ook te komen ;-). Beide mannen, heel veel succes met jullie wetenschappelijke carrière!

Ook de mensen in de groep van Pim wil ik graag bedanken voor al hun hulp en voor de fijne werksfeer die er altijd heerst. Gerben, onze spierenman! Bedankt voor je hulp en voor de gezellige praatjes in de trein naar Utrecht. Merel, heel veel succes met je onderzoek,

en natuurlijk veel geluk met je mannetjes! Monica, Qiushi and Douglas, best of luck in obtaining your PhDs! Mike, ondanks de titel van nr. 1 ...., het is wel altijd gezellig in de kamer. Succes met het afmaken van je tuin (en PhD), dat ga ik in Utrecht ook doen. Tom, ik hoop dat die knie van je het binnen niet al te lange tijd weer toestaat om een balletje te trappen, succes met het herstel. Stijn, super bedankt voor alles wat je hebt gedaan voor mijn onderzoek! Zonder jou had ik er waarschijnlijk nog wat tijd aan vast moeten plakken. Maar let alsjeblieft een keer op je Nederlands ;-)! Joon, nadat ik je kaarttrucjes zag doen en piano zag spelen wist ik waarom je zo goed kan pipetteren. Succes met het verdere onderzoek allen!
Of course I would also like to thank my previous Master students Willemieke and Andrea for their support in my research. Also thanks to all other students that have worked in our lab. I wish all of you best of luck in your future careers.
Marian en Marianne, bedankt voor alle hulp met het leren van de technieken die in het Pompe-lab worden gebruikt, en voor het vergaren van alle databases en cellijnen. Ook Lies, Frans, Mark en Leontine bedankt voor jullie hulp!

Alle andere mensen op de research afdeling van de klinische genetica wil ik ook graag bedanken. Vincenzo, Annelies, Renate, Tjakko, Bedankt voor de fijne discussies en ik wens jullie groepen een hoop goede wetenschap en publicaties toe. De mensen van kamer 928, Katherine, Nynke, Laura en Rhiana, bedankt voor de leuke tijd die we hebben gehad en voor de zeer wetenschappelijke, maar ook zeer niet wetenschappelijke praatjes.
Thanks to Wim, Fenne, Roy, Shimriet, Danny, Rajendra, Judith, William, Guido, Michelle, Shami, Herma, Helen, Lies-anne, Rachel, Bianca, Maria, Erwin, Isa, Jeanette, Bep, and also the people from the group of Joost, for the lovely work atmosphere and nice discussions at the coffee machine at the ninth floor.
Ook iedereen bij het centrum voor lysosomale en metabole ziekten en de afdeling kindergeneeskunde bedankt voor de informatieve en leerzame werkbesprekingen en journal clubs!

Ik wil ook graag de mensen bedanken die belangrijk waren voordat ik begon aan mijn PhD. Patrick, bedankt voor je goede begeleiding en altijd grappige insteek tijdens mijn tweede masterstage en ook Wouter bedankt dat ik dat in jou lab mocht doen. Ook Marc bedankt voor de kans om mijn eerste masterstage bij je te lopen. Iedereen van het MBO en HBO bedankt voor jullie insteek in mijn “carrière”. Eén persoon in het bijzonder, Claire, bedankt voor alles wat je voor mij hebt gedaan. Die spreekwoordelijke schop onder m’n kont had ik echt nodig. En we hebben toch een geweldige ervaring gehad in het buitenland. Ook de rest van de familie Keller bedankt dat jullie er altijd voor me waren.

Ook de “etenopdinsdaggroep” wil ik via deze weg bedanken. TimB, TimF, Feike, Tom,

Paul, Sander, Maarten, Tinus, Geert, JanV, Nol en TimO, het is altijd weer een feest om op dinsdag aan te schuiven, en dat het eten dan voor je klaar staat ;-). Daarvoor moet je natuurlijk wel af en toe wat sarcastische grappen kunnen tolereren, maar met Linda en Hansi op de achtergrond is dat allemaal geen probleem. Wel gelachen hoor!
Ook de dames in de groep die dit jaar weer “het” kamp regelt, al die zeer nuttige vergaderingen die de laatste jaren hebben geleid tot een heerlijk rustgevend weekendje weg… Dina, Maaike, Melissa, Rianne, Sanne en Malou, we maken er weer een mooi weekend van dit jaar in Heino!

Harrie en Marita, Johan en Deborah, ook jullie bedankt voor alles, en nu is het mijn tijd om jullie te helpen met alle verhuizingen die er aan zitten te komen!

Pa en ma, jullie verdienen de grootste credits! Zo’n 12 jaar geleden zag het er totaal niet naar uit dat ik nu hier zou zijn. En zonder jullie onvoorwaardelijke steun was dat waarschijnlijk ook niet het geval geweest. Ik wil jullie daar echt ontzettend en uit de grond van m’n hart voor bedanken! En pa, natuurlijk ook bedankt voor de hulp met de layout van dit proefschrift. Natuurlijk ook Okke, Sytske, Tessa en Lisa, bedankt voor jullie bijdrage en steun, ik hou van jullie allemaal!

En dan, lieve schat, wat ben ik blij dat ik jou heb mogen ontmoeten (met dank aan Kirsten en Bastijn!). We hebben in de laatste 3 jaar al zoveel samen gedaan, en dat heeft geen moment verveeld! Ik weet dat ik de afgelopen maanden wat minder gezellig ben geweest, maar je staat zelfs dan nog altijd voor me klaar. Ontzettend bedankt voor alles wat je voor me hebt gedaan, en alvast bedankt voor wat er nog komen gaat. We gaan samen nog een mooie tijd tegemoet, en we gaan rustig aan die bucket lists afwerken. Dank je lieverd, hou ontzettend veel van je!!!