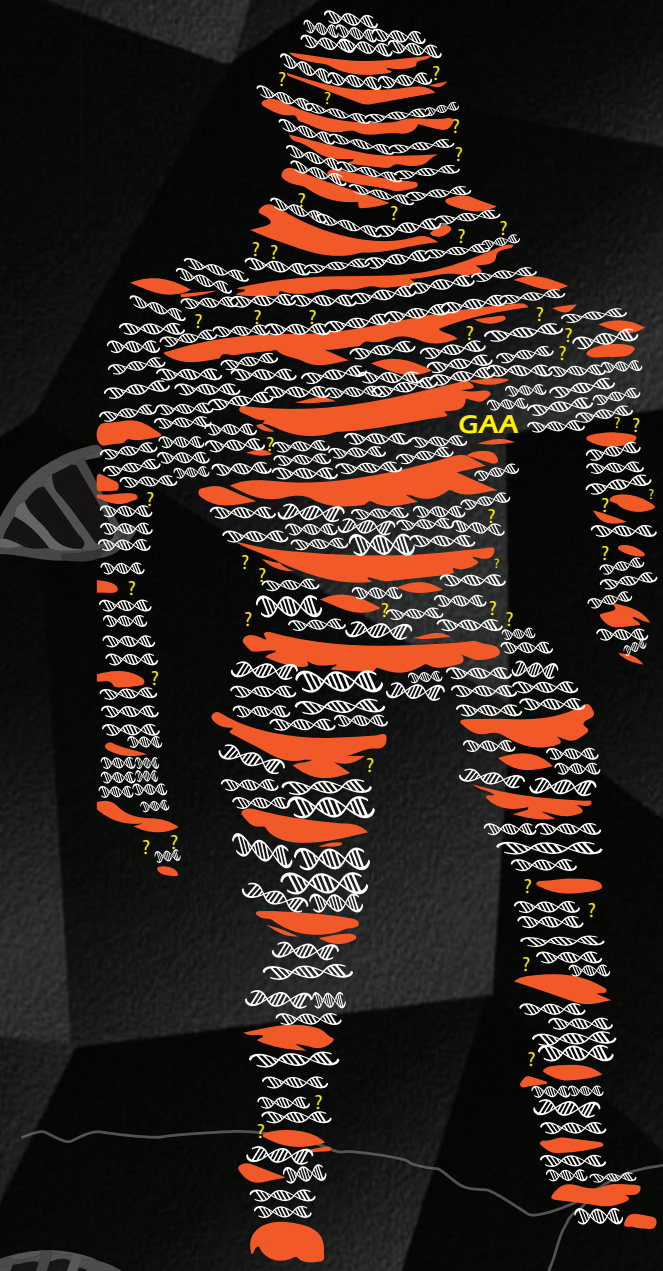
GENETICS OF POMPE DISEASE

The secrets within the coding region of the *GAA* gene and beyond



Douglas O. S. de Faria

Genetics of Pompe Disease

The secrets within the coding region of the *GAA* gene and beyond

Douglas Oliveira Soares de Faria

The research presented in this thesis was financially supported by grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil; grant number: 234407/2014-0) awarded to Douglas Oliveira Soares de Faria and Erasmus Medical Center Rotterdam (EMC).

The studies described in this thesis were performed in the Center for Lysosomal and Metabolic Diseases, Department of Pediatrics and Department of Clinical Genetics of the Erasmus University Medical Center, Rotterdam, The Netherlands.

ISBN: 978-94-6416-835-8

Author: Douglas Oliveira Soares de Faria

Cover: Douglas O. S. de Faria and Mike Broeders

Layout: Douglas O. S. de Faria and Mike Broeders

Printed by: Ridderprint

Genetics of Pompe Disease

The secrets within the coding region of the *GAA* gene and beyond

Genetica van de ziekte van Pompe

De geheimen binnen en buiten het coderende gebied van het *GAA* gen

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. A.L. Bredenoord

and in accordance with the decision of the Doctoral Board.

The public defense will be held on

Tuesday 2nd November 2021 at 10.30 hours

by

Douglas Oliveira Soares de Faria

born in Uberaba, Brazil

Erasmus University Rotterdam

Doctoral Committee

Promotor:	Prof. dr. A.T. van der Ploeg
Copromotor:	Dr. W.W.M. P. Pijnappel
Other members:	Prof. dr. R.M.H. Wijnen Prof. dr. P.A. van Doorn Prof. dr. C.D.M. van Karnebeek

Table of contents

List of Abbreviations

Abbreviation	Name
4MUG	4-methylumbelliferyl-α-D-glucopyranoside
ACE	angiotensin converting enzyme gene
ACTN3	alpha-actin-3 gene
ALT	alanine aminotrasnferase
AST	aspartate aminotrasnferase
BGAL	β-galactosidase
CADD	Combined Annotation-Dependent Depletion
CCDC40	Coiled-Coil Domain Containing 40 gene
cDNA	complementary DNA
CEU	Utah Residents with Northern and Western European Ancestry
CK	creatine kinase
CMV	cytomegalovirus
CNV	copy number variations
CRIM	cross reactive immunological material
DBS	dried blood spot
DEGs	differentially expressed genes
DMF	digital microfluids
EPOC	European Pompe Consortium
eQTL	expression quantitative trait loci
ER	endoplasmic reticulum
ERT	enzyme replacement therapy
FDA	Food and Drug Administration
FDR	false discovery rate
FISH	Fluorescence in situ hybridization
FPKM	fragments per kilobase of transcript per million
GAA	acid α-glucosidase
GAA	acid α-glucosidase gene
Glc4	tetrasaccharide 6-α-D-glucopyranosyl-maltotriose
GnoamAD	Genome Aggregation Database
GO	gene ontology
GoNL	genome of the Netherlands
GSD	glycogen storage disease
GTEx	Genotype-Tissue Expression project
HGMD	Human Gene Mutation database
HGVS	Human Genome Variation Society

HPLC	High Performance Liquid Chromatography
IBS	identical by state
IGF-1	Insulin-like growth factor 1 gene
IPA	Ingenuity Pathway Analysis
IVS1	c.-32-13T>G
LD	linkage disequilibrium
LEPR	Leptin Receptor gene
LOVD	Leiden Open Variation Database
LVH	left ventricular hypertrophy
M6P	mannose 6-phosphate
MAF	minor allele frequencies
MAOA	Monoamine oxidase A gene
MDS	Multidimensional Scaling
MLPA	Multiplex ligation-dependent probe amplification
MRI	Magnetic resonance imaging
mRNA	messenger RNA
MS/MS	Tandem Mass Spectrometry
MYBPC1	Myosin Binding Protein C gene
NBS	newborn screening
NGS	Next Generation Sequencing
NRK-1	Nicotinamide riboside kinase gene
PANTHER	Protein Analysis Through Evolutionary Relationships database
PAS	periodic acid Schiff
QC	quality control
qRT-PCR	Quantitative real-time PCR
QTL	quantitative trait loci
rhGAA	recombinant human GAA
RNA-seq	RNA sequencing
RoH	Regions of homozygosity
SDM	site-directed mutagenesis
SNP	single nucleotide polymorphisms
SNV	single nucleotide variant
STC-1	Stanniocalcin-1 gene
TCF7L2	Transcription Factor 7 Like 2 gene
TGLC	tetrasaccharide Glc4
TSS	transcription start site
UPD	uniparental disomy
VUS	variant of unknown significance

CHAPTER 1

General Introduction

GENERAL INTRODUCTION

The lysosomal enzyme acid α-glucosidase (*GAA*), also called acid maltase, is responsible for the breakdown of the intra-lysosomal glycogen into glucose, which represents approximately 10-15% of the cellular glycogen content [1]–[3]. Pompe disease (OMIM 232300), also known as glycogen storage disease type II, is an inborn error of metabolism caused by the deficiency of *GAA* and is transmitted in an autosomal recessive manner. The *GAA* gene is considered to be a housekeeping gene since it is expressed in every cell type [4]. However, skeletal and cardiac muscles are mainly affected. Disease-associated variants in the *GAA* can lead to dysfunctional or absent GAA production causing excessive glycogen accumulation within lysosomes resulting in structural damage of cells and tissues [5]–[8].

CLINICAL FORMS

The primary Pompe disease classification was based on the age at start of symptoms, and/or the presence of cardiomyopathy. The classic infantile form is characterized by age at symptom onset before 1 year old and the presence of hypertrophic cardiomyopathy (cardiomegaly). The non-classic forms may start at any age [9]. In most of these cases, the heart is not involved. Based on the age of symptom onset, further subdivisions have been made in childhood and adult Pompe disease, but in fact the disease presents as a continuous spectrum [7], [10], [11]. In general, early onset of symptoms precludes a more severe form of disease. Due to the deficiency of GAA in Pompe disease, the initial idea was to treat the disease by administering functional GAA intravenously, known as enzyme replacement therapy (ERT) [12]. The clinical trials of ERT with recombinant human GAA (rhGAA) in classic infantile Pompe patients showed improvement on survival for these patients [13], [14]. In 2006, the first treatment of Pompe disease with ERT was approved by the Food and Drug Administration (FDA). ERT results in substantial clinical improvement for classic infantile patients, eliminating the cardiac symptoms and improving the motor function [13], [15]. The features of classic infantile, childhood and adult Pompe disease are described below.

Classic infantile Pompe disease

Patients with the classic infantile form, the most severe type of the disease develop symptoms within the first months of life. They completely lack GAA enzyme activity. Classic infantile patients can be further subdivided into two groups: patients that do not produce any GAA protein so called cross- reactive immunological material (CRIM) negative patients

and patients who produce catalytically inactive GAA so called CRIM positive patients [16], [17]. CRIM negative patients have a higher likelihood of forming anti-rhGAA antibodies when treated with ERT since their immune system does not recognize the GAA protein. The primary phenotypic features of classic infantile patients are progressive general muscle weakness (floppy baby appearance) and hypertrophic cardiomyopathy. When patients do not receive ERT, they will die within the first year of life due to cardiorespiratory failure [18]. Additional common symptoms of untreated infants are feeding problems, failure to thrive, and respiratory problems including airway infections and respiratory insufficiency [19]. Early postmortem studies have shown that glycogen storage in infants is not limited to muscle. The glycogen stores in other organs and tissues as well, such as in the nervous system [20], [21].

Childhood and Adult Pompe disease

Patients with the childhood form of the disease experience the first symptoms from early infancy (some patients within the first 6 months of life) up to 16 years of age, but mostly without developing progressive hypertrophic cardiomyopathy. They show some residual GAA enzymatic activity, and the disease progression is slower than in classic infantile Pompe disease. Childhood patients present with a slowly progressive proximal skeletal muscle weakness (limb-girdle weakness). This leads to mobility problems and children often show delayed motor development [22], [23]. Other features related to muscle weakness are scapular winging, scoliosis and increased lumbar lordosis [24]. The respiratory muscles are also affected, which may cause respiratory complications and may also contribute to the fatigue that most patients experience. Without treatment, the life expectancy of these children is variable. They can survive from several years to several decades [25]. In a long-term follow-up study of 17 childhood patients, ERT had a positive effect by increasing muscle strength and stabilizing muscle function. However, the effects on lung function were more variable [26].

Patients with adult onset Pompe disease usually present with slowly progressive proximal myopathy and/or reduced pulmonary function [22], [27]. Cardiac involvement is rarely seen [28]. As in childhood Pompe disease, the respiratory muscles, in particular the diaphragm, are involved. This results in reduced pulmonary function and hypoventilation especially when patients are in supine position. Most patients ultimately require mechanical ventilation [29]. Respiratory insufficiency is the leading cause of death [30]. Adult patients can benefit from ERT. In a placebo controlled multicenter trial, ERT was shown to improve walking distance (assessed by the six-minute walk test) and to stabilize lung function. Further studies have shown that ERT changes the natural course when it comes to loss of pulmonary function and muscle strength [31]–[34].

DISEASE PATHOLOGY

For understanding the pathology of Pompe disease it is easiest to start with explaining the function and processing of GAA. GAA is co-translationally imported in the endoplasmic reticulum (ER), folded and glycosylated, whereafter it is transported through the Golgi complex *en route* to the lysosomes. Phosphorylation of mannose residues provides the mannose 6-phosphate (M6P) recognition marker and directs transport to the lysosomes by binding to M6P-receptors in the trans Golgi network. Once GAA is inside the lysosome, it degrades lysosomal glycogen to glucose.

Lysosomes accumulate glycogen through autophagy. The autophagosome envelops cytoplasmic material, which contains glycogen, after which it fuses with the lysosome. Deficiency of GAA causes accumulation of glycogen in lysosomes since glycogen cannot cross the lysosomal membrane. Cellular pathology is mainly seen in muscle tissue but can be seen in other tissue as well [35], [36]. Glycogen accumulation increases the size and number of lysosomes. The increase of size causes susceptibility of the lysosomal membrane to rupture, releasing degradative enzymes and metabolites into the cytoplasm resulting in cellular damage [37]. Also, when the lysosomal function is compromised due to glycogen accumulation, autophagosomal dysfunction ensues as a consequence followed by total cellular dysfunction [38], [39]. These (secondary) cellular disturbances involve abnormal calcium homeostasis, mitochondrial abnormalities and accumulation of lipofuscin deposits, which has been associated with oxidative stress [40]-[43].

MOLECULAR ASPECTS OF POMPE DISEASE

The GAA gene

In 1979, the human acid α-glucosidase gene with gene symbol *GAA* was discovered to reside on chromosome 17 [44]. Ten 10 years later, the *GAA* locus was determined to be on the distal portion of the long arm around position q23 [45]. The exact location is 17q25.2-q25.3 [7]. The *GAA* gene was characterized in 1990 to consist of approximately 20kb containing 20 exons [46]. The first exon is noncoding, and has recently been found to consist of exon 1A and 1B, interspersed by a 195 bp long intron 1A [47]. The ATG translation start codon is located in exon 2 following 32 base pairs (bp) of untranslated sequence counting from the 5' end of exon 2. According to the conventional nomenclature, the base A of the ATG codon is c.1. *GAA* exon 2 contains the sequence for the signal peptide that directs the ribosome to the membrane of the ER [37].

The GAA *protein*

GAA mRNA is translated at the ER, emerging as a 110 kDa glycoprotein. Mannose

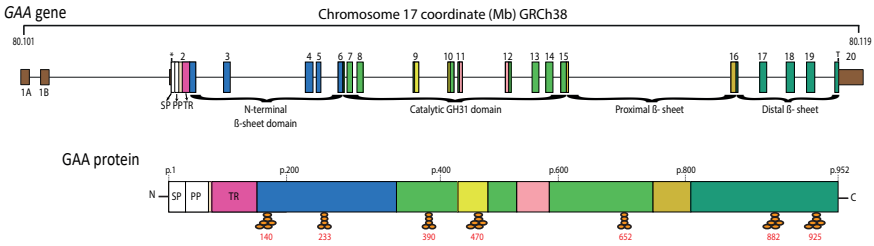


Figure 1. Schematic representation of the *GAA* gene and GAA protein. Upper cartoon: the *GAA* gene located in chromosome 17, encompassing 18 kb. There are 20 exons, of which 19 (2-20) are coding. Dark brown regions indicate the 5' and 3' untranslated regions. The exonic regions corresponding to the predicted protein domains are indicated as follows: the trefoil-P domain (pink), the N-terminal β-sheet domain (blue), the catalytic GH31 domain (green), catalytic site 1 (yellow), catalytic site 2 (salmon), the proximal (dark yellow) and distal (emerald green) β-sheet domains. (*, start codon; T, stop codon; SP, signal peptide; PP, propeptide; TR, trefoil). Lower cartoon: the GAA protein encompassing 952 amino acids with the representation of the protein domains and the 7 glycosylation sites, some carrying the mannose 6-phosphate recognition marker (indicate as brown structures) (Based on [8], [50], [83]).

6-phosphate groups are acquired, the GAA protein is subsequently transported to the Golgi and transverses the *cis*- and *trans*-Golgi cisternae. At the exit of the *trans*-Golgi, GAA is captured by the mannose 6-phosphate receptor, whereby it is transported to the lysosomal system [48]. During transport, the 110 kD precursor form of GAA undergoes N and C terminal proteolytic processing and is activated, resulting in the formation of the 95 kD intermediate form and mature forms of 76 kD and 70kD [49].

The mature GAA protein contains several domains, including a trefoil Type-P domain, an N-terminal β-sheet domain, the catalytic GH31 $(\beta/\alpha)_8$ domain, and proximal and distal β-sheet domains [50]. A schematic representation of the *GAA* sequence and its domains is shown in **Figure 1**.

ENZYMATIC AND MOLECULAR DIAGNOSTIC OF POMPE DISEASE

The diagnosis of Pompe disease can be easily missed due to the rarity of the disease and the unawareness of the condition by general practitioners. Furthermore, the phenotypic features can be similar as in other neuromuscular disorders, particularly in affected children and adults [51]. Among the first indicative symptoms is limb-girdle weakness. Blood analysis may reveal high creatine kinase (CK) serum levels in symptomatic and presymptomatic patients [52], [53]. Non-invasive tests, such as the analysis of the tetrasaccharide GLc4 in the urine, can also be performed [54]-[56]. However, Glc4 is not specific for Pompe

disease as elevated Glc4 also occurs in other glycogen storage diseases, in some cases of Duchenne muscular dystrophy and after disease-unrelated muscle trauma [56], [57]. Electromyography may reveal a myopathic pattern [7]. Muscle or skin biopsies can be very useful for the diagnosis of metabolic myopathies. Biopsy specimens are tested for histology, immunohistochemistry, electron microscopy, and biochemistry. Patients with Pompe disease may show a higher concentration of intramuscular glycogen detected by periodic acid Schiff (PAS) in the biopsy. However, the absence of vacuolization is seen in some cases of childhood or adult Pompe disease [58]–[60]. In addition, the presence of PAS positive material may vary from fiber to fiber, and glycogen can easily be washed out when the material is not properly fixated. Increased staining of acid phosphatase may point to a positive diagnosis. To confirm the diagnosis of Pompe disease, GAA enzyme activity in leukocytes or fibroblasts are commonly used, preferably combined with DNA analysis, as described below.

Enzymatic analysis

Lysosomal acid α-glucosidase (GAA) is present in all human tissues and is expressed e.g. in leukocytes and cultured skin fibroblasts, which sources can be used for enzymatic diagnosis [61]. Prenatal diagnosis in families at risk can be performed by *GAA* DNA analysis to identify known familial *GAA* variants combined with enzymatic assays using chorionic villi as a sample source [62], [63]. Since ERT became available in 2006 to treat Pompe disease, newborn screening (NBS) has become an option for neonatal patient detection. Biomarkers such as CK and Glc4 display low specificity and are mainly used for additional diagnostic purposes [64].

High throughput methods to quantify the GAA activity in dried blood spots (DBS) are based on Tandem Mass Spectrometry (MS/MS) and Fluorometric assays. MS/MS is a highly sensitive and specific technique. This technique is becoming more popular for the analysis of biomarkers for different types of lysosomal storage disorders [65]. In 2002, Chamoles and colleagues showed that many lysosomal enzymes remain active in dried blood spots [66]. Since this finding, DBS has been increasingly implemented as an NBS method for lysosomal storage disorders [67]. Nowadays, many NBS programs worldwide use DBS samples to perform enzymatic analysis using a glycogen-similar substrate that is quantified by MS/MS. The use of digital microfluidics (DMF) is also being performed. DMF is a multiplex advancement of the classical enzymatic fluorescent assay based on synthetic 4-methylumbelliferyl-α-D-glucopyranoside (4MUG) that processes microdroplets. Droplets are dispensed, moved, stored, mixed, reacted and analysed in an automated manner, taking less than 4 hours with minimal crosstalk between analyses [68], [69]. A study reported the use of MS/MS to distinguish patients with Pompe disease from healthy individuals with the known Asian pseudodeficiency allele c.[1726G>A;2065G>A] [70]. However, the use of this technique requires in all cases a second tier enzyme assay to decrease the false-positive

rate [71]. Therefore, the final diagnosis is made by fluorometric GAA assay using mixed leukocytes and/or cultured primary skin fibroblasts combined with molecular analysis to detect the disease-associated variants in the *GAA* that cause the deficiency. The enzymatic activity can be measured using the natural substrate glycogen or the artificial substrate 4MUG. The addition of acarbose to the reaction mixture is essential to inhibit the interfering glucoamylase activity when using leukocytes as an enzyme source. According to our recent analysis (Chapter 2), cultured skin fibroblasts provides the only source that enables us to distinguish between the different phenotypes of Pompe disease [72]-[74].

Molecular analysis and Pompe disease in a genomic era

As part of the diagnostic workup for Pompe disease, DNA analysis is usually performed to determine the causative *GAA* disease-associated variants after the diagnosis has been confirmed enzymatically. The diagnosis is confirmed when two disease-associated *GAA* variants are identified, one in each *GAA* allele. In some cases, the definitive diagnosis can be delayed when the pathogenicity of a variant is unknown, or when only one disease-associated variant is detected. DNA analysis is advantageous because it can discriminate between carriers and affected individuals. However, there are cases that require more extensive analysis, such as deletions that are too big to be detected by Sanger sequencing [75], [76] or when complex situations are encountered as described in Chapter 3.

The past decades have seen the advent of powerful and innovative methods for the sequencing of nucleic acids (DNA and RNA). These Next Generation Sequencing (NGS) techniques provide new ways to analyze the genetic bases of diseases. NGS methods range from gene panel sequencing, as is the case in MotorPlex sequencing, and whole exome and genome sequencing [77], [78]. These sequencing techniques are so powerful that they have led to the discussion of whether NGS can replace conventional biochemical tests for newborn screening [79]. One of the major driving forces of this debate is the decreasing price of NGS techniques compared to the prices of more traditional diagnostic methodology [80]. There are many possible applications of utilizing NGS techniques for the diagnosis of Pompe disease and other disorders, which are shown in **Table 1**.

However, the major challenge of NGS technology is the interpretation of the clinical significance of the DNA variants that are identified. The interpretation of a variant of unknown

Table 1. Applications for the use of NGS in diagnostics
Differential diagnosis
Mosaicism
Prenatal diagnosis
Newborn screening
Carrier screening
Identification of biomarkers or genetic modifiers
Screening when enzymatic test is unavailable

significance (VUS) is discussed in Chapter 4, in which benign variants can be misinterpreted as disease-associated. The disease-associated genotype alone is not always sufficient to predict the disease phenotype. A deeper understanding of the effects of disease-associated variants is needed to clarify the correlation between the patient's genotype and clinical phenotype in inherited diseases such as Pompe disease. The maintenance of databases that include both genotypic as well as phenotypic information of affected individuals is essential to obtain a better understanding of the link between genotype and phenotype and to investigate the role of modifying factors.

POMPE DISEASE *GAA* VARIANT DATABASE

Databases

Many disease-associated *GAA* variants have been reported in the literature, and various different sources provide an overview of the discovered variants in the *GAA* gene. The Human Gene Mutation Database (HGMD) at http://www.hgmd.cf.ac.uk/ac/ currently lists 658 variants distributed throughout the *GAA* gene. The HGMD collects variants present in published scientific articles using automated procedures and manual curation [81]. The Leiden Open Variation Database (LOVD) at www.lovd.nl/*GAA* reports 756 unique *GAA* sequence variants, of which 404 are classified as pathogenic. LOVD is a freely accessible gene-centered collection platform that allows for the inclusion of data at the individual level but can also accommodate data from NGS studies. The LOVD provides a public web interface in which registered users can submit sequence variants prior to publication. These submissions are subsequently reviewed by a curator to ensure the correct use of Human Genome Variation Society (HGVS) nomenclature [82].

Above mentioned databases provide a global overview of variants that have been described in the literature. However, these databases do not clearly define the impact of a specific variant on the disease severity in terms of clinical presentations. The Pompe disease *GAA* variant database at www.pompevariantdatabase.nl is a comprehensive disease-specific database that provides information on 911 variants that have been described in the literature. Furthermore, it provides the predicted clinical severity of the variant based on the phenotypes of patients carrying the variant and displays information regarding the effect of the sequence variants based on functional analysis and *in silico* predictions. Regular updates of the database are performed in which new variants are added and existing variants are curated in order to update variant severity scores. From the total of 911 variants, 336 variants could be linked to classic infantile, childhood or adult phenotypes when combined with a null allele [8], [83].

Disease-associated variants in the GAA gene

Disease-associated variants in the *GAA* gene can affect the GAA protein production and/or activity in several ways. In some instances, the type of disease-associated variant can be used to predict the phenotype [7], [8], [84]. The most common types of disease-associated variants that cause Pompe disease are missense variants [8]. These variants often affect the structure of GAA, thereby destabilizing the protein structure, which can subsequently lead to endoplasmic-reticulum-associated protein degradation [8], [50], [85]. Exonic deletions are the second most common type of variants identified to cause Pompe disease. An example is the c.525del variant. This variant, which is the most common deletion in Dutch patients, is a severe variant that generates a premature stop codon in *GAA* exon 2 [86]. Variants that affect pre-mRNA splicing are also common in Pompe disease; they amount to 13% of the disease-associated *GAA* variants reported to date. The most common variant in Pompe disease, which occurs in 90% of Caucasian patients, is c.-32-13T>G (commonly referred to as IVS1). It is an intronic splicing variant that will be discussed in more detail below [87].

During the standard diagnostic screening of monogenic diseases, exonic sequences and exon-intron boundaries are analyzed for detection of disease-associated variants since they are most common. However, variants located deep into intronic regions, promoter regions and UTRs have also been identified and can be missed [47]. Although these variants are not located in the coding region of the gene, they may still have an impact on pre-mRNA splicing. An example of such a deep intronic variant that causes aberrant splicing was identified in *GAA* intron 15. The c.2190-345A>G variant generates a new splice acceptor site, leading to the production of three additional products that contain a premature stop codon [6]. The HGMD database reports a similar percentage (12.7%) of splice variants as the Pompe variant database (13%). The pathogenic effect of these variants has been analyzed previously [88]. One common disease-associated variant, which occurs in 50% of children and 90% of adults with Pompe disease in the Caucasian population, is the IVS1 variant (c.-32-13T>G) [89]–[91]. This variant is located in the polypyrimidine tract of *GAA* intron 1B and causes partial or total skipping of exon 2, thereby removing the translation start site [89], [92]. Some residual normal splicing remains, which allows the production of 10-15% of canonically spliced *GAA* mRNA, leading to a late-onset phenotype [93]. Patients that carry the IVS1 variant have been reported to have a broad clinical spectrum, indicating that modifying factors play a role in the Pompe disease phenotype [4], [87], [94], [95].

Extended molecular analysis

Traditional Sanger sequencing for the detection of disease-associated variants might fail to detect larger deletions. For example, two reports have described gross deletions in the *GAA* that were caused by the recombination of two Alu elements. Sanger sequencing was unable to identify these variants, which were only found after additional molecular analysis [76], [96]. Furthermore, uniparental disomy (UPD), which stands for the inheritance of two

homologous chromosomes from the same parent, was diagnosed in a patient with Pompe disease after extended diagnosis [97]. In another case report, an infant was presented with symptoms suggestive of Pompe disease, but with an uncertain diagnosis. The patient had left ventricular hypertrophy, consistent with the classic infantile form of Pompe disease, but did not have muscle weakness. In this case, the routine molecular diagnosis did not either provide a clear answer. Additional studies implementing extensive molecular screening and the use of SNP array (searching for single nucleotide polymorphisms (SNP)) established a chromosomal imbalance on chromosome 17, discussed in Chapter 3. This chromosomal imbalance was present in a mosaic pattern [47].

Common variants in the GAA gene

Many SNPs with minor allele frequencies (MAF) above 1% have already been reported to reside within the *GAA* and can be ruled out to cause Pompe disease due to their high MAF in healthy individuals [94], [98], [99]. Nonetheless, it cannot be excluded with certainty that these variants affect disease progression when present *in cis* with disease-associated variants [83]. Several *in cis* disease-modifying variants in the *GAA* have been reported. These variants cause a decrease in GAA enzymatic activity but do not cause Pompe disease by themselves. They cause GAA pseudodeficiency. An example is the Caucasian c.271G>A (p.Asp91Asn) variant (also termed *GAA*2), which is located in *GAA* exon 2 [100], [101]. This variant has a MAF of 3% in the European population. Alleles with this variant encode the GAA2 alloenzyme, which has reduced affinity for the glycogen substrate [100], [102]. Another well described pseudodeficiency allele is the combination c.[1726G>A;2065G>A] with MAFs of 15% and 24%, respectively, in East Asia. Homozygotes have an estimated 60-80% reduction of acid α-glucosidase enzyme activity when tested *in vitro* [103]. However, considering the MAF and the effect on GAA activity, these variants are insufficiently deleterious to cause Pompe disease when no other disease-associated variants are present *in cis* [8], [83].

Is there an SNP signature for Pompe disease?

An allele can be described as a variant of a gene at a specific chromosomal location, termed a locus. A locus consisting of a combination of distinct alleles is called a haplotype. An important aim in the field of inherited diseases is to identify the relationship between alleles and/or haplotypes and their respective effect on the expression of genes that could result in a specific phenotype. Haplotypes are often defined by the presence of a specific combination of SNPs. Individual SNPs can affect the gene they reside in depending on their location (e.g. coding region, enhancer, silencer, promoter)[104], [105]. The combination of multiple SNPs can potentially influence certain traits [106], [107]. Often such traits are

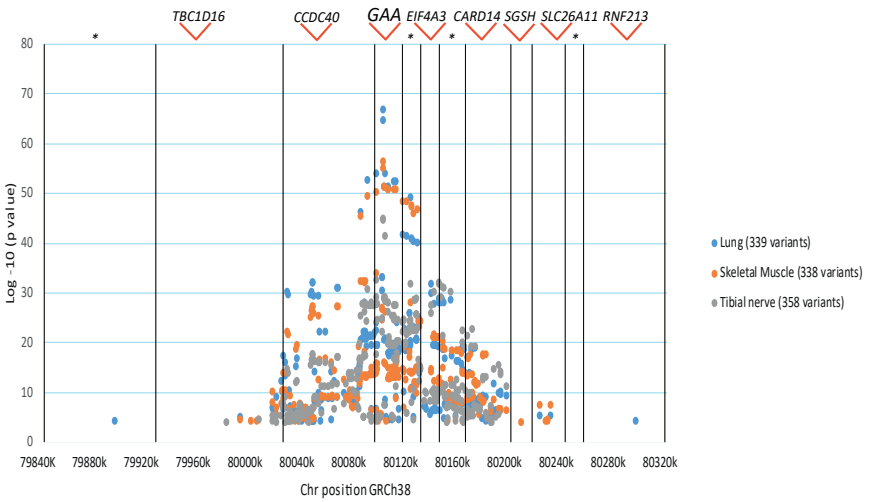


Figure 2. *cis*-eQTLs for *GAA*. Significant variants associated with GAA expression in lung, skeletal muscle and tibial nerve tissues. The strength of the eQTL is expressed as the -10 log of the p-value of the ratio between the expression of the alternative variant and the reference variant. The *cis*-eQTLs are present at both upstream and downstream locations relative to the *GAA* transcription start site.
(* intergenic region) (source: GTExportal.org)

specific to geographical regions and can differ by ethnicity since they are more likely to be inherited within such a population [108].

Many SNPs have been described in the *GAA* gene. However, little information about the potential impact of these variants on GAA protein production and/or activity is available [8], [99]. Understanding the effect of these SNPs can yield additional information for a specific phenotype. Some of these effects have already been established and are described as quantitative trait loci (QTLs). QTLs are sections of DNA that contain a specific combination of genetic variants that correlate with a specific trait. Sections of DNA that affect gene expression of certain genes are known as expression QTLs (eQTLs) [109]. eQTLs for the *GAA* can be found in the Genotype-Tissue Expression project (GTEx) (https://www.gtexportal.org/home/). eQTLs in different tissues from *post mortem* healthy donors shows variants *in cis* that influence the expression of GAA in the lung, skeletal muscle, and other tissues. Therefore, common variants residing in the *GAA* itself and in the nearby (<250 kilobases) surrounding area influence GAA expression (**Figure 2**). Moreover, variants that are located further away from the gene (>5 megabases) or on a different chromosome, could also affect gene expression. These variants are known as *trans*-eQTLs. However, to study the impact of *trans*-eQTLs, a considerable number of samples needs to be analysed, since the overall effects are usually small [109].

As mentioned above, patients who carry the IVS1 variant on one *GAA* allele next

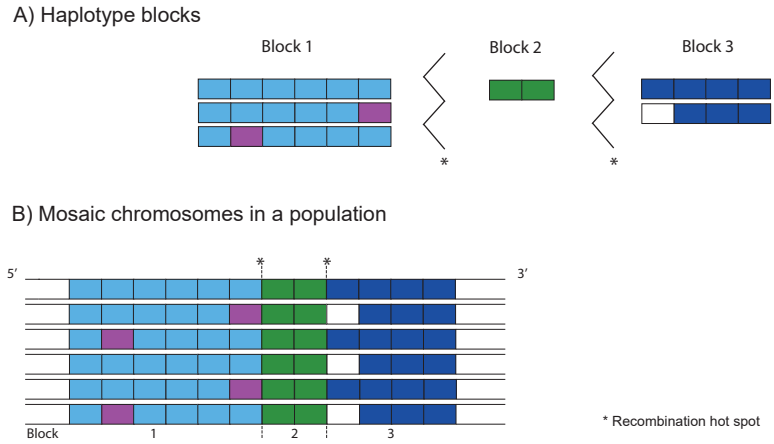


Figure 3. Example of haplotype blocks and their combinations. The different colours in a block show alternative alleles. A) Within each block, there is little or no evidence for recombination. Recombination hotspots are represented as zigzag lines. In a specific population, a limited number of distinct alleles within a given haplotype block can occur. B) Most chromosomes are composed of a mosaic arrangement of haplotype blocks. Each population will differ in the frequency of those specific arrangements (adjusted from [112]).

to a null variant on the other allele present a broad spectrum of phenotypes with respect to ages at symptom onset and severity of clinical features. This suggests that modifying factors play a role [4], [87], [94], [95]. A comparative study performed among affected children and adults both carrying the IVS1 variant has identified the synonymous c.510C>T variant as a disease modifying factor [87]. Splice assays and RT-qPCR analyses have shown that this variant negatively influences the correct *GAA* pre-mRNA splicing when present *in cis* with IVS1.

Kroos *et al* 2007 published a major core *GAA* haplotype based on 17 variants (some of which are common (MAF >1%), while others are rare (MAF <1%)) in IVS1 haplotypes. This core haplotype was defined by the amino acids DHRGEIVT present in the majority of 174 unaffected individuals from different ethnicities [94]. Analysis using samples from 71 patients that carried the IVS1 variant showed that 68% of IVS1 haplotypes did not harbour any of the 17 variants analyzed. Just one patient carried the major *GAA* core haplotype [94]. It has been hypothesized that the IVS1 allele was propagated as a founder mutation through inheritance and migration, although other scenarios also are possible. Another haplotype study has been performed based on 10 *GAA* SNPs in which Taiwanese newborns with low GAA activity were screened [98]. Analysis in 93 unaffected Taiwanese individuals revealed two core *GAA* haplotypes, DRHGHEIVST (24% of the individuals) and DHRGHEVVST (41%). The IVS1 variant was not present in those individuals. Interestingly, newborns with

low *GAA* activity had another haplotype (DRHSHKIVS), harbouring the common Asian pseudo-deficiency allele c.[1726G>A;2065G>A] in 81% of cases.

GAA haplotypes often vary between patients from different ethnicities. A common way to determine the ancestral chromosomal segment is through the analysis of haplotype blocks. Haplotype blocks are regions with high linkage disequilibrium (LD, nonrandom assortment of alleles at different loci in a population) and low variability in a certain population, passed from a shared ancestor [110] (**Figure 3**). Factors that affect LD are recombination, genetic drift, natural selection and how the population size changes over time. On average, haplotype blocks span from 5kb to 20kb, but they can reach up to 100kb [110]. Haplotype blocks have been used to identify tag SNPs, which can help to predict which other SNPs will be present in the block by calculating the LD [111]. Analyzing haplotypes of disease-associated variants in Pompe disease could help to understand if the presence of certain variants (common or rare) within a haplotype interferes with GAA expression and/or activity.

Aims and scope of this thesis

Since the development of ERT for Pompe disease, substantial progress has been made with the identification of novel disease-associated variants in relation to the clinical phenotypes. A better understanding of the subjacent genetics of Pompe disease combined with the application of improved tools enables the molecular diagnosis of patients, the unravelling of complex cases, and the prediction of phenotypes. The aims of the research described in this thesis are two-fold: Firstly, to improve the enzymatic and molecular diagnosis of Pompe disease (Chapters 2 and 3); Secondly, to investigate and reveal the genotype-phenotype correlation in Pompe disease (Chapters 4-7).

Chapter 2 is a study on enzymatic diagnostic practice over a period of 28 years using different sample sources and methods with which the “gold standard” could be defined of how to diagnose Pompe disease enzymatically. **Chapter 3** illustrates the need for additional analyses to detect new disease-associated *GAA* variants when conventional methods are insufficient. After a diagnosis of a patient with Pompe disease by enzymatic and molecular analysis, a broad spectrum of disease severity is observed in the childhood and adult onset forms. This stresses the need to collect data from many patients to understand the genotype-phenotype correlation. This relation is investigated in the second part of this thesis that is focused on the importance of maintaining a versatile and up-to-date version of the Pompe variant database (**Chapters 4 and 5**). **Chapter 6** analyses the impact of disease-associated variants in the context of *GAA* haplotypes. The research described in **Chapter 7** explores the genotype-phenotype correlation in Pompe disease by comparing mRNA expression levels of genes of affected children and adults with similar *GAA* genotypes. The overall results are discussed in the last chapter of this thesis (**Chapter 8**).

REFERENCES

[1] R. Geddes and G. C. Straton, “The influence of lysosomes on glycogen metabolism,” *Biochem. J.*, vol. 163, no. 2, pp. 193–200, 1977, doi: 10.1042/bj1630193.

[2] P. J. Roach, A. A. Depaoli-Roach, T. D. Hurley, and V. S. Tagliabracci, “Glycogen and its metabolism: Some new developments and old themes,” *Biochem. J.*, vol. 441, no. 3, pp. 763–787, 2012, doi: 10.1042/BJ20111416.

[3] A. T. van der Ploeg *et al.*, “Breakdown of lysosomal glycogen in cultured fibroblasts from glycogenosis type II patients after uptake of acid α-glucosidase,” *J. Neurol. Sci.*, vol. 79, no. 3, pp. 327–336, Jul. 1987, doi: 10.1016/0022-510X(87)90239-5.

[4] M. Kroos, M. Hoogeveen-Westerveld, A. van der Ploeg, and A. J. J. Reuser, “The genotype-phenotype correlation in Pompe disease,” *Am. J. Med. Genet. Part C Semin. Med. Genet.*, vol. 160 C, no. 1, pp. 59–68, Feb. 2012, doi: 10.1002/ajmg.c.31318.

[5] N. Zhong, F. Martiniuk, S. Tzall, and R. Hirschhorn, “Identification of a missense mutation in one allele of a patient with Pompe disease, and use of endonuclease digestion of PCR-amplified RNA to demonstrate lack of mRNA expression from the second allele,” *Am. J. Hum. Genet.*, vol. 49, no. 3, pp. 635–645, 1991, Accessed: Oct. 09, 2020. [Online]. Available: /pmc/articles/PMC1683123/?report=abstract.

[6] A. J. Bergsma, S. L. M. In’t Groen, F. W. Verheijen, A. T. Van Der Ploeg, and W. W. M. P. Pijnappel, “From Cryptic Toward Canonical Pre-mRNA Splicing in Pompe Disease: A Pipeline for the Development of Antisense Oligonucleotides,” *Mol. Ther. - Nucleic Acids*, vol. 5, no. 9, p. e361, Sep. 2016, doi: 10.1038/mtna.2016.75.

[7] A. Reuser, R. Hirschhorn, and M. Kroos, *Pompe Disease: Glycogen Storage Disease Type II, Acid α-Glucosidase (Acid Maltase) Deficiency*. McGraw-Hill Education: New York, NY: The Online Metabolic and Molecular Bases of Inherited Disease, 2018.

[8] M. Y. Niño *et al.*, “Extension of the Pompe mutation database by linking disease-associated variants to clinical severity,” *Hum. Mutat.*, vol. 40, no. 11, pp. 1954–1967, Nov. 2019, doi: 10.1002/humu.23854.

[9] D. Güngör and A. J. J. Reuser, “How to describe the clinical spectrum in Pompe disease?,” *American Journal of Medical Genetics, Part A*, vol. 161, no. 2. Am J Med Genet A, pp. 399–400, Feb. 2013, doi: 10.1002/ajmg.a.35662.

[10] F. Martiniuk *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.*, vol. 79, no. 1, pp. 69–72, 1998, doi: 10.1002/(SICI)1096-8628(19980827)79:1<69::AID-AJMG16>3.0.CO;2-K.

[11] A. E. Slonim, L. Bulone, S. Ritz, T. Goldberg, A. Chen, and F. Martiniuk, “Identification of two subtypes of infantile acid maltase deficiency,” *J. Pediatr.*, vol. 137, no. 2, pp. 283–285, 2000, doi: 10.1067/mpd.2000.107112.

[12] T. de Barsy, P. Jacquemin, F. Van Hoof, and H. G. Hers, “Enzyme replacement in Pompe disease: an attempt with purified human acid alpha-glucosidase.,” *Birth Defects Orig. Artic. Ser.*, vol. 9, no. 2, pp. 184–90, Mar. 1973, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/4611528.

[13] J. M. P. Van den Hout, A. J. J. Reuser, J. B. C. de Klerk, W. F. Arts, J. A. M. Smeitink, and A. T. Van der Ploeg, “Enzyme therapy for Pompe disease with recombinant human α-glucosidase from rabbit milk,” *J. Inherit. Metab. Dis.*, vol. 24, no. 2, pp. 266–274, 2001, doi: 10.1023/A:1010383421286.

[14] P. S. Kishnani *et al.*, “Chinese hamster ovary cell-derived recombinant human acid α-glucosidase in infantile-onset Pompe disease,” *J. Pediatr.*, vol. 149, no. 1, pp. 89–97, Jul. 2006, doi: 10.1016/j.jpeds.2006.02.035.

[15] A. Toscano and B. Schoser, “Enzyme replacement therapy in late-onset Pompe disease: A systematic literature review,” *Journal of Neurology*, vol. 260, no. 4. J Neurol, pp. 951–959, Apr. 2013, doi: 10.1007/s00415-012-6636-x.

[16] P. S. Kishnani *et al.*, “Cross-reactive immunologic material status affects treatment outcomes in Pompe disease infants,” *Mol. Genet. Metab.*, vol. 99, no. 1, pp. 26–33, Jan. 2010, doi: 10.1016/j.ymgme.2009.08.003.

[17] C. M. van Gelder *et al.*, “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.*, vol. 39, no. 3, pp. 383–390, May 2016, doi: 10.1007/s10545-015-9912-y.

[18] E. Poelman *et al.*, "Effects of higher and more frequent dosing of alglucosidase alfa and immunomodulation on long-term clinical outcome of classic infantile Pompe patients," *J. Inherit. Metab. Dis.*, 2020, doi: 10.1002/jimd.12268.

[19] R. R. Howell, B. Byrne, B. T. Darras, P. Kishnani, M. Nicolino, and A. Van Der Ploeg, "Diagnostic challenges for Pompe disease: An under-recognized cause of floppy baby syndrome," *Genet. Med.*, vol. 8, no. 5, pp. 289–296, May 2006, doi: 10.1097/01.gim.0000204462.42910.b8.

[20] B. J. Ebbink *et al.*, "Cognitive decline in classic infantile Pompe disease: An underacknowledged challenge," *Neurology*, vol. 86, no. 13, pp. 1260–1261, Mar. 2016, doi: 10.1212/WNL.0000000000002523.

[21] A. Korlimarla, J.-A. Lim, P. S. Kishnani, and B. Sun, "An emerging phenotype of central nervous system involvement in Pompe disease: from bench to bedside and beyond," *Ann. Transl. Med.*, vol. 7, no. 13, pp. 289–289, Jul. 2019, doi: 10.21037/atm.2019.04.49.

[22] M. L. C. Hagemans, L. P. F. Winkel, W. C. J. Hop, A. J. J. Reuser, P. A. Van Doorn, and A. T. Van Der Ploeg, "Disease severity in children and adults with Pompe disease related to age and disease duration," *Neurology*, vol. 64, no. 12, pp. 2139–2141, Jun. 2005, doi: 10.1212/01.WNL.0000165979.46537.56.

[23] P. S. Kishnani and R. R. Howell, "Pompe disease in infants and children," in *Journal of Pediatrics*, 2004, vol. 144, no. 5 SUPPL., doi: 10.1016/j.jpeds.2004.01.053.

[24] M. Roberts *et al.*, "The prevalence and impact of scoliosis in Pompe disease: Lessons learned from the Pompe Registry," *Mol. Genet. Metab.*, vol. 104, no. 4, pp. 574–582, Dec. 2011, doi: 10.1016/j.ymgme.2011.08.011.

[25] L. P. F. Winkel *et al.*, "The natural course of non-classic Pompe's disease; a review of 225 published cases," *Journal of Neurology*, vol. 252, no. 8. J Neurol, pp. 875–884, Aug. 2005, doi: 10.1007/s00415-005-0922-9.

[26] J. C. van der Meijden, M. E. Kruijshaar, L. Harlaar, D. Rizopoulos, N. A. M. E. van der Beek, and A. T. van der Ploeg, "Long-term follow-up of 17 patients with childhood Pompe disease treated with enzyme replacement therapy," *J. Inherit. Metab. Dis.*, vol. 41, no. 6, pp. 1205–1214, Dec. 2018, doi: 10.1007/s10545-018-0166-3.

[27] N. Pellegrini *et al.*, "Respiratory insufficiency and limb muscle weakness in adults with Pompe's disease," *Eur. Respir. J.*, vol. 26, no. 6, pp. 1024–1031, Dec. 2005, doi: 10.1183/09031936.05.00020005.

[28] N. A. M. E. Van Der Beek *et al.*, "Clinical features and predictors for disease natural progression in adults with Pompe disease: A nationwide prospective observational study," *Orphanet J. Rare Dis.*, vol. 7, no. 1, 2012, doi: 10.1186/1750-1172-7-88.

[29] B. de S. Sixel *et al.*, "Respiratory manifestations in late-onset Pompe disease: a case series conducted in Brazil," *J. Bras. Pneumol.*, vol. 43, no. 1, pp. 54–59, 2017, doi: 10.1590/S1806-37562015000000343.

[30] N. A. M. E. Van der Beek *et al.*, "Rate of disease progression during long-term follow-up of patients with late-onset Pompe disease," *Neuromuscul. Disord.*, vol. 19, no. 2, pp. 113–117, Feb. 2009, doi: 10.1016/j.nmd.2008.11.007.

[31] E. Kuperus *et al.*, "Long-term benefit of enzyme replacement therapy in Pompe disease: A 5-year prospective study," *Neurology*, vol. 89, no. 23, pp. 2365–2373, Dec. 2017, doi: 10.1212/WNL.0000000000004711.

[32] A. T. van der Ploeg *et al.*, "A Randomized Study of Alglucosidase Alfa in Late-Onset Pompe's Disease," *N. Engl. J. Med.*, vol. 362, no. 15, pp. 1396–1406, Apr. 2010, doi: 10.1056/nejmoa0909859.

[33] A. T. Van der Ploeg *et al.*, "Open-label extension study following the Late-Onset Treatment Study (LOTS) of alglucosidase alfa," *Mol. Genet. Metab.*, vol. 107, no. 3, pp. 456–461, Nov. 2012, doi: 10.1016/j.ymgme.2012.09.015.

[34] L. Harlaar *et al.*, "Large variation in effects during 10 years of enzyme therapy in adults with Pompe disease," *Neurology*, vol. 93, no. 19, pp. e1756–e1767, Nov. 2019, doi: 10.1212/WNL.0000000000008441.

[35] H. A. Wisselaar, M. A. Kroos, M. M. P. Hermans, J. Van Beeumen, and A. J. J. Reuser, "Structural and functional changes of lysosomal acid α-glucosidase during intracellular transport and maturation," *J. Biol. Chem.*, 1993.

[36] A. J. J. Reuser, M. Kroos, R. Willemsen, D. Swallow, J. M. Tager, and H. Galjaard, "Clinical Diversity in glycogenosis type II. Biosynthesis and in situ localization of acid α-glucosidase in mutant fibroblasts," *J. Clin. Invest.*, vol. 79, no. 6, pp. 1689–1699, 1987, doi: 10.1172/JCI113008.

[37] N. Raben, P. Plotz, and B. Byrne, “Acid a-Glucosidase Deficiency (Glycogenosis Type II, Pompe Disease),” *Curr. Mol. Med.*, vol. 2, no. 2, pp. 145–166, Jul. 2005, doi: 10.2174/1566524024605789.

[38] J.-A. Lim, O. Kakhlon, L. Li, R. Myerowitz, and N. Raben, “Pompe disease: Shared and unshared features of lysosomal storage disorders,” *Rare Dis.*, vol. 3, no. 1, p. e1068978, Jan. 2015, doi: 10.1080/21675511.2015.1068978.

[39] C. M. Bellettato and M. Scarpa, “Pathophysiology of neuropathic lysosomal storage disorders,” *Journal of Inherited Metabolic Disease*, vol. 33, no. 4. Springer, pp. 347–362, Aug. 29, 2010, doi: 10.1007/s10545-010-9075-9.

[40] J. A. Lim, L. Li, O. Kakhlon, R. Myerowitz, and N. Raben, “Defects in calcium homeostasis and mitochondria can be reversed in Pompe disease,” *Autophagy*, vol. 11, no. 2, pp. 385–402, 2015, doi: 10.1080/15548627.2015.1009779.

[41] A. P. Lieberman, R. Puertollano, N. Raben, S. Slaugenhaupt, S. U. Walkley, and A. Ballabio, “Autophagy in lysosomal storage disorders,” *Autophagy*, vol. 8, no. 5. Taylor and Francis Inc., pp. 719–730, 2012, doi: 10.4161/auto.19469.

[42] T. Jung, N. Bader, and T. Grune, “Lipofuscin: Formation, distribution, and metabolic consequences,” 2007, doi: 10.1196/annals.1404.008.

[43] J. A. Lim, L. Li, and N. Raben, “Pompe disease: From pathophysiology to therapy and back again,” *Frontiers in Aging Neuroscience*, vol. 6, no. JUL. Frontiers Media S.A., 2014, doi: 10.3389/fnagi.2014.00177.

[44] E. Solomon, D. Swallow, S. Burguess, and L. Evans, “Assignment of the human acid α-glucosidase gene (αGLU) to chromosome 17 using somatic cell hybrids,” *Ann. Hum. Genet.*, vol. 42, no. 3, pp. 273–281, 1979, doi: 10.1111/j.1469-1809.1979.tb00661.x.

[45] E. Solomon and D. F. Barker, “Report of the committee on the genetic constitution of chromosome 17,” *Cytogenet. Genome Res.*, vol. 51, no. 1–4, pp. 319–337, 1989, doi: 10.1159/000132797.

[46] L. H. Hoefsloot, M. Hoogeveen-Westerveld, A. J. J. Reuser, and B. A. Oostra, “Characterization of the human lysosomal α-glucosidase gene,” *Biochem. J.*, vol. 272, no. 2, pp. 493–497, 1990, doi: 10.1042/bj2720493.

[47] S. L. M. in ’t Groen *et al.*, “Novel *GAA* Variants and Mosaicism in Pompe Disease Identified by Extended Analyses of Patients with an Incomplete DNA Diagnosis,” *Mol. Ther. - Methods Clin. Dev.*, vol. 17, pp. 337–348, Jun. 2020, doi: 10.1016/j.omtm.2019.12.016.

[48] M. F. Coutinho, M. J. Prata, and S. Alves, “Mannose-6-phosphate pathway: A review on its role in lysosomal function and dysfunction,” *Molecular Genetics and Metabolism*, vol. 105, no. 4. Academic Press, pp. 542–550, Apr. 01, 2012, doi: 10.1016/j.ymgme.2011.12.012.

[49] R. J. Moreland *et al.*, “Lysosomal acid α-glucosidase consists of four different peptides processed from a single chain precursor,” *J. Biol. Chem.*, vol. 280, no. 8, pp. 6780–6791, Feb. 2005, doi: 10.1074/jbc.M404008200.

[50] V. Roig-Zamboni *et al.*, “Structure of human lysosomal acid α-glucosidase-A guide for the treatment of Pompe disease,” *Nat. Commun.*, vol. 8, no. 1, Dec. 2017, doi: 10.1038/s41467-017-01263-3.

[51] P. S. Kishnani *et al.*, “Pompe disease diagnosis and management guideline,” *Genetics in Medicine*, vol. 8, no. 5. Various, pp. 267–288, May 2006, doi: 10.1097/01.gim.0000218152.87434.f3.

[52] N. Preisler *et al.*, “Late-onset Pompe disease is prevalent in unclassified limb-girdle muscular dystrophies,” *Mol. Genet. Metab.*, vol. 110, no. 3, pp. 287–289, Nov. 2013, doi: 10.1016/j.ymgme.2013.08.005.

[53] Z. Lukacs *et al.*, “Prevalence of Pompe disease in 3,076 patients with hyperCKemia and limb-girdle muscular weakness,” *Neurology*, vol. 87, no. 3, pp. 295–298, Jul. 2016, doi: 10.1212/WNL.0000000000002758.

[54] S. P. Young, R. D. Stevens, Y. An, Y. T. Chen, and D. S. Millington, “Analysis of a glucose tetrasaccharide elevated in Pompe disease by stable isotope dilution-electrospray ionization tandem mass spectrometry,” *Anal. Biochem.*, vol. 316, no. 2, pp. 175–180, May 2003, doi: 10.1016/S0003-2697(03)00056-3.

[55] Y. An *et al.*, “Glucose tetrasaccharide as a biomarker for monitoring the therapeutic response to enzyme replacement therapy for Pompe disease,” *Mol. Genet. Metab.*, vol. 85, no. 4, pp. 247–254, Aug. 2005, doi: 10.1016/j.ymgme.2005.03.010.

[56] W. Sluiter *et al.*, “Rapid ultraperformance liquid chromatography-tandem mass spectrometry assay for a characteristic glycogen-derived tetrasaccharide in pompe disease and other glycogen storage diseases,” *Clin. Chem.*, vol. 58, no. 7, pp. 1139–1147, Jul. 2012, doi:

10.1373/clinchem.2011.178319.
[57] M. R. Heiner-Fokkema *et al.*, "The multiple faces of urinary glucose tetrasaccharide as biomarker for patients with hepatic glycogen storage diseases," *Genetics in Medicine*. Springer Nature, pp. 1-2, Jul. 13, 2020, doi: 10.1038/s41436-020-0878-2.
[58] J. Finsterer, "An update on diagnosis and therapy of metabolic myopathies," *Expert Rev. Neurother.*, vol. 18, no. 12, pp. 933-943, Dec. 2018, doi: 10.1080/14737175.2018.1550360.
[59] A. Toscano, E. Barca, and O. Musumeci, "Update on diagnostics of metabolic myopathies," *Current Opinion in Neurology*, vol. 30, no. 5. Lippincott Williams and Wilkins, pp. 553-562, Oct. 01, 2017, doi: 10.1097/WCO.0000000000000483.
[60] A. Golsari, A. Nasimzadah, G. Thomalla, S. Keller, C. Gerloff, and T. Magnus, "Prevalence of adult Pompe disease in patients with proximal myopathic syndrome and undiagnosed muscle biopsy," *Neuromuscul. Disord.*, vol. 28, no. 3, pp. 257-261, Mar. 2018, doi: 10.1016/j.nmd.2017.12.001.
[61] B. Winchester *et al.*, "Methods for a prompt and reliable laboratory diagnosis of Pompe disease: Report from an international consensus meeting," *Molecular Genetics and Metabolism*, vol. 93, no. 3. Academic Press Inc., pp. 275-281, 2008, doi: 10.1016/j.ymgme.2007.09.006.
[62] J. Sheth *et al.*, "Prenatal Diagnosis of Lysosomal Storage Disorders by Enzymes Study Using Chorionic Villus and Amniotic Fluid," *J. Fetal Med.*, vol. 1, no. 1, pp. 17-24, Mar. 2014, doi: 10.1007/s40556-014-0001-3.
[63] D. Li *et al.*, "Early prenatal diagnosis of lysosomal storage disorders by enzymatic and molecular analysis," *Prenat. Diagn.*, vol. 38, no. 10, pp. 779-787, Sep. 2018, doi: 10.1002/pd.5329.
[64] A. J. J. Reuser *et al.*, "Enzymatic and molecular strategies to diagnose Pompe disease," *Expert Opinion on Medical Diagnostics*, vol. 4, no. 1. Expert Opin Med Diagn, pp. 79-89, Jan. 2010, doi: 10.1517/17530050903460300.
[65] M. Piraud *et al.*, "Contribution of tandem mass spectrometry to the diagnosis of lysosomal storage disorders," *J. Inherit. Metab. Dis.*, vol. 41, no. 3, pp. 457-477, May 2018, doi: 10.1007/s10545-017-0126-3.
[66] N. A. Chamoles, M. Blanco, D. Gaggioli, and C. Casentini, "Tay-Sachs and Sandhoff diseases: Enzymatic diagnosis in dried blood spots on filter paper: Retrospective diagnoses in newborn-screening cards," *Clin. Chim. Acta*, vol. 318, no. 1-2, pp. 133-137, 2002, doi: 10.1016/S0009-8981(02)00002-5.
[67] M. H. Gelb, F. Turecek, C. R. Scott, and N. A. Chamoles, "Direct multiplex assay of enzymes in dried blood spots by tandem mass spectrometry for the newborn screening of lysosomal storage disorders," *J. Inherit. Metab. Dis.*, vol. 29, no. 2-3, pp. 397-404, Apr. 2006, doi: 10.1007/s10545-006-0265-4.
[68] P. V. Hopkins, C. Campbell, T. Klug, S. Rogers, J. Raburn-Miller, and J. Kiesling, "Lysosomal storage disorder screening implementation: Findings from the first six months of full population pilot testing in Missouri," *J. Pediatr.*, vol. 166, no. 1, pp. 172-177, Jan. 2015, doi: 10.1016/j.jpeds.2014.09.023.
[69] V. La Cognata, M. Guarnaccia, A. Polizzi, M. Ruggieri, and S. Cavallaro, "Highlights on Genomics Applications for Lysosomal Storage Diseases," *Cells*, vol. 9, no. 8, p. 1902, Aug. 2020, doi: 10.3390/cells9081902.
[70] H. C. Liao *et al.*, "Mass spectrometry but not fluorimetry distinguishes affected and pseudodeficiency patients in newborn screening for pompe disease," *Clin. Chem.*, vol. 63, no. 7, pp. 1271-1277, Jul. 2017, doi: 10.1373/clinchem.2016.269027.
[71] K. A. Sanders *et al.*, "A comparative effectiveness study of newborn screening methods for four lysosomal storage disorders," *Int. J. Neonatal Screen.*, vol. 6, no. 2, Jun. 2020, doi: 10.3390/ijns6020044.
[72] M. Y. Niño *et al.*, "Enzymatic diagnosis of Pompe disease: lessons from 28 years of experience," *Eur. J. Hum. Genet.*, vol. 29, no. 3, pp. 434-446, Mar. 2021, doi: 10.1038/s41431-020-00752-2.
[73] S. C. A. Wens *et al.*, "Remarkably low fibroblast acid α-glucosidase activity in three adults with Pompe disease," *Mol. Genet. Metab.*, vol. 107, no. 3, pp. 485-489, Nov. 2012, doi: 10.1016/j.ymgme.2012.09.003.
[74] O. Musumeci and A. Toscano, "Diagnostic tools in late onset Pompe disease (LOPD)," *Ann. Transl. Med.*, vol. 7, no. 13, pp. 286-286, Jul. 2019, doi: 10.21037/atm.2019.06.60.
[75] M. L. Huie, K. Anyane-Yeboa, E. Guzman, and R. Hirschhorn, "Homozygosity for multiple contiguous single-nucleotide polymorphisms as an indicator of large heterozygous deletions:

Identification of a novel heterozygous 8-kb intragenic deletion (IVS7-19 to IVS15-17) in a patient with glycogen storage disease type II," *Am. J. Hum. Genet.*, vol. 70, no. 4, pp. 1054–1057, Apr. 2002, doi: 10.1086/339691.

[76] C. Amiñoso *et al.*, "Identification of the first deletion-insertion involving the complete structure of *GAA* gene and part of CCDC40 gene mediated by an Alu element," *Gene*, vol. 519, no. 1, pp. 169–172, Apr. 2013, doi: 10.1016/j.gene.2013.01.051.

[77] M. Savarese *et al.*, "MotorPlex provides accurate variant detection across large muscle genes both in single myopathic patients and in pools of DNA samples," *Acta Neuropathol. Commun.*, vol. 2, no. 1, p. 100, 2014, doi: 10.1186/preaccept-1020679654138560.

[78] K. Komlosi, A. Sólyom, and M. Beck, "The Role of Next-Generation Sequencing in the Diagnosis of Lysosomal Storage Disorders," *J. Inborn Errors Metab. Screen.*, vol. 4, Jul. 2016, doi: 10.1177/2326409816669376.

[79] J. S. Berg *et al.*, "Newborn sequencing in genomic medicine and public health," *Pediatrics*, vol. 139, no. 2. American Academy of Pediatrics, Feb. 01, 2017, doi: 10.1542/peds.2016-2252.

[80] E. Friedman, "Next generation sequencing for newborn screening: are we there yet?," *Genet. Res. (Camb).*, vol. 97, p. e17, 2015, doi: 10.1017/S001667231500018X.

[81] P. D. Stenson, M. Mort, E. V. Ball, K. Shaw, A. D. Phillips, and D. N. Cooper, "The Human Gene Mutation Database: Building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine," *Human Genetics*, vol. 133, no. 1. Springer, pp. 1-9, Jan. 28, 2014, doi: 10.1007/s00439-013-1358-4.

[82] I. F. A. C. Fokkema, P. E. M. Taschner, G. C. P. Schaafsma, J. Celli, J. F. J. Laros, and J. T. den Dunnen, "LOVD v.2.0: The next generation in gene variant databases," *Hum. Mutat.*, vol. 32, no. 5, pp. 557–563, May 2011, doi: 10.1002/humu.21438.

[83] D. O. S. de Faria *et al.*, "Update of the Pompe variant database for the prediction of clinical phenotypes: Novel disease-associated variants, common sequence variants, and results from newborn screening," *Hum. Mutat.*, vol. 42, no. 2, pp. 119–134, Feb. 2021, doi: 10.1002/humu.24148.

[84] C. Semplicini *et al.*, "Late-onset Pompe disease in France: molecular features and epidemiology from a nationwide study," *J. Inherit. Metab. Dis.*, vol. 41, no. 6, pp. 937–946, Nov. 2018, doi: 10.1007/s10545-018-0243-7.

[85] F. Wang, W. Song, G. Brancati, and L. Segatori, "Inhibition of endoplasmic reticulum-associated degradation rescues native folding in loss of function protein misfolding diseases," *J. Biol. Chem.*, vol. 286, no. 50, pp. 43454–43464, Dec. 2011, doi: 10.1074/jbc.M111.274332.

[86] M. G. E. M. Ausems *et al.*, "Dutch patients with glycogen storage disease type II show common ancestry for the 525delT and del exon 18 mutations [1]," *Journal of Medical Genetics*, vol. 38, no. 8. BMJ Publishing Group, pp. 527–529, 2001, doi: 10.1136/jmg.38.8.527.

[87] A. J. Bergsma *et al.*, "A genetic modifier of symptom onset in Pompe disease," *EBioMedicine*, vol. 43, pp. 553–561, May 2019, doi: 10.1016/j.ebiom.2019.03.048.

[88] S. Zampieri *et al.*, "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.*, vol. 19, no. 4, pp. 422–431, Apr. 2011, doi: 10.1038/ejhg.2010.188.

[89] C. F. Boerkoel *et al.*, "Leaky splicing mutation in the acid maltase gene is associated with delayed onset of glycogenosis type II," *Am. J. Hum. Genet.*, vol. 56, no. 4, pp. 887–897, 1995, Accessed: Aug. 12, 2020. [Online]. Available: /pmc/articles/PMC1801206/?report=abstract.

[90] A. Dardis *et al.*, "Functional characterization of the common c.-32-13T>G mutation of *GAA* gene: Identification of potential therapeutic agents," *Nucleic Acids Res.*, vol. 42, no. 2, pp. 1291–1302, 2014, doi: 10.1093/nar/gkt987.

[91] A. J. Bergsma, M. Kroos, M. Hoogeveen-Westerveld, D. Halley, A. T. van der Ploeg, and W. W. Pijnappel, "Identification and Characterization of Aberrant *GAA* Pre-mRNA Splicing in Pompe Disease Using a Generic Approach," *Hum. Mutat.*, vol. 36, no. 1, pp. 57–68, Jan. 2015, doi: 10.1002/humu.22705.

[92] E. van der Wal, A. J. Bergsma, J. M. Pijnenburg, A. T. van der Ploeg, and W. W. M. P. Pijnappel, "Antisense Oligonucleotides Promote Exon Inclusion and Correct the Common c.-32-13T>G *GAA* Splicing Variant in Pompe Disease," *Mol. Ther. - Nucleic Acids*, vol. 7, pp. 90–100, Jun. 2017, doi: 10.1016/j.omtn.2017.03.001.

[93] N. Raben, R. C. Nichols, F. Martiniuk, and P. H. Plotz, "A model of mRNA splicing in adult

lysosomal storage disease (glycogenosis type II)," *Hum. Mol. Genet.*, vol. 5, no. 7, pp. 995–1000, Jul. 1996, doi: 10.1093/hmg/5.7.995.

[94] M. A. Kroos *et al.*, "Broad spectrum of Pompe disease in patients with the same c.-32-13T→G haplotype," *Neurology*, vol. 68, no. 2, pp. 110–115, Jan. 2007, doi: 10.1212/01.wnl.0000252798.25690.76.

[95] S. C. A. Wens *et al.*, "Phenotypical variation within 22 families with Pompe disease," *Orphanet J. Rare Dis.*, vol. 8, no. 1, 2013, doi: 10.1186/1750-1172-8-182.

[96] M. L. Huie, A. L. Shanske, J. S. Kasper, R. W. Marion, and R. Hirschhorn, "A large Alu-mediated deletion, identified by PCR, as the molecular basis for glycogen storage disease type II (GSDII)," *Hum. Genet.*, vol. 104, no. 1, pp. 94–98, 1999, doi: 10.1007/s004390050916.

[97] I. Labrijn-Marks *et al.*, "Segmental and total uniparental isodisomy (UPiD) as a disease mechanism in autosomal recessive lysosomal disorders: evidence from SNP arrays," *Eur. J. Hum. Genet.*, vol. 27, no. 6, pp. 919–927, Jun. 2019, doi: 10.1038/s41431-019-0348-y.

[98] P. Labrousse *et al.*, "Genetic heterozygosity and pseudodeficiency in the Pompe disease newborn screening pilot program," *Mol. Genet. Metab.*, vol. 99, no. 4, pp. 379–383, Apr. 2010, doi: 10.1016/j.ymgme.2009.12.014.

[99] L. T. Turaça *et al.*, "Novel *GAA* mutations in patients with Pompe disease," *Gene*, vol. 561, no. 1, 2015, doi: 10.1016/j.gene.2015.02.023.

[100] F. Martiniuk, M. Bodkin, S. Tzall, and R. Hirschhorn, "Identification of the base-pair substitution responsible for a human acid alpha glucosidase allele with lower 'affinity' for glycogen (*GAA* 2) and transient gene expression in deficient cells," *Am. J. Hum. Genet.*, vol. 47, no. 3, pp. 440–445, 1990, Accessed: Aug. 13, 2020. [Online]. Available: /pmc/articles/PMC1683879/?report=abstract.

[101] D. M. Swallow *et al.*, "An investigation of the properties and possible clinical significance of the lysosomal α-glucosidase *GAA* 2 allele," *Ann. Hum. Genet.*, vol. 53, no. 2, pp. 177–184, 1989, doi: 10.1111/j.1469-1809.1989.tb01782.x.

[102] T. Okumiya *et al.*, "A new diagnostic assay for glycogen storage disease type II in mixed leukocytes," *Mol. Genet. Metab.*, vol. 88, no. 1, pp. 22–28, May 2006, doi: 10.1016/j.ymgme.2005.10.016.

[103] M. A. Kroos *et al.*, "p.[G576S; E689K]: Pathogenic combination or polymorphism in Pompe disease?," *Eur. J. Hum. Genet.*, vol. 16, no. 8, pp. 875–879, Aug. 2008, doi: 10.1038/ejhg.2008.34.

[104] I. Dunham *et al.*, "An integrated encyclopedia of DNA elements in the human genome," *Nature*, vol. 489, no. 7414, pp. 57–74, Sep. 2012, doi: 10.1038/nature11247.

[105] F. Robert and J. Pelletier, "Exploring the Impact of Single-Nucleotide Polymorphisms on Translation," *Front. Genet.*, vol. 9, p. 507, Oct. 2018, doi: 10.3389/fgene.2018.00507.

[106] J. H. Moore and S. M. Williams, "Epistasis and Its Implications for Personal Genetics," *American Journal of Human Genetics*, vol. 85, no. 3. Cell Press, pp. 309–320, Sep. 11, 2009, doi: 10.1016/j.ajhg.2009.08.006.

[107] P. Li, M. Guo, C. Wang, X. Liu, and Q. Zou, "An overview of SNP interactions in genome-wide association studies," *Brief. Funct. Genomics*, vol. 14, no. 2, pp. 143–155, Mar. 2015, doi: 10.1093/bfgp/elu036.

[108] R. S. Spielman, L. A. Bastone, J. T. Burdick, M. Morley, W. J. Ewens, and V. G. Cheung, "Common genetic variants account for differences in gene expression among ethnic groups," *Nat. Genet.*, vol. 39, no. 2, pp. 226–231, Feb. 2007, doi: 10.1038/ng1955.

[109] H. J. Westra and L. Franke, "From genome to function by studying eQTLs," *Biochimica et Biophysica Acta - Molecular Basis of Disease*, vol. 1842, no. 10. Elsevier, pp. 1896–1902, Oct. 01, 2014, doi: 10.1016/j.bbadis.2014.04.024.

[110] J. D. Wall and J. K. Pritchard, "Haplotype blocks and linkage disequilibrium in the human genome," *Nature Reviews Genetics*, vol. 4, no. 8. Nature Publishing Group, pp. 587–597, Aug. 01, 2003, doi: 10.1038/nrg1123.

[111] D. O. Stram, "Tag SNP selection for association studies," *Genetic Epidemiology*, vol. 27, no. 4. Genet Epidemiol, pp. 365–374, Dec. 2004, doi: 10.1002/gepi.20028.

[112] L. R. Cardon and G. R. Abecasis, "Using haplotype blocks to map human complex trait loci," *Trends in Genetics*, vol. 19, no. 3. Elsevier Ltd, pp. 135–140, Mar. 01, 2003, doi: 10.1016/S0168-9525(03)00022-2.

CHAPTER 2

Enzymatic diagnosis of Pompe disease: lessons from 28 years of experience

Monica Y. Niño, Mark Wijgerde* , **Douglas Oliveira Soares de Faria***, Marianne Hoogeveen Westerveld, Atze J. Bergsma, Mike Broeders, Nadine A. M. E. van der Beek, J. M. van den Hout, Ans T. van der Ploeg, Frans W. Verheijen, W. W. M. Pim Pijnappel

*These authors contributed equally to this study
Eur J Hum Genet. 2021 Mar;29(3):434-446.
doi:10.1038/s41431-020-00752-2. Epub 2020 Nov 8

European Journal of Human Genetics (2021) 29:434–446
https://doi.org/10.1038/s41431-020-00752-2

ESHG

ARTICLE

Enzymatic diagnosis of Pompe disease: lessons from 28 years of experience

Monica Y. Niño[1,2,3] · Mark Wijgerde[2] · Douglas Oliveira Soares de Faria[1,2,3] · Marianne Hoogeveen-Westerveld[2] · Atze J. Bergsma[1,2,3] · Mike Broeders[1,2,3] · Nadine A. M. E. van der Beek[3,4] · Hannerieke J. M. van den Hout[1,3] · Ans T. van der Ploeg[1,3] · Frans W. Verheijen[2] · W. W. M. Pim Pijnappel[1,2,3]

Received: 10 March 2020 / Revised: 3 October 2020 / Accepted: 20 October 2020 / Published online: 8 November 2020

Abstract

Pompe disease is a lysosomal and neuromuscular disorder caused by deficiency of acid alpha-glucosidase (GAA), and causes classic infantile, childhood onset, or adulthood onset phenotypes. The biochemical diagnosis is based on GAA activity assays in dried blood spots, leukocytes, or fibroblasts. Diagnosis can be complicated by the existence of pseudodeficiencies, i.e., *GAA* variants that lower GAA activity but do not cause Pompe disease. A large-scale comparison between these assays for patient samples, including exceptions and borderline cases, along with clinical diagnoses has not been reported so far. Here we analyzed GAA activity in a total of 1709 diagnostic cases over the past 28 years using a total of 2591 analyses and we confirmed the clinical diagnosis in 174 patients. We compared the following assays: leukocytes using glycogen or 4MUG as substrate, fibroblasts using 4MUG as substrate, and dried blood spots using 4MUG as substrate. In 794 individuals, two or more assays were performed. We found that phenotypes could only be distinguished using fibroblasts with 4MUG as substrate. Pseudodeficiencies caused by the *GAA2* allele could be ruled out using 4MUG rather than glycogen as substrate in leukocytes or fibroblasts. The Asian pseudodeficiency could only be ruled out in fibroblasts using 4MUG as substrate. We conclude that fibroblasts using 4MUG as substrate provides the most reliable assay for biochemical diagnosis and can serve to validate results from leukocytes or dried blood spots.

Introduction

Pompe disease is an autosomal recessive metabolic disorder caused by acid α-glucosidase (GAA) deficiency, which leads to intralysosomal accumulation of glycogen causing progressive damage especially to cardiac and skeletal muscles [1–3]. Pompe disease presents as a spectrum of phenotypes. The classic infantile form is associated with rapidly progressive general muscle weakness and hypertrophic cardiomyopathy, culminating in death within the first year of life if left untreated. Less severe and less progressive forms can manifest at any age from infancy to late adulthood and present with proximal muscle weakness and/or respiratory problems with minimal or no cardiac involvement [4–7]. These patients may eventually lose ambulation and/or become ventilator dependent. Patients with less severe forms and with the same *GAA* genotype can have broad phenotypic variation, with symptom onset ranging from early infantile to late adult [8]. This indicates the presence of environmental, epigenetic, or genetic modifying factors. Recently we identified the silent, cis-acting

These authors contributed equally: Mark Wijgerde, Douglas Oliveira Soares de Faria

Supplementary information The online version of this article (https://doi.org/10.1038/s41431-020-00752-2) contains supplementary material, which is available to authorized users.

✉ W. W. M. Pim Pijnappel
w.pijnappel@erasmusmc.nl

[1] Department of Pediatrics, Erasmus MC University Medical Center, Rotterdam, The Netherlands

[2] Department of Clinical Genetics, 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 Neurology, Erasmus MC University Medical Center, Rotterdam, The Netherlands

c.510C>T *GAA* variant as a genetic modifier of symptom onset and splicing in Pompe disease [9]. Symptoms in other tissues and organs also occur including in smooth muscle, visceral organs, and, in the classic infantile form, the central nervous system. Enzyme replacement therapy with recombinant human GAA is available and is often applied in combination with immunomodulation in classic infantile cases to reduce the chance of antibody formation [10–13]. Early diagnosis including phenotype prediction is required to optimize counseling and to ensure a timely start of treatment.

Current diagnostic guidelines recommend the establishment of GAA enzyme deficiency with additional confirmation of two disease-associated *GAA* variants [14, 15]. Disease-associated variants, recently linked to clinical phenotypes, are listed in the open access Pompe disease *GAA* variant database [www.pompevariantdatabase.nl] [16, 17]. Different enzymatic diagnostic assays are available in which the biological material and the choice of substrate are variables.

Biological materials can be leukocytes, dried blood spots (DBSs), fibroblasts derived from skin biopsies, and muscle tissue. Leukocytes and DBSs are easily obtained, which is important for timely screening, diagnosis, and treatment of classic infantile patients. Newborn screening (NBS) programs are based on DBS assays and positive cases require a second-tier assay for confirmation [18]. Muscle biopsies can be used for diagnosis however, these are not always available and involve a rather invasive and sometimes painful procedure. In addition, in the case of fat replacement of muscle, which can occur in Pompe disease, it can be difficult to obtain muscle cells from a muscle biopsy. Substrates include the natural substrate glycogen, and the artificial substrates 4-methylumbelliferyl-α-D-glucopyranoside (4MUG) [19] and (GAA-S), a substrate used in tandem mass spectrometry for NBS [20]. In certain assays such as using leukocytes or DBSs, the neutral hydrolase glucoamylase activity needs to be inhibited using acarbose as it interferes with the measurement of GAA activity [21–23].

Additional evidence for the diagnosis of Pompe disease can be obtained by measuring urine tetrasaccharide Glc4 (TGLC), which can be measured using HPLC with UV detection or using mass spectrometry. This biomarker has been found to be sensitive but it is not specific to Pompe disease, as it is also elevated in liver abnormalities or in response to food intake. In borderline cases, it might help to establish the diagnosis [24–26].

For the interpretation of diagnostic outcome, it is important to consider the following information on *GAA* DNA variants. (1) At least two (combinations of) *GAA* variants can lead to pseudodeficiency. In the Caucasian population, the *GAA2* (c.271G>A) pseudodeficiency variant lowers the activity of GAA for the natural substrate glycogen, but not for the artificial substrate 4MUG [27, 28]. In the Asian population, the common *GAA* pseudodeficiency variants c.[1726G>A; 2065G>A] can lower GAA activity to levels that come close to the disease threshold, and their presence can lead to a false positive diagnosis in certain assays [23, 29, 30]. (2) Standard DNA diagnostic analysis may fail to identify *GAA* variants, in these cases extended analysis is required as described [31–33].

Over the past 28 years, our laboratory has processed 1709 diagnostic cases and has diagnosed over 250 patients with Pompe disease using various assays separately and in parallel. In this article we are presenting and reviewing all our test results of blood-based and fibroblast-based assays from 1990 until 2018 allowing to compare the various methods for enzymatic diagnosis and their pitfalls.

Materials and methods

Nomenclature and Pompe variant database

The variant nomenclature is according to HGVS standards [34]. The reference sequences used for the annotation of *GAA* cDNA variants were NM_000152.5 and LRG_673t1.1. Position c.1 represents the first nucleotide of the translation start codon ATG located in exon 2. Exon numbering was according to Niño et al., 2019 [17]. NP_000143.2 was used for annotation of GAA protein variants. Variant information was retrieved from the Pompe disease *GAA* variant database at www.pompevariantdatabase.nl and the Leiden open variation database (http://lovd.nl/gaa).

Diagnostic materials and assay procedures

GAA activity assays were performed when there was a suspicion of Pompe disease based on clinical symptoms (described below). Results were obtained over the period 1990–2018. Reference ranges reflect those of the Diagnostic Department of the Erasmus MC and were established for each assay based on at least 20 individuals per phenotype (classic infantile, childhood, adulthood, healthy controls) with a confirmed clinical diagnosis. Values in between these ranges are specified as ‘gray zone’, which is defined as GAA enzyme activities above the patient range but below the control range. For fibroblasts, the reference ranges were derived from data collected over a 40 years period. For the other assays, including GAA activity in leukocytes and bloodspots, and tetrasaccharide 6-α-D-glucopyranosyl-maltotriose (Glc4) concentration in urine by mass spectrometry, data collection of at least 8-years was used [24]. All reference ranges for patients, healthy controls, and the gray zones are indicated in the legends of the figures. The study

was conducted according to the Declaration of Helsinki. The Medical Ethical Committee at Erasmus University MC approved the study protocol, and all patients, or their parents or legal guardians, provided written informed consent.

Activity assays were performed as previously described [22, 24, 35–37]. A final concentration of 3 µmol/L acarbose (in the reaction mixture) was used for the glycogen assay and 8 µmol/L acarbose for the 4MUG assay to inhibit glucoamylase activity present in mixed leukocytes and which interferes with GAA activity measurements. Performance of the assays was determined for each individual assay by including an internal standard sample. According to standard diagnostic procedures, a maximal deviation of less than 10% of the internal standard was required to use values as diagnostic values. For all assays, no false positive or false negative values of the internal standard were obtained. Typical standard deviations were: leukocytes/4MU: 5.1% ($n = 29$); leukocytes/glycogen: 6.2% ($n = 72$); fibroblasts/4MUG: 7.1% ($n = 78$); DBS (in our Diagnostic Center not used as diagnostic assay): 11.3% ($n = 7$). The GAA activity in DBS was determined in the presence of 8 µmol/L acarbose for glucoamylase inhibition and after precipitation of hemoglobin as previously described [35, 36].

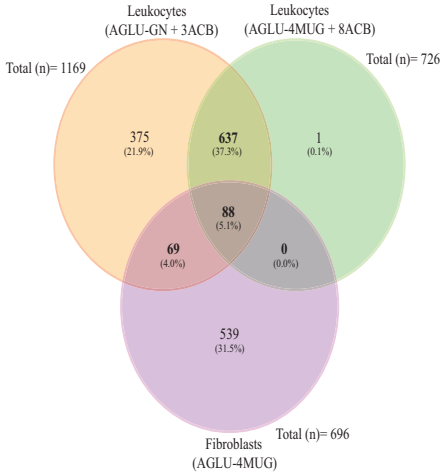


Fig. 1 Venn diagram showing how many diagnostic cases were processed with each of the three different enzymatic methods. The figure shows the similarities and differences of individuals whereby the GAA activity was measured in leukocytes with glycogen as well as with 4MUG as substrate, and in fibroblast with 4MUG. Of note 11 assay results (five from leukocytes with glycogen four from leukocytes with 4MUG, and two from fibroblast with 4MUG) were not included in this comparison since they are biological replicates. DBS assays (20 assays from 18 individuals) are not shown.

Clinical diagnosis

We classified patients with classic-infantile Pompe disease when they presented symptoms at or under 12 months of age. Symptoms in these patients consist of muscle weakness combined with a hypertrophic cardiomyopathy. The phenotype of childhood and adulthood onset Pompe disease is dominated by a progressive limb-girdle myopathy that leads to severe functional limitations, while respiratory muscle dysfunction limits patients' average life span. Childhood onset is here defined as symptom onset before the age of 18 years, adult onset at 18 years or older. The diagnosis was established based on combined evidence consisting of these clinical symptoms, GAA enzyme deficiency, and DNA analysis.

Results

Over the past 28 years, we analyzed a total of 1709 individuals using a total of 2591 assays (Fig. 1 and Table S1). In the majority of individuals, leukocytes were used with both glycogen and 4MUG as substrate ($n = 637$). In 539 individuals, fibroblasts using 4MUG as substrate were used as single assay, while in 375 individuals, leukocytes with glycogen were used as single assay. Only one individual was analyzed with leukocytes using 4MUG as substrate as single assay. Eighty-eight individuals were analyzed with all three assays, while 69 individuals were analyzed with both leukocytes/glycogen and fibroblasts/4MUG. As part of this study we traced back the clinical data of all patients with a confirmed enzymatic diagnosis. The following categories for the diagnoses were defined: 'Classic Infantile', 'Childhood onset' (symptom onset before the age of 18 years), 'Adult onset' (symptom onset at 18 years and older), 'no Pompe disease', and 'Unknown'. For enzymatic assays, the following additional categories were defined: 'Unknown/Deficient' (i.e., the value of the assay was deficient but it was unknown whether the individual had Pompe disease), 'No Pompe/Deficient' (i.e., the value of the assay was deficient but Pompe disease was excluded based on other assays), 'Asymptomatic/Deficient' (i.e., the value of the assay was deficient but the individual was asymptomatic), and 'Gray Zone' (the values above patient ranges but below normal ranges).

GAA activity assay in leukocytes using glycogen as substrate

Figure 2A shows GAA activities using glycogen as substrate that were measured in 1169 individuals. Within the range of 0–3.5 nmol/mg/h for classic infantile patients, there were 20 classic infantile patients, 15 with childhood onset, and 111 with adult onset Pompe disease. Two individuals were 'unknown deficient'. Within the range of 3.5–10 nmol/mg/h, which is the higher activity range for

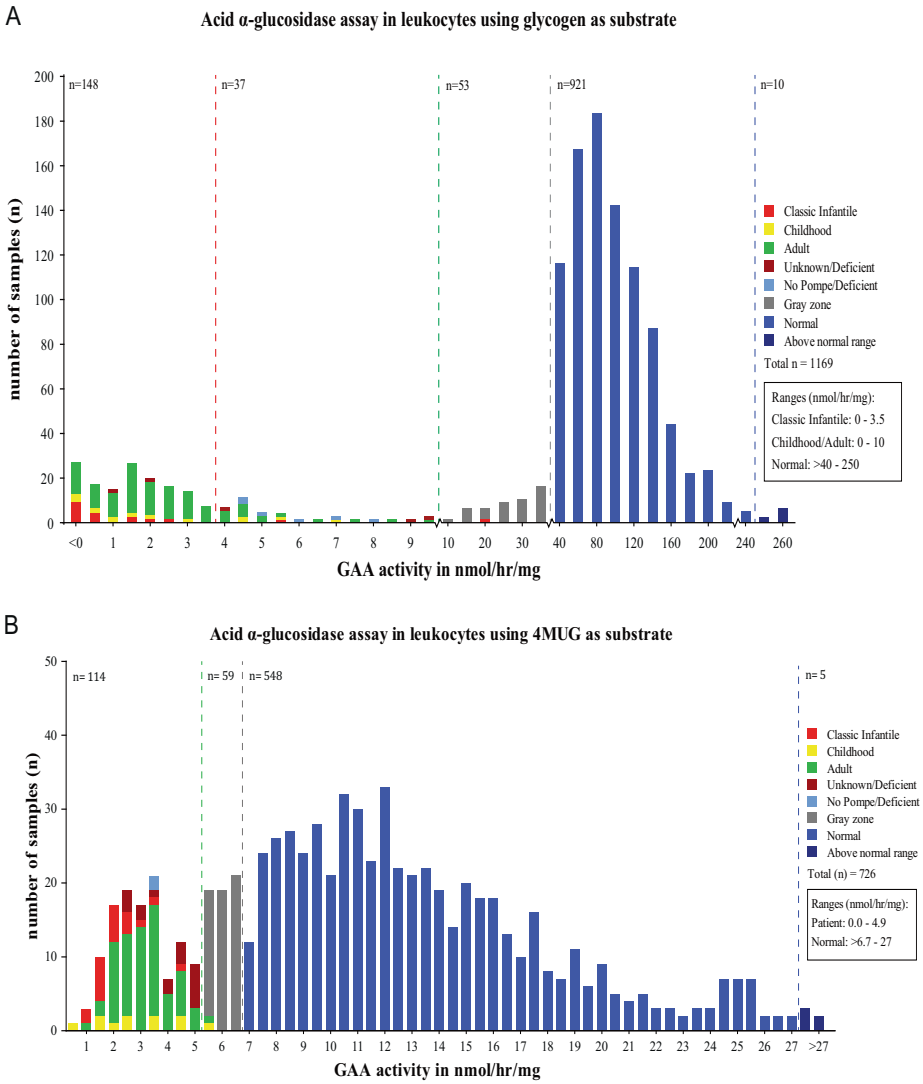


Fig. 2 Distribution of GAA activity in leukocytes using. A Glycogen as substrate in the presence of 3 µmol/L acarbose (AGLU-GN+ 3ACB assay). A total of 1169 individuals suspected of Pompe disease were included. **B** 4MUG as substrate in the presence of 8 µmol/L acarbose (AGLU-4MUG+ 8ACB assay). A total of 726 individuals suspected of Pompe disease were included in this series.

childhood/adult onset patients (range 0–10 nmol/h/mg), there were 1 classic infantile patient, 6 with childhood onset, 20 with adult onset, 7 'no Pompe/deficient', and 3 'unknown/deficient'. This indicated that the range of 0–3.5 nmol/mg/h contained 95% of the classic infantile, 71% of the childhood onset, and 84% of the adulthood onset patients. The 3.5–10 nmol/mg/h range contained 5% of the classic infantile, 29% of the childhood onset, and 16% of the adulthood onset patients. In both ranges, putative false positives were found to a total of 12 individuals. A subset of these false positives was followed up using leukocytes/4MUG and fibroblasts/4MUG assays, and these showed mainly gray zone (seven assays) and normal (nine assays) values, while one assay in leukocytes/4MUG showed an unknown/deficient value (see Table S1 for raw data). Within the range of 10–40 nmol/mg/h, 52 individuals were classified with gray zone (above the patient range but below the control range) values, while 1 classic infantile

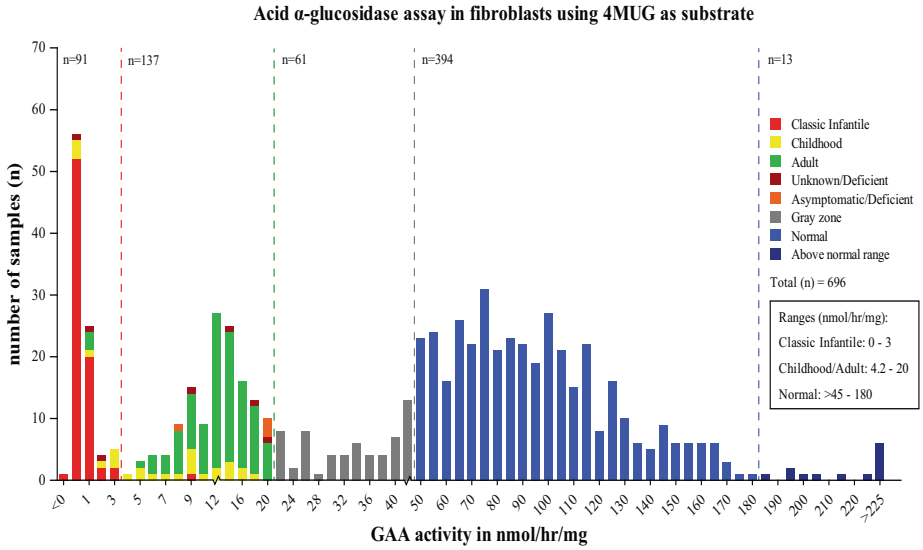


Fig. 3 Distribution of GAA activity in fibroblast using the artificial substrate 4-methylumbelliferyl-α-D-glucoside (4MUG) (AGLU-4MUG assay). A total of 695 individuals suspected of Pompe disease were included in this series.

patient was present. This patient showed values in the patient range in leukocytes/4MUG (3.46 nmol/mg/h) and in the classic infantile range in fibroblasts/4MUG (0.2 nmol/mg/h). In the normal ranges above 40 nmol/mg/h, a total of 921 healthy individuals were present. In conclusion, the leukocytes/glycogen assay has limited ability to discriminate between patients with classic infantile and late onset phenotypes, has a considerable risk of false positive, and a low risk of false negative outcomes.

GAA activity assay in leukocytes using 4MUG as substrate

In Fig. 2B, the activities in leukocyte assays using 4MUG as substrate, measured in 726 individuals, are shown. In the patient range of 0.0–4.9 nmol/h/mg, there were 19 classic infantile patients, 10 with childhood and 66 with adulthood onset, 2 were no Pompe/deficient, and 17 were unknown/deficient. The two individuals that were no Pompe/deficient and the 17 individuals that were unknown/deficient showed either gray zone and/or normal values when tested in leukocytes/4MUG and/or fibroblasts/4MUG. This showed that the leukocyte/4MUG assay did not distinguish classic infantile from late onset phenotypes. The range 4.9–6.7 showed 1 childhood onset patient, 1 adult onset patient, and 57 gray zone individuals. The childhood onset patient had borderline activity just above the threshold (5.14 nmol/mg/h) and tested in the patient range (0.365 nmol/mg/h) in the leukocyte/glycogen assay. The healthy range > 6.7 nmol/mg/h included 548 healthy individuals and 5 individuals with somewhat increased enzyme activities. This indicated that the leukocyte/4MUG assay could faithfully discriminate between healthy and diseased individuals but could also result in false positive outcomes.

GAA activity assay in fibroblasts with 4MUG as substrate

In Fig. 3, the activities in fibroblasts using 4MUG as substrate are shown. In the range 0–3 nmol/h/mg for classic infantile patients, there were 77 classic infantile patients, 8 patients with childhood and 3 with adulthood onset. One patient with childhood onset had an activity in between 3 and 4.2. In the range 4.2–20 nmol/mg/h for childhood/adult onset patients, there were 1 classic infantile patient, 18 with childhood onset and 108 with adulthood onset. There were six asymptomatic/deficient individuals. These individuals may develop symptoms later in life, as symptom onset in Pompe disease can be highly variable. Two of these individuals were also tested using leukocytes/glycogen and 4MUG, and in both individuals, these two assays yielded gray zone values. This indicated that the fibroblasts/4MUG assay could distinguish classic infantile from childhood/adulthood onset patients to a large extent, and between childhood and adulthood patients to a lesser extent. There were 61 individuals with enzyme activities in the gray zone range of >20–45 nmol/h/mg. All individuals with activities in

Table 1 GAA activity in Pompe patients measured in DBS and leukocytes.

Patient	Dried blood spot (DBS)		LEUKOCYTES		Clinical diagnosis
	AGLU-4MUG + 8ACB (pmol/17 h/punch)	BGAL (pmol/17 h/punch)	AGLU-GN+3ACB (nmol/h/mg)	BGAL (nmol/h/mg)	
15	*30.8*	4730	*1.6*	152	Classic infantile Pompe disease
16	*48.8*	2480	*1.0*	131	Childhood onset Pompe disease
17	*13.8*	2860	*–0.1*	162	Classic infantile Pompe disease
18	*25.2*	2520	*0.8*	129	Classic infantile Pompe disease
19	*17.4*	1660	*5.0*	147	Classic infantile Pompe disease
20	*11.4*	1380	*1.2*	192	Classic infantile Pompe disease
21	*10.6*	1780	*–1.1*	174	Classic infantile Pompe disease
22	*27.8*	3100	*–0.2*	192	Classic infantile Pompe disease
23	*14.6*	2040	*–1.6*	167	Classic infantile Pompe disease
24	*12.9*	2950	*0.3*	207	Classic infantile Pompe disease
25	*25.4*	3920	*0.1*	172	Classic infantile Pompe disease
26	*28.6*	3710	*2.8*	104	Childhood onset Pompe disease
27	*25*	4990	*0.9*	222	Unknown
28	*8.51*	3060	*1.2*	161	Classic infantile Pompe disease
29	*41.5*	4700	*–0.4*	181	Classic infantile Pompe disease
30	59.4	3620	*–2.0*	225	Classic infantile Pompe disease
31	58.6	6050	*–1.1*	169	Classic infantile Pompe disease
32*	85.3	4710	*–0.7*	139	Classic infantile Pompe disease
	68	5230			Classic infantile Pompe disease
	48.6	3250			Classic infantile Pompe disease
Normal range	94–448	476–4680	>40–250	50–326	
Patient range	11–56	88	Classic infantile: 0–3.5	0.6–6.3	
			Childhood/Adult: 0–10		

Cases 30–32 were borderline inconclusive using the DBS, but were positively confirmed with leukocytes as sample source. *Patient 32 showed variable measurements of GAA activity and BGAL in DBS (technical replicates were performed). Patient 16 and 26 were diagnosed with childhood Pompe disease, and individual 27 was classified as Unknown/Deficient due lack of clinical information.

Italic: values within the patient range.

Table 2 Follow up of three cases with inconclusive DBS data.

Patient	Genotype DNA (protein)	Predicted severity (Pompe disease *GAA* variant database)	ACMG variant classification	Fibroblast (GAA activity)	Clinical diagnosis
				AGLU-4MUG (nmol/h/mg)	
30	c.1460T>C p.(Phe487Ser);	Potentially less severe	Likely pathogenic	0.432	Classic infantile
	c.1460T>C p.(Phe487Ser)	Potentially less severe	Likely pathogenic		Pompe disease
31	c.379_380del p.(Cys127Leufs*18);	Very severe	Pathogenic	0.35	Classic infantile
	c.525del p.(Glu176Argfs*45)	Very severe	Pathogenic		Pompe disease
32	c.2481+102_2646+31del p.(Gly828_Asn882del);	Very severe	Pathogenic	0.415	Classic infantile
	c.525del p.(Glu176Argfs*45)	Very severe	Pathogenic		Pompe disease
Ranges (GAA activity)		Normal range		>45–180	
		Classic infantile		0–3	
		Childhood/Adult		4.2–20	

DNA analysis and GAA activity in cultured fibroblasts left no doubt that all three patients had classic infantile Pompe disease.

the healthy range >45 nmol/h/mg were healthy. This showed that the fibroblast/4MUG assay can distinguish both healthy from diseased individuals and classic infantile from childhood/adulthood onset phenotypes.

GAA activity assay in DBS

Table 1 shows the results of DBS assays in 18 individuals with Pompe disease. In 15 cases (12 classic infantile, 2 childhood, 1 with an unknown phenotype), GAA activity was in the patient range of 11–56 pmol/17 h/punch, while three individuals (patients 30–32, who had classic infantile Pompe disease) were in the gray zone of 56–94 pmol/17 h/punch. However, all individuals including individuals 30–32 tested in the patient range in leukocytes with glycogen as substrate. One sample, from patient 32 was measured three times, of which one was in the patient range and 2 in the gray zone. Fibroblasts from this patient showed activity in the patient range and DNA analysis of this patient and patients 30 and 31 confirmed the presence of two severe *GAA* disease-associated variants (Table 2 and Fig S1). These results indicated that in this cohort the DBS assay resulted in relatively frequent false negative outcomes (16.7% in this analysis) although the number of DBS samples analyzed here ($n = 18$) was low.

GAA2 pseudodeficiency

Individuals with pseudodeficiencies can give false positive outcomes in GAA enzyme assays. To illustrate this, we compared individuals with confirmed presence of the *GAA2* (c.271G>A) and Asian c.[1726G>A; 2065G>A] pseudodeficiencies using different GAA activity assays.

In four individuals with the *GAA2* variant at heterozygous state without other known GAA disease-associated variants, activities were in the gray zone (two cases) or in the normal range (two cases) in leukocytes/glycogen (Table 3). In leukocytes/4MUG, one case had activity in the gray zone, and three in the normal range. In fibroblasts/4MUG, three patients were tested, two of whom were in the normal range and one (individual 4) in the patient range. This indicated that the *GAA2* allele at heterozygous state can already lower GAA activity in leukocyte-based assays to values in the gray zone. Individual 4 had a medical record of hypertrophic cardiomyopathy. This individual underwent metabolic screening as no cause for his hypertrophic cardiomyopathy could initially be found. Genetic testing using WES led to the identification of a disease-associated variant in exon 19 (c.1831G>A, p.Gly611Ser) of the *MYBPC1* gene (Myosin Binding Protein C, slow type), which plays an important role in muscle contraction and cardiac conduction. No muscle biopsy or MRI was performed, CK was normal (91 U/l). He had an incomplete traumatic cervical spinal cord lesion; he can walk but suffers from spasms due to this condition. General and neurological examination showed no other abnormalities, and there were no signs or symptoms of Pompe disease.

In individuals that were homozygous for *GAA2* (number 5 and 6), GAA activities were in the patient range in leukocytes/glycogen, but in the normal range in leukocytes/4MUG and in fibroblasts/4MUG (Table 3). In individuals that were compound heterozygous for *GAA2* in combination with a second disease-associated *GAA* variant (numbers 7–11), GAA activities were in the patient range in leukocytes/glycogen in all cases (individual 11 contained an additional *GAA* variant classified as childhood when combined with a null allele according to www.pompevariantdatabase.nl) (Table 3). In leukocytes/4MUG and fibroblasts/4MUG, values were either in the gray zone or in the normal range. This highlights that the *GAA2* pseudodeficiency at homozygous or compound heterozygous state can give false positive outcomes when using the leukocyte/glycogen assay, while correct outcomes can be obtained by using either the leukocytes/4MUG or fibroblast/4MUG assays. Patient 12 contained 2 *GAA* disease-

Table 3 GAA enzyme activity in leukocytes and fibroblast from individuals carrying the *GAA2* pseudodeficiency allele (c.271G>A p.(Asp91Asn).

Individual/patient (gender)	Genotype DNA (protein)	Enzymatic assay			Clinical diagnosis
		Leukocytes		Fibroblast	
		AGLU-GN + 3ACB (nmol/h/mg)	AGLU-4MUG+ 8ACB (nmol/h/mg)	AGLU-4MUG (nmol/h/mg)	
1 (M)	c.271G>A p.(Asp91Asn)	36.2	7.9	61.7	No Pompe disease
2 (M)	c.271G>A p.(Asp91Asn)	46.6	8.5	50.4	No Pompe disease
3 (M)	c.271G>A p.(Asp91Asn)	77.6	12.2	N.D.	No Pompe disease
4 (M)[a]	c.271G>A p.(Asp91Asn)	24.0	5.7	*19.5*	No Pompe disease
5 (M)	c.271G>A p.(Asp91Asn); c.271G>A p.(Asp91Asn)	*7.6*	10.9	49.8	No Pompe disease
6 (M)[b]	c.271G>A p.(Asp91Asn); c.271G>A p.(Asp91Asn)	*4.4/7.7*	11.1	86.9	No Pompe disease
7 (F)	c.271G>A p.(Asp91Asn); c.-32-13T>G p.(=), p.(0)	*2.2*	6.1	25.9	No Pompe disease
8 (M)	c.271G>A p.(Asp91Asn); c.-32-13T>G p.(=), p.(0)	*4.1*	N.D.	51.0	No Pompe disease
9 (F)	c.271G>A p.(Asp91Asn); c.-32-13T>G p.(=), p.(0)	*4.7*	6.5	53.5	No Pompe disease
10 (F)	c.271G>A p.(Asp91Asn); c.-32-13T>G p.(=), p.(0)	*6.6*	N.D.	53.1	No Pompe disease
11 (F)[c]	c.271G>A p.(Asp91Asn); c.-32-13T>G p.(=), p.(0); c.1076-22T>G p.?	*5.7*	N.D.	33.3	No Pompe disease
12 (F)	c.271G>A p.(Asp91Asn); c.-32-13T>G p.(=), p.(0); c.1447G>A p.(Gly483Arg)	*1.5*	*1.9*	*7.6*	Adult onset Pompe disease
Ranges	Normal range	>40–250	>6.7–27	>45–180	
	Patient range	Classic infantile: 0–3.5 Childhood/Adult: 0–10	1.1–4.9	Classic infantile: 0–3 Childhood/Adult: 4.2–20	

Italic: values within the patient range.

[a]Unaffected by Pompe disease, but a disease-associated variant was found in the MYBPC1 gene, which is associated with cardiac hypertrophy.

[b]The biological replicate was 7.7 nmol/h/mg in leukocytes using glycogen as substrate.

[c]Unaffected by Pompe disease.

Table 4 Lack of diagnosis in cases with the Asian pseudodeficiency (*c.1726G>A and c.2065G>A*).

Individual (gender, age in years)	Genotype DNA (protein)	Enzymatic assay			Ancillary studies				Clinical diagnosis
		Leukocytes		Fibroblast	Plasma			Urine	
		AGLU-GN+3ACB (nmol/h/mg)	AGLU-4MUG +8ACB (nmol/h/mg)	AGLU-4MUG (nmol/h/mg)	CK (U/l)	ASAT (U/l)	ALAT (U/l)	TGLC (mmol/mol creatinine)	
13 (F, 74 years)	c.1726G>A p.(Gly576Ser); c.2065G>A p.(Glu689Lys); c.-32-13T>G p.(=), p.(0)	*9.6*	*3*	22	58–104	27–34	19–32	*2.3*[a]	No Pompe disease
14 (M, 53 years)	c.1726G>A p.(Gly576Ser); c.2065G>A p.(Glu689Lys); c.1726G>A p.(Gly576Ser); c.2065G>A p.(Glu689Lys)	*18.7*	*3.9*	20.75	992	50	57	0.8	No Pompe disease
Ranges	Normal range	>40–250	>6.7–27	>45–180	W > 17 years: <170 M > 17 years: <200	W > 17 years: <31 M > 17 years: <37	W > 17 years: <31 M > 17 years: <41	>20 years: 0–2.2	
	Patient range	Classic infantile: 0–3.5 Childhood/Adult: 0–10	1.1–4.9	Classic infantile: 0–3 Childhood/ Adult: 4.2–20				>20 years: 2.3–130	

GAA activities were measured both in leukocytes using the natural and artificial substrates as well as in cultured fibroblasts.

Italic: values within the patient range.

[a]2th measurement of TGLC was 5.1 and a 3rd measurement was 11.6 mmol/mol creatinine.

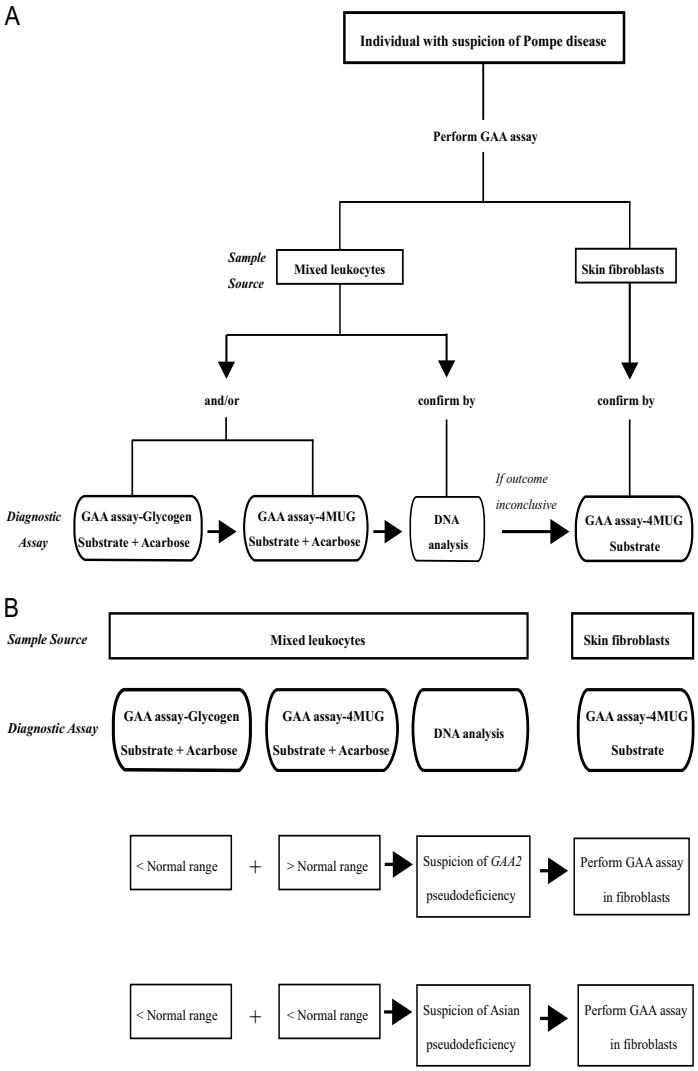


Fig. 4 Flow chart for diagnosis of Pompe disease. A General flow chart. **B** Flow chart for cases with pseudodeficiency.

associated variants in addition to the *GAA2* allele at heterozygous state, and showed GAA values in the patient range in all three assays consistent with the diagnosis of Pompe disease (adulthood onset).

Asian c.[1726G>A; 2065G>A] pseudodeficiency

Individual 13 was compound heterozygous and individual 14 homozygous for the Asian c.[1726G>A; 2065G>A] pseudodeficiency (Table 4). In both cases, assays using leukocytes, either with glycogen or with 4MUG as substrate, resulted in activities in the patient range, while activities in fibroblasts/4MUG were in the gray zone, just above the patient range. Additional diagnostic analyses showed that individual 13 had normal CK, ASAT, and ALAT levels, but slightly elevated TGLC levels. Individual 14 did have elevated CK, ASAT, and ALT levels but normal TGLC levels. Both individuals were diagnosed not to have Pompe disease based on the diagnostic information and the lack of clinical signs associated with Pompe disease. This indicated that the presence of the Asian pseudodeficiency can seriously affect the diagnostic outcome of enzymatic assays that are based on leukocytes, independent of the substrate used.

Discussion

We compared the outcome of different assay procedures for measuring the GAA activity in various sample types using different substrates measured at the Erasmus MC in the last 28 years. The pro's and con's of the different diagnostic procedures are discussed below.

Blood-based assays

Leukocytes

We used leukocyte pellets for diagnostic purposes and applied both glycogen as well as 4MUG as substrate [36]. The results presented in Fig. 2A, B illustrate that a diagnosis can be established in most cases and with both substrates if proper cutoff values are chosen. As previously suggested by van Diggelen et al. [36], our long-term data also demonstrate that the dynamic range, i.e., the difference in values between patients and healthy individuals, is broader with glycogen than with 4MUG. None of the two assays fully discriminate between classic and milder (childhood, adult) phenotypes, but by using both assays in parallel, pseudodeficiencies caused by the Caucasian *GAA2* allele can be excluded (Fig. 4B) [27, 28]. Another pseudodeficiency that complicates the assay is the Asian c.[1726G>A; 2065G>A] pseudodeficiency [23, 29, 30]. This problem applies to all enzymatic procedures used.

Bloodspots

In our small patient cohort, 3 (16.7%) of 18 patients (assays 30–32 in Fig. S1) came out as false negative in the DBS assay but were correctly diagnosed with other diagnostic methods. Analysis of the activity of a second lysosomal enzyme (for instance BGAL) as reference enzyme may help to judge the outcome of the assay. For example, a normal GAA activity should not be trusted if the reference enzyme shows activity outside of the normal ranges. In such an event additional assays are required. In our diagnostic center we do not use the DBS assay routinely as we have fast and standard procedures for leukocyte-based assays. DBS-based assays prove to be very valuable tools for NBS. In both individual cases and in NBS, a second assay is required to confirm the diagnosis.

Despite the ten times higher activity of GAA for glycogen compared to 4MUG, the small sample size of the bloodspot assay precludes the use of glycogen for DBS testing due to the lower sensitivity of the colorimetric assay used with glycogen compared to the fluorimetric detection using 4MUG. The tandem mass spectrometry methodology, employing GAA-S has proven to be a valuable alternative [20]. A recent refinement of that method was reported to separate 96% of the Taiwanese newborns with GAA pseudodeficiency and all Pompe disease carriers from patients with Pompe disease [38]. A recent study from Japan supports this claim [39].

Skin fibroblast assays

Figures 2A, B and 3 clearly demonstrate that the combination of fibroblasts as diagnostic material and the fluorimetric assay using 4MUG as artificial substrate provide the most robust and reliable enzymatic assay for diagnosing Pompe disease. In the majority of cases it distinguishes classic infantile Pompe disease from childhood/adulthood onset phenotypes on the basis of residual activity (although exceptions exist), while none of the other methods do. Assays with intermediate GAA activity were likely derived from Pompe disease carriers. DNA analysis usually provides a clear answer in these cases, but some remained unsolved. For instance, in individual 13, that carried the Asian pseudodeficiency c.[1726G>A; 2065G>A] in combination with the IVS1 variant, biochemical results suggested Pompe disease while clinical signs did not, but it leaves doubt whether adult/late onset Pompe disease can be fully ruled out in this patient.

Conclusions and recommendations

We conclude from our data that cultured skin fibroblasts provide the only sample source that can distinguish classic infantile from childhood/adulthood phenotypes, although exceptions can occur, as noted previously [40]. The laborious procedure of biopsy and culturing makes this assay not suitable for a fast diagnosis. Leukocytes isolated from peripheral blood offer a fast diagnostic sample source, provided that the reaction mixture contains acarbose to inhibit glucoamylase. The use of glycogen as natural substrate enhances the resolution between affected and unaffected, but *GAA2* pseudodeficiency occurs in the Caucasian population and has to be excluded by also using 4MUG as substrate in case of exceptionally low GAA activity (Fig. 4B). The implementation of DBSs has enabled (newborn) screening, and several analytical methods have proven their value as first-tier test. There is general agreement that additional tests remain necessary to finally establish the diagnosis. For reasons mentioned in the text, we advise to conduct genetic testing in addition to any type of enzymatic testing, as recommended recently by the EPOC consortium and others [18, 41].

The recommended enzymatic and molecular diagnostic flow at our Center is as follows: (1) start with leukocytes using 4MUG and glycogen as substrate; (2) confirm with DNA analysis, in which two disease-associated variants should be identified, while a pseudodeficiency does not

qualify as disease-associated variant; (3) if there is doubt, for example when values are close to the gray zone or when a DNA variant is unidentified or novel with unknown significance unknown, use fibroblasts as confirmation (Fig. 4A, B).

Acknowledgements We thank Dr. Arnold Reuser for invaluable discussions.

Funding This work was supported by Departamento Administrativo de Ciencia, Tecnología e Innovación (Colciencias), Colombia (Administrative Department of Science, Technology and Innovation, Colciencias, Colombia), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brasil), Zeldzame Ziekten Fonds/WE Foundation, the Sophia Foundation for Medical Research (SSWO; project number s17-32), Metakids (project number 2016–063 and 2018–082), and Stofwisselkracht.

Compliance with ethical standards

Conflict of interest 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 other authors do not declare any conflict of interest.

References

1. Dasouki M, Jawdat O, Almadhoun O, Pasnoor M, McVey AL, Abuzinadah A, et al. Pompe disease: literature review and case series. Neurol Clin. 2014;32:751–76.
2. Hers HG. alpha-Glucosidase deficiency in generalized glycogen-storage disease (Pompe's disease). Biochem J. 1963;86:11–6.
3. Reuser A, Hirschhorn R, Kroos M. Pompe disease: glycogen storage disease type II, acid α-glucosidase (acid maltase) deficiency. NY: McGraw-Hill Medical; 2018.
4. Herzog A, Hartung R, Reuser AJ, Hermanns P, Runz H, Karabul N, et al. 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. 2012;7:35.
5. Laforet P, Laloui K, Granger B, Hamroun D, Taouagh N, Hogrel JY, et al. The French Pompe registry. Baseline characteristics of a cohort of 126 patients with adult Pompe disease. Rev Neurol. 2013;169:595–602.
6. van Capelle CI, van der Meijden JC, van den Hout JM, Jaeken J, Baethmann M, Voit T, et al. Childhood Pompe disease: clinical spectrum and genotype in 31 patients. Orphanet J Rare Dis. 2016;11:65.
7. van der Ploeg AT, Reuser AJ. Pompe's disease. Lancet. 2008;372:1342–53.
8. Wens SC, van Gelder CM, Kruijshaar ME, de Vries JM, van der Beek NA, Reuser AJ, et al. Phenotypical variation within 22 families with Pompe disease. Orphanet J Rare Dis. 2013;8:182.
9. Bergsma AJ, In 't Groen SLM, van den Dorpel JJA, van den Hout H, van der Beek N, Schoser B, et al. A genetic modifier of symptom onset in Pompe disease. EBioMedicine. 2019;43:553–61.
10. Elder ME, Nayak S, Collins SW, Lawson LA, Kelley JS, Herzog RW, et al. B-Cell depletion and immunomodulation before initiation of enzyme replacement therapy blocks the immune response to acid alpha-glucosidase in infantile-onset Pompe disease. J Pediatr. 2013;163:847–54. e841.
11. Poelman E, Hoogeveen-Westerveld M, Kroos-de Haan MA, van den Hout JMP, Bronsema KJ, van de Merbel NC, et al. High sustained antibody titers in patients with classic infantile Pompe disease following immunomodulation at start of enzyme replacement therapy. J Pediatr. 2018;195:236–43. e233.
12. Poelman E, Hoogeveen-Westerveld M, van den Hout JMP, Bredius RGM, Lankester AC, Driessen GJA, et al. Effects of immunomodulation in classic infantile Pompe patients with high antibody titers. Orphanet J Rare Dis. 2019;14:71.
13. Stenger EO, Kazi Z, Lisi E, Gambello MJ, Kishnani P. Immune tolerance strategies in siblings with infantile Pompe disease-advantages for a preemptive approach to high-sustained antibody titers. Mol Genet Metab Rep. 2015;4:30–4.
14. Musumeci O, Toscano A. Diagnostic tools in late onset Pompe disease (LOPD). Ann Transl Med. 2019;7:286.
15. van der Ploeg AT, Kruijshaar ME, Toscano A, Laforet P, Angelini C, Lachmann RH, et al. European consensus for starting and stopping enzyme replacement therapy in adult patients with Pompe disease: a 10-year experience. Eur J Neurol. 2017;24:768–e31.
16. Kroos M, Hoogeveen-Westerveld M, Michelakakis H, Pomponio R, Van der Ploeg A, Halley D, et al. 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. 2012;33:1161–5.
17. Nino MY, In 't Groen SLM, Bergsma AJ, van der Beek N, Kroos M, Hoogeveen-Westerveld M, et al. Extension of the Pompe mutation database by linking disease-associated variants to clinical severity. Hum Mutat. 2019;40:1954–67.
18. Bodamer OA, Scott CR, Giugliani R. Pompe disease newborn screening working G. Newborn screening for Pompe disease. Pediatrics. 2017;140:S4–13.
19. Robinson D, Price RG, Dance N. Rat-urine glycosidases and kidney damage. Biochem J. 1967;102:533–8.
20. Li Y, Scott CR, Chamoles NA, Ghavami A, Pinto BM, Turecek F, et al. Direct multiplex assay of lysosomal enzymes in dried blood spots for newborn screening. Clin Chem. 2004;50:1785–96.
21. Gelb MH, Turecek F, Scott CR, Chamoles NA. Direct multiplex assay of enzymes in dried blood spots by tandem mass spectrometry for the newborn screening of lysosomal storage disorders. J Inherit Metab Dis. 2006;29:397–404.
22. Okumiya T, Keulemans JL, Kroos MA, Van der Beek NM, Boer MA, Takeuchi H, et al. A new diagnostic assay for glycogen storage disease type II in mixed leukocytes. Mol Genet Metab. 2006;88:22–8.
23. Shigeto S, Katafuchi T, Okada Y, Nakamura K, Endo F, Okuyama T, et al. Improved assay for differential diagnosis between Pompe

disease and acid alpha-glucosidase pseudodeficiency on dried blood spots. Mol Genet Metab. 2011;103:12–7.
24. Sluiter W, van den Bosch JC, Goudriaan DA, van Gelder CM, de Vries JM, Huijmans JG, et al. Rapid ultraperformance liquid chromatography-tandem mass spectrometry assay for a characteristic glycogen-derived tetrasaccharide in Pompe disease and other glycogen storage diseases. Clin Chem. 2012;58:1139–47.
25. Piraud M, Pettazzoni M, de Antonio M, Vianey-Saban C, Froissart R, Chabrol B, et al. Urine glucose tetrasaccharide: a good biomarker for glycogenoses type II and III? A study of the French cohort. Mol Genet Metab Rep. 2020;23:100583.
26. Young SP, Zhang H, Corzo D, Thurberg BL, Bali D, Kishnani PS, et al. Long-term monitoring of patients with infantile-onset Pompe disease on enzyme replacement therapy using a urinary glucose tetrasaccharide biomarker. Genet Med. 2009;11:536–41.
27. Martiniuk F, Bodkin M, Tzall S, Hirschhorn R. Identification of the base-pair substitution responsible for a human acid alpha glucosidase allele with lower "affinity" for glycogen (GAA 2) and transient gene expression in deficient cells. Am J Hum Genet. 1990;47:440–5.
28. Swallow DM, Kroos M, Van der Ploeg AT, Griffiths B, Islam I, Marenah CB, et al. An investigation of the properties and possible clinical significance of the lysosomal alpha-glucosidase GAA*2 allele. Ann Hum Genet. 1989;53:177–84.
29. Kumamoto S, Katafuchi T, Nakamura K, Endo F, Oda E, Okuyama T, et al. High frequency of acid alpha-glucosidase pseudodeficiency complicates newborn screening for glycogen storage disease type II in the Japanese population. Mol Genet Metab. 2009;97:190–5.
30. Tajima Y, Matsuzawa F, Aikawa SI, Okumiya T, Yoshimizu M, Tsukimura T, et al. Structural and biochemical studies on Pompe disease and a "pseudodeficiency of acid alpha-glucosidase". J Hum Genet. 2007;52:898–906.
31. Bergsma AJ, Kroos M, Hoogeveen-Westerveld M, Halley D, van der Ploeg AT, Pijnappel WW. Identification and characterization of aberrant splicing in Pompe disease using a generic approach. Hum Mutat. 2015;36:57–68.
32. Hauser F, Gokce S, Werner G, Danckwardt S, Sollfrank S, Neukirch C, et al. A non-invasive diagnostic assay for rapid detection and characterization of aberrant mRNA-splicing by nonsense mediated decay inhibition. Mol Genet Metab. 2020;130:27–35.
33. In 't Groen SLM, de Faria DOS, Iuliano A, van den Hout JMP, Douben H, Dijkhuizen T, et al. Novel GAA variants and mosaicism in Pompe disease identified by extended analyses of patients with an incomplete DNA diagnosis. Mol Ther Methods Clin Dev. 2020;17:337–48.
34. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et al. HGVS recommendations for the description of sequence variants: 2016 update. Hum Mutat. 2016;37:564–9.
35. Oemardien LF, Boer AM, Ruijter GJ, van der Ploeg AT, de Klerk JB, Reuser AJ, et al. Hemoglobin precipitation greatly improves 4-methylumbelliferone-based diagnostic assays for lysosomal storage diseases in dried blood spots. Mol Genet Metab. 2011;102:44–8.
36. van Diggelen OP, Oemardien LF, van der Beek NA, Kroos MA, Wind HK, Voznyi YV, et al. Enzyme analysis for Pompe disease in leukocytes; superior results with natural substrate compared with artificial substrates. J Inherit Metab Dis. 2009;32:416–23.
37. Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, et al. Measurement of protein using bicinchoninic acid. Anal Biochem. 1985;150:76–85.
38. Liao HC, Chan MJ, Yang CF, Chiang CC, Niu DM, Huang CK, et al. Mass spectrometry but not fluorimetry distinguishes affected and pseudodeficiency patients in newborn screening for Pompe disease. Clin Chem. 2017;63:1271–7.
39. Mashima R, Okuyama T. Enzyme activities of alpha-glucosidase in Japanese neonates with pseudodeficiency alleles. Mol Genet Metab Rep. 2017;12:110–4.
40. Wens SC, Kroos MA, de Vries JM, Hoogeveen-Westerveld M, Wijgerde MG, van Doorn PA, et al. Remarkably low fibroblast acid alpha-glucosidase activity in three adults with Pompe disease. Mol Genet Metab. 2012;107:485–9.
41. van der Ploeg AT, Kruijshaar ME, Toscano A, Laforêt P, Angelini C, Lachmann RH, et al. European consensus for starting and stopping enzyme replacement therapy in adult patients with Pompe disease: a 10-year experience. Eur J Neurol. 2017;24:768–e-31. 2017.

Supplementary Information

Supplementary text

Pseudodeficiencies and borderline cases

c.271G>A (GAA2): common Caucasian background

Table 3 lists individuals that contained the *GAA2* pseudodeficiency allele. In individuals 1-4, only the *GAA2* variant was detected in heterozygous state but no other disease-associated *GAA* variant were detected. GAA enzyme activities were generally low and in some cases in the gray zone, but above the patient range in all three assays (leukocytes using 4MUG or glycogen as substrate and fibroblasts using 4MUG as substrate), except for individual 4. This person showed an activity that was within the patient range in fibroblasts using 4MUG a substrate (19.5 nmol/mg/hr, patient range 4.2-20), while the activity was slightly above the disease threshold in leukocytes using 4MUG as substrate. This individual was diagnosed not to have Pompe disease. The genetic testing led to the identification of a disease-associated variant in exon 19 (c.1831G>A, p.E611K) of the MYBPC1 gene (Myosin Binding Protein C, Slow Type). The GAA enzymatic diagnostic outcome remained enigmatic.

Individual 5 and 6 were homozygous for *GAA2*. In both cases, the activity for glycogen was within the patients' range, but for 4MUG in the normal range. Also based on the normal activity in cultured fibroblasts, both cases were diagnosed not to have Pompe disease.

Individual 7-10 were compound heterozygous for *GAA2* and c.-32-13T>G (IVS1). Also in these cases, the activities using glycogen as substrate were in the patient range, but activities using 4MUG as substrate were in the normal range or gray zone.

Individuals 11 and 12 contained the *GAA2* allele in heterozygous state in addition to two disease-associated variants. Case 11 concerned an unaffected of six children, two females of whom have the adult Pompe disease phenotype. DNA analysis of the mother revealed the presence of two disease-associated variants on one *GAA* allele c.[-32-13T>G; 1076-22T>G] and the *GAA2* variant on the other. The activity in leukocytes was in the patient range when measured with glycogen as substrate, but in the gray zone (33.3 nmol MU/hr/mg) when

measured in fibroblasts with 4MUG as substrate. She was diagnosed not to have Pompe disease. Case 12 demonstrated GAA deficiency in all three assays and was diagnosed with adult onset of Pompe disease. She contained [c.271G>A; c.-32-13T>G on one allele, and the disease-associated variant c.1447G>A p.(Gly483Arg) on the other. This was supported by her clinical history, mentioning unbalanced walking at the age of 26 years, progressing to difficulty with running, respiratory involvement at the age of 43, and weak back and walking difficulties one decade later. She developed walking disability and started ERT at the age of 64 years.

In conclusion, the pseudodeficiency *GAA2* allele can lower GAA enzyme activity measurements when using glycogen as substrate, but not when using 4MUG as substrate. This can result in values within the patient range when using leukocytes with glycogen as substrate in patients that carry a disease-associated variant on one allele and *GAA2* on the second allele.

c.1726G>A and c.2065G>A: common Asian background

Individual 13 was referred at the age of 73 and was compound heterozygous for c.[1726G>A, c.2065G>A] and c.-32-13T>G (IVS1) (Table 4). Activities in leukocytes were in the patient range, both when using glycogen and 4MUG as substrate. The activity in fibroblasts was just above the patient range in the gray zone. This individual also had elevated urinary glucose tetrasaccharide levels, but no pathology in a muscle biopsy. Individual 14 was referred at the age of 60 and was homozygous for c.[1726G>A, c.2065G>A]. GAA enzyme activity was in the patient range in leukocytes with 4MUG as substrate but not with glycogen as substrate; in fibroblasts the activity was just above the patient range. This case also had elevated CK, ASAT, and ALAT levels. Both cases did not display clinical signs of Pompe disease and were diagnosed not to have Pompe disease. They illustrate the difficulty of GAA diagnosis solely based on biochemical assays.

Genetic analysis

DNA sequence analysis has become an alternative or supportive diagnostic tool in Pompe disease. We recently extended the Pompe mutation database, now termed the Pompe disease *GAA* variant database, with clinical information [www.pompevariantdatabase.nl] [1]. There are

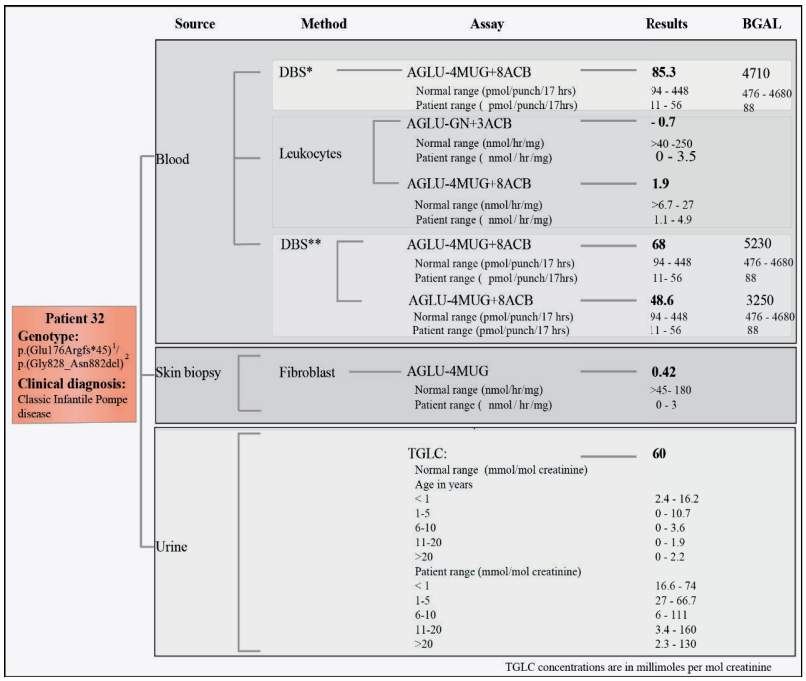
at least two reasons not to put DNA analysis first in line. First, two disease-associated *GAA* variants cannot always be identified by standard diagnostics, for example when the variant is located in an intron [2]. Second, it is not always known whether the variant is disease-associated, despite the large number of variants listed at www.pompevariantdatabase.nl and at the LOVD.

Our present studies focused on the performance of various enzymatic procedures for diagnosing Pompe disease. The pro's and con's of various genetic procedures are largely known and we mention them only briefly. Sequencing of only the *GAA* exons and adjacent regions bears the risk that disease-associated variants in the promoter sequences or deeper intronic sequences are missed. This may be approached using analysis at the mRNA level as shown by us previously [2, 3]. Another limitation is the identification of variants of uncertain significance (VUS) necessitating the performance of an expression assay [4, 5]. NGS procedures, developed for disease specific gene panels, are gaining field, but their application is expected to increase the number of VUSs. The great advantage of an enzymatic assay over DNA analysis is that the enzymatic assay captures in principle all disease-associated *GAA* variants in one. However, when it comes to carrier detection and genetic counseling in families with an index Pompe case, DNA analysis is essential since the activity ranges of carriers and healthy individuals show overlap [6].

References supplementary data

1. Niño, MY, In 't Groen, SLM, Bergsma, AJ, van der Beek, N, Kroos, M, Hoogeveen-Westerveld, M*, et al.* (2019). Extension of the Pompe mutation database by linking disease-associated variants to clinical severity. *Hum Mutat* **40**:1954-1967
2. Bergsma, AJ, In 't Groen, SL, Verheijen, FW, van der Ploeg, AT, and Pijnappel, W (2016). From Cryptic Toward Canonical Pre-mRNA Splicing in Pompe Disease: a Pipeline for the Development of Antisense Oligonucleotides. *Mol Ther Nucleic Acids* **5**: e361.
3. Bergsma, AJ, Kroos, M, Hoogeveen-Westerveld, M, Halley, D, van der Ploeg, AT, and Pijnappel, WW (2015). Identification and characterization of aberrant GAA pre-mRNA splicing in pompe disease using a generic approach. *Hum Mutat* **36**: 57-68.
4. Kroos, M, Pomponio, RJ, van Vliet, L, Palmer, RE, Phipps, M, Van der Helm, R*, et al.* (2008). Update of the Pompe disease mutation database with 107 sequence variants and a format for severity rating. *Hum Mutat* **29**: E13-26.
5. Kroos, M, Hoogeveen-Westerveld, M, Michelakakis, H, Pomponio, R, Van der Ploeg, A, Halley, D*, et al.* (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.
6. Reuser, Verheijen, F, Kroos, M, Okumiya, T, Van Diggelen, O, Van der Ploeg, A*, et al.* (2010). Enzymatic and molecular strategies to diagnose Pompe disease. *Expert Opin Med Diagn* **4**: 79-89.

Supplementary Figure



Supplementary Figure S1. Series of assays used for diagnosis of patient 32. DBSs, leukocytes and fibroblasts were analyzed as well as the Glc4 content of urine. [1] = pathogenic variant from the mother, [2]= pathogenic variant from the father, * a pilot study by DBS showed GAA activity in the grey zone, and BGAL activity higher than normal ranges. Technical replicates were therefore performed (**), as well as assays in leukocytes using glycogen or 4MUG (plus acarbose) as substrate and in fibroblast with 4MUG as substrate.

Supplementary Figure S1, Chapter 2

CHAPTER 3

Novel *GAA* Variants and Mosaicism in Pompe Disease Identified by Extended Analyses of Patients with an Incomplete DNA Diagnosis

Stijn L.M. in 't Groen, **Douglas O.S. de Faria**, Alessandro Iuliano, Johanna M.P. van den Hout, Hannie Douben, Trijnie Dijkhuizen, David Cassiman, Peter Witters, Miguel-Ángel Barba Romero, Annelies de Klein, Galhana M. Somers-Bolman, Jasper J. Saris, Lies H. Hoefsloot, Ans T. van der Ploeg, Atze J. Bergsma, and W.W.M. Pim Pijnappel

Mol Ther Methods Clin Dev. 2020 Jan 13;17:337-348.
doi: 10.1016/j.omtm.2019.12.016

Molecular Therapy
Methods & Clinical Development
Original Article

Novel *GAA* Variants and Mosaicism in Pompe Disease Identified by Extended Analyses of Patients with an Incomplete DNA Diagnosis

Stijn L.M. in 't Groen,[1,2,3] Douglas O.S. de Faria,[1,2,3] Alessandro Iuliano,[1,2,3] Johanna M.P. van den Hout,[1,3] Hannie Douben,[2] Trijnie Dijkhuizen,[4] David Cassiman,[5] Peter Witters,[5] Miguel-Ángel Barba Romero,[6] Annelies de Klein,[2] Galhana M. Somers-Bolman,[2] Jasper J. Saris,[2] Lies H. Hoefsloot,[2] Ans T. van der Ploeg,[1,3] Atze J. Bergsma,[1,2,3] and W.W.M. Pim Pijnappel[1,2,3]

[1]Department of Pediatrics, Erasmus University Medical Center, 3015 GE Rotterdam, the Netherlands; [2]Department of Clinical Genetics, Erasmus University Medical Center, 3015 GE Rotterdam, the Netherlands; [3]Center for Lysosomal and Metabolic Diseases, Erasmus University Medical Center, 3015 GE Rotterdam, the Netherlands; [4]Department of Genetics, University of Groningen, University Medical Center Groningen, 9713 GZ Groningen, the Netherlands; [5]Center for Metabolic Diseases, UZ and KU Leuven, 3000 Leuven, Belgium; [6]Internal Medicine Department, Albacete University Hospital, 02006 Albacete, Spain

Pompe disease is a metabolic disorder caused by a deficiency of the glycogen-hydrolyzing lysosomal enzyme acid α-glucosidase (GAA), which leads to progressive muscle wasting. This autosomal-recessive disorder is the result of disease-associated variants located in the *GAA* gene. In the present study, we performed extended molecular diagnostic analysis to identify novel disease-associated variants in six suspected Pompe patients from four different families for which conventional diagnostic assays were insufficient. Additional assays, such as a generic-splicing assay, minigene analysis, SNP array analysis, and targeted Sanger sequencing, allowed the identification of an exonic deletion, a promoter deletion, and a novel splicing variant located in the 5′ UTR. Furthermore, we describe the diagnostic process for an infantile patient with an atypical phenotype, consisting of left ventricular hypertrophy but no signs of muscle weakness or motor problems. This led to the identification of a genetic mosaicism for a very severe GAA variant caused by a segmental uniparental isodisomy (UPD). With this study, we aim to emphasize the need for additional analyses to detect new disease-associated *GAA* variants and non-Mendelian genotypes in Pompe disease where conventional DNA diagnostic assays are insufficient.

INTRODUCTION

Pompe disease (glycogen storage disease type II, OMIM 232300) is a rare metabolic disorder that leads to progressive muscle wasting caused by acid α-glucosidase (GAA) enzyme deficiency. This autosomal-recessive disorder is caused by disease-associated variants located in the *GAA* gene.[1] Classic infantile Pompe patients have a complete deficiency of GAA enzymatic activity and present hypertrophic cardiomyopathy and severe muscle weakness shortly after birth. Patients with residual GAA activity can be classified as childhood- or adult-onset Pompe patients, based on the age at symptom onset, and generally do not show cardiac involvement.

Enzyme replacement therapy (ERT) is currently the only available treatment. Without treatment, classic infantile Pompe patients do not survive the first year of life due to cardiorespiratory failure.[2,3] In order to become eligible for ERT, a patient needs to present a clinical phenotype related to Pompe disease and have a GAA deficiency, and certain countries require the identification of two disease-associated variants in the *GAA* gene.[4] GAA activity measurements are performed using dried blood with spot sampling or with primary fibroblasts or leukocytes using either glycogen or 4-methylumbelliferone-α-d-glucopyranoside (4MU) as a substrate.[5] False-positive results are known to occur using blood-based assays, for example, those caused by known pseudodeficiency alleles; therefore, DNA analysis is recommended in order to identify disease-associated variants.[5–7] Routine diagnostic DNA analysis usually focuses only on the coding regions of *GAA* using PCR reactions and subsequent Sanger sequencing.[6,8] This method detects variants in the coding regions and in close proximity to splice sites but not in the promoter, UTRs, and most of the intronic regions.[4,9] Reported variants are listed in the "Pompe disease *GAA* variant database" (http://www.pompevariantdatabase.nl), which contains over 400 disease-associated variants in *GAA* and has4 recently been extended to include clinical phenotypes.[10] When the disease-associated variants of both alleles are identified using DNA sequencing, a general prediction of the patient phenotype is possible using the information in this database. Recent findings indicate that analysis of the modifier c.510C> T variant, which is silent but modulates splicing in patients carrying the common c.-32-13T>G (IVS1) variants, is also important.[11]

Received 9 November 2019; accepted 31 December 2019;
https://doi.org/10.1016/j.omtm.2019.12.016.

Correspondence: W.W.M. Pim Pijnappel, Department of Clinical Genetics, Erasmus University Medical Center, Wytemaweg 80, Rotterdam 3015 CN, the Netherlands.
E-mail: w.pijnappel@erasmusmc.nl

Table 1. Basic Information for All Patients at Start of Analysis

	Age at Symptom Onset	GAA Activity in Fibroblasts (Patient Range: 0–20 nmol/h/mg)	Disease-Associated Variant 1	Disease-Associated Variant 2	Initial Symptoms
Patient 1	<1 month	0 nmol/h/mg	c.2331+2T>A p.0?	?	muscle weakness and cardiomyopathy
Patient 2	15 years	13.7 nmol/h/mg	c.-32-13T>G p.=, p.0?	?	muscle weakness
Patient 3 (sibling 1)	24 years	7.5 nmol/h/mg	? + c.2065G>A ? + p.(Glu689Lys)	? + c.2065G>A ? + p.(Glu689Lys)	general fatigue
Patient 4 (sibling 2)	41 years	deficient in lymphocytes	? + c.2065G>A ? + p.(Glu689Lys)	? + c.2065G>A ? + p.(Glu689Lys)	muscle weakness and general fatigue
Patient 5 (sibling 3)	21 years	deficient in lymphocytes	? + c.2065G>A ? + p.(Glu689Lys)	? + c.2065G>A ? + p.(Glu689Lys)	muscle weakness and general fatigue
Patient 6	<1 month	8.1 nmol/h/mg	c.925G>A p.(Gly309Arg)	?	left ventricular hypertrophy

Previously, we described an exon-flanking RT-PCR that can be used to detect novel disease-associated variants that affect pre-mRNA splicing, irrespective of their location.[9,12] This assay is important for diagnosis but also to allow the design of antisense oligonucleotides that can restore splicing.[12–15] Other work has shown single-nucleotide polymorphism (SNP) array analysis to be capable of explaining phenotypes by detecting large deletions or genomic copy number variations (CNVs).[16–18] In addition, SNP array analysis can be used to elucidate homozygote variants incompatible with Mendelian inheritance by detecting uniparental isodisomies (UPDs) and regions of homozygosity (RoHs).[16,17]

Here, we utilized the additional diagnostic assays described above to identify the genetic cause of Pompe disease in six patients from four families with incomplete DNA sequencing results. This resulted in the identification of three novel disease-associated variants located in non-coding regions of *GAA*. Additionally, we describe the genetic analysis performed for a patient with an atypical Pompe phenotype, first described in Labrijn-Marks et al.[17] We provide the clinical symptoms and the series of experiments that were performed in order to make this diagnosis and to strengthen the conclusion that this is a mosaic patient.

RESULTS

The six patients (including three siblings) described here had symptoms that could be attributed to a GAA deficiency. Enzymatic activity of GAA was found deficient in primary fibroblasts and indicated that Pompe disease is a probable cause of symptoms. Genetic screenings of the coding regions of GAA were insufficient to provide a genotype that corresponds with the observed phenotype (Table 1).

Patient 1

The first indication of Pompe disease was observed during prenatal ultrasound, which found evidence of cardiac hypertrophy. Muscle weakness presented shortly after birth, and cultured primary fibroblasts showed no GAA enzyme activity (Table 1). Sanger sequencing of the coding regions identified the heterozygous c.2331+2T>A variant (Table 1; Figure 1A). The intronic c.2331+2T>A variant has been classified as "very severe" and is associated with the classic infantile phenotype when combined with a null allele.[10,19] This variant is located close to the exon 16 splice donor site and affects splicing of intron 16, ultimately resulting in a frameshift, the generation of an early stop codon, and mRNA degradation. The second disease-associated variant was not identified by standard DNA sequencing.

Exon-flanking RT-PCR for all *GAA* exons was performed as previously described[9] to test for the presence of a non-coding intronic variant and revealed an aberrant product of exons 8, 9, and 10 (Figure 1B; product 2). Sanger sequence analysis of this product showed the exclusion of exon 9 from the transcripts, but it failed to identify a candidate splicing variant around the splice junctions (Figure 1C). To test the possibility of a deletion that could have been obscured by the presence of the other allele, we performed flanking exon PCR analysis of exon 9 (ranging from exon 8 to exon 10) using genomic DNA as a template (Figure 1D). This showed the presence of an additional band in the patient DNA that was not found in the healthy control, one that corresponded with the expected product ranging from exon 8 to 10 and one that contained a 343-bp deletion that spanned exon 9 and parts of the neighboring introns (band A in Figure 1E).

The resulting c.1327-61_1437+171del variant explains the exclusion of exon 9 at the RNA level and why the initial DNA sequencing analysis was unable to detect this variant, as the annealing sites for PCR amplification were included in the deleted region. The deletion of exon 9 has been shown to result in a total disruption of GAA enzyme activity in previous work.[9] The combination of c.1327-61_1437+171del and c.2331+2T>A explains the lack of enzyme activity and the classic infantile phenotype of the patient (Figure 1F).

Patient 2

The second patient presented mild muscle weakness at 15 years of age. GAA activity in fibroblasts was found to be reduced to 13.7 nmol/h/mg (patient range: 0–20 nmol/h/mg). Sanger sequencing analysis performed on the patient's DNA identified the c.-32-13T>G

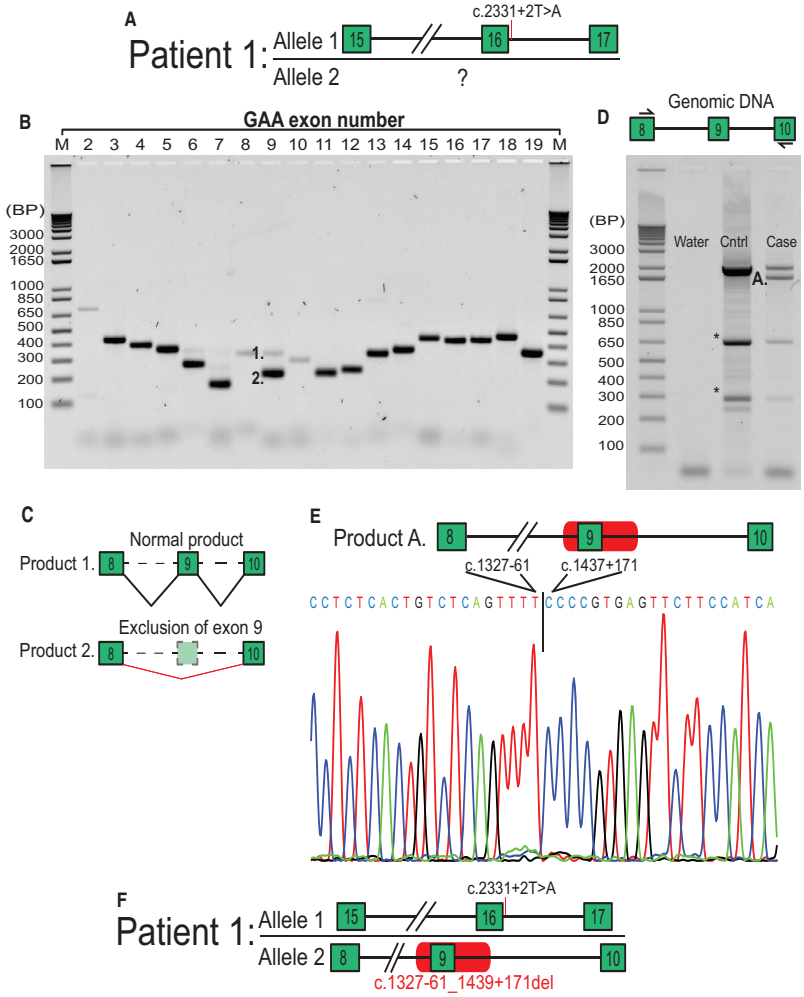


Figure 1. Identification of a Disease-Associated Deletion in Patient 1 Using the Splicing Assay

(A) Genotype of patient 1 at the start of analysis. (B) Splicing assay for *GAA* exons 2–19 performed on cDNA obtained from fibroblasts. Bands marked 1. and 2. indicate two different products detected around exon 9. (C) Cartoons of products 1 and 2 identified using Sanger sequencing of the PCR product. (D) PCR product from *GAA* exon 8 to exon 10, performed on DNA. A., An additional band from patient 1 that was 343 bp smaller compared to the normal product; *, non-specific bands as a result of the PCR. (E) Sanger sequencing of product A. The new junction resulting from the deletion is indicated. (F) Cartoon of both disease-associated variants identified in patient 1.

(IVS1) variant in a heterozygous state but failed to identify a second disease-associated variant (Table 1; Figure 2A)

Exon-flanking RT-PCR of cDNA, derived from the patient's fibroblasts, showed no aberrant splicing products, except for those caused by the IVS1 variant around exon 2 (Figure 2B).[13,20,21] Next, we performed SNP array analysis to test for the presence of large genomic aberrations. This analysis revealed a deletion on chromosome 17 (GRCh37/hg19 [Chr17:78.059.821–78.076.592]) (Figure 2C). Further evaluation of this 17kb deletion indicated that it starts upstream of *GAA* in the *CCDC40* gene and included the promoter, transcription start site (TSS), and non-coding exons 1A and 1B of *GAA*.

Formally proving the pathogenicity of this deletion was complicated, due to the residual GAA expression and activity originating

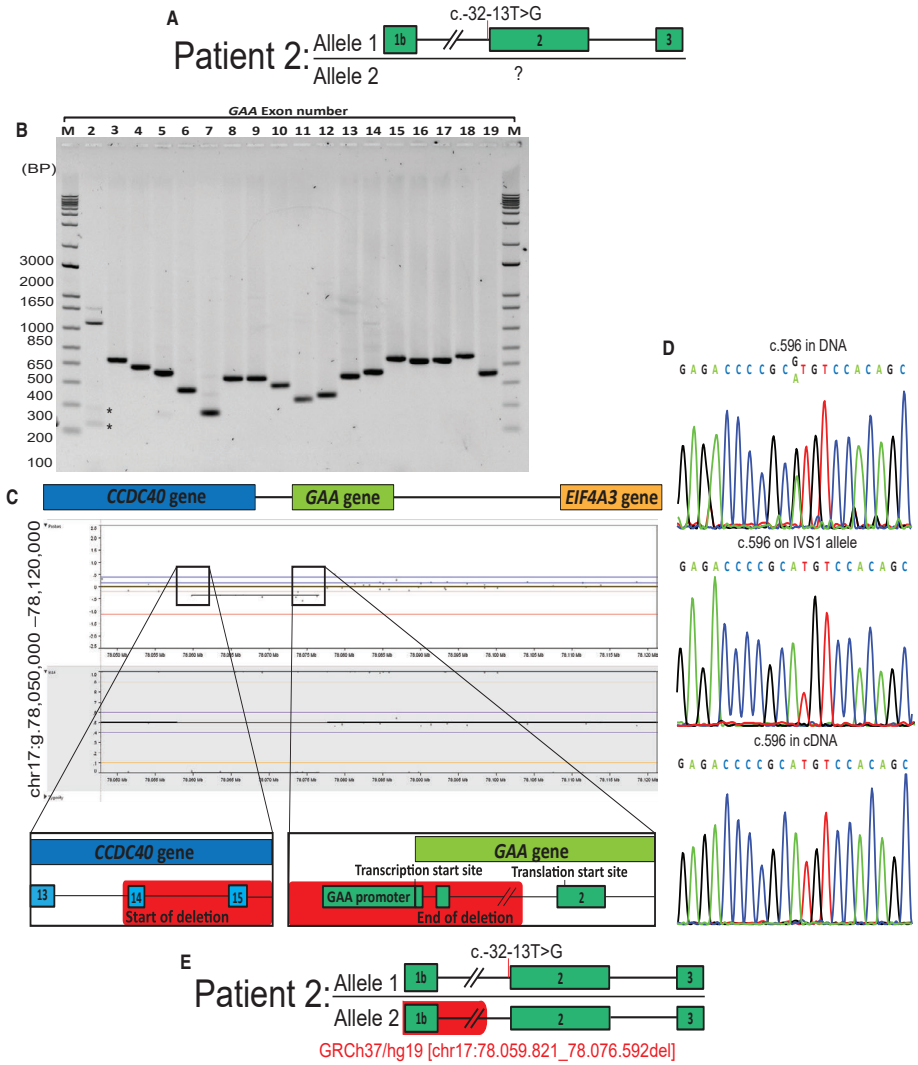
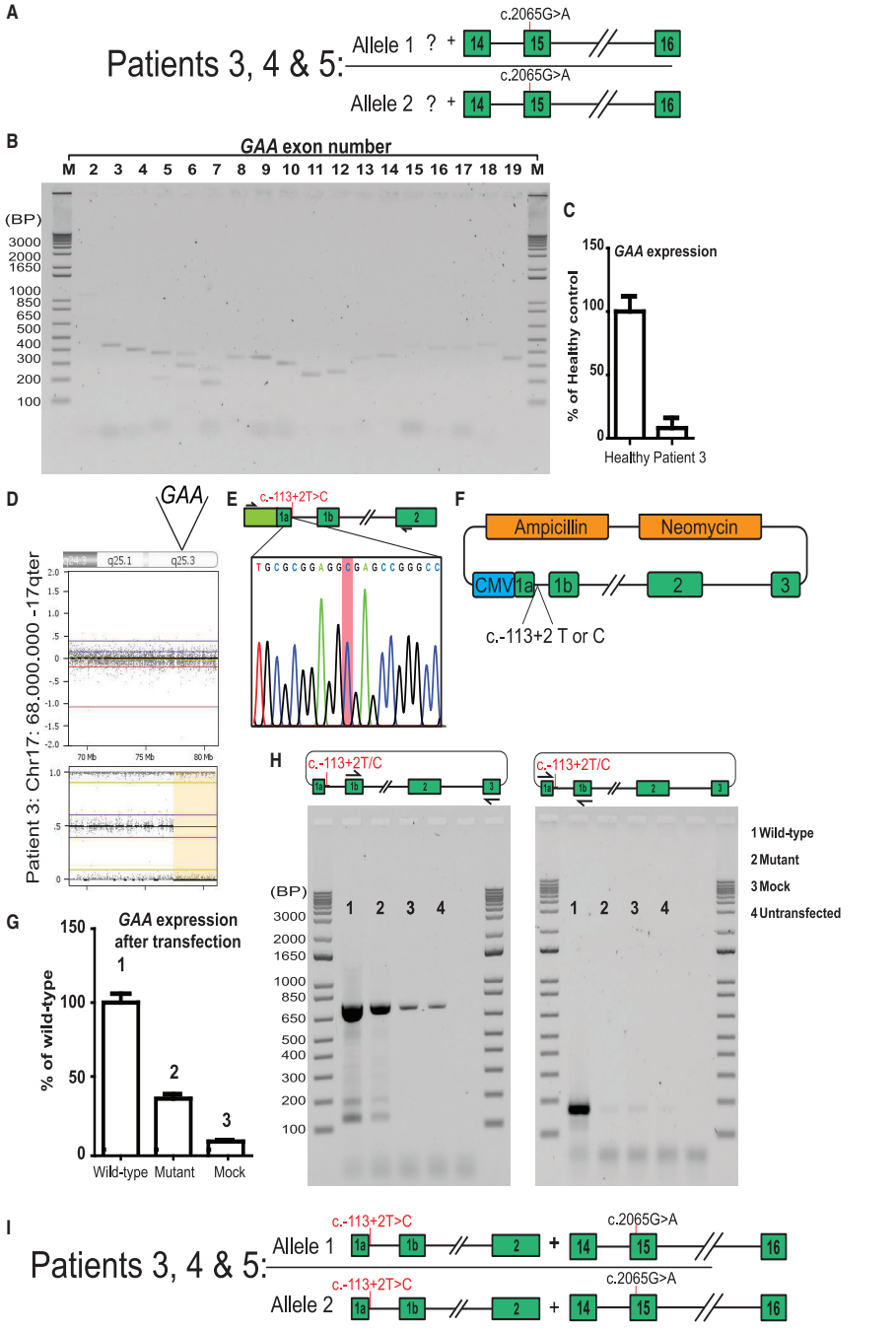


Figure 2. Identification of a *GAA* Promoter Deletion in Patient 2 Using a SNP Array

(A) Genotype of patient 2 at the start of analysis. (B) Splicing assay for *GAA* exons 2–19 performed on cDNA obtained from fibroblasts. *Known mis-spliced transcripts caused by the IVS1 variant. (C) Cartoon of the genomic region spanning the *CCDC40*, *GAA*, and *EIF4A3* genes in the chromosome (top) and the SNP array analysis (bottom) visualized for Chr17:78,050,000–78,120,000. Zooms are also shown on the 5′ and 3′ ends of the 17-kb deletion. (D) Sanger sequencing results around the c.596A>G SNP using a non allele-specific PCR of DNA (top), an IVS1 allele-specific PCR of DNA (middle), and a non allele-specific PCR of cDNA obtained from fibroblasts (bottom). (E) Cartoon of both disease-associated variants identified in patient 2.

from the IVS1 allele. To address this, the sequence of exon 3 in the patient's DNA and mRNA was analyzed for allele-specific SNPs (Figure 2D). We identified the c.596A>G SNP (rs1042393, minor allele frequency [MAF] in Dutch population: 75%) in a heterozygous state in the patient's DNA.[22] With the use of an IVS1-specific PCR (Figure S1), we found the c.596A>G SNP not to be present on the allele containing the IVS1 variant. Sequence analysis of the cDNA failed to detect transcripts that contained the c.596A>G SNP, indicating that the allele containing the deletion is not expressed.



(legend on next page)

In conclusion, the 17kb deletion, including the promoter and TSS, completely abrogates *GAA* expression, and the combination with the late-onset IVS1 variant on the other allele (Figure 2E) explains the late-onset phenotype presented by the patient.

Patients 3, 4, and 5

Patients 3, 4, and 5 are siblings, all of whom showed progressive muscle weakness later in life. Primary fibroblasts were only available for patient 3. GAA activity in lymphocytes was deficient in all three siblings; this was confirmed in the primary fibroblasts of patient 3 (Table 1). Sanger sequence analysis revealed the 2065G>A variant in a homozygous state for all three siblings but was unable to identify any disease-associated variants (Figure 3A). Previous work has shown that the c.2065G>A variant reduces GAA activity to 54% in a *GAA* cDNA expression construct.[23] This reduction is not sufficient for a patient to present symptoms associated with Pompe disease but is known to lead to a pseudodeficiency of GAA.[23]

Exon-flanking RT-PCR of cDNA derived from fibroblasts of patient 3 did not detect any aberrant splicing events but showed an overall low expression of *GAA* (Figure 3B). Quantitative real-time PCR (qRT-PCR) analysis confirmed this, showing a reduction of *GAA* expression to 8% of healthy control values (Figure 3C). SNP array analysis showed several large RoHs in all siblings that were likely derived from regions that were homologous in both parents, suggesting consanguinity between the parents. Interestingly, all three siblings were found to have a RoH ranging from 17q25.3 to the 17qter (including *GAA*) but this was not present in the paternal DNA (Figure 3D; Figure S2). The consanguinity of the parents explains the presence of the multiple RoHs in the three siblings via a combination of several identical blocks. However, this finding does not explain the decreased expression of *GAA* mRNA, and we hypothesized the presence of a variant in the 5′ UTR or the promoter. To this end, Sanger sequencing was performed for the non-coding first exons and the *GAA* promoter (Figure 3E). Sanger sequence analysis of this region revealed the presence of the homozygous variant c.-113+2T>C in patient 3. This variant is located close to the splice acceptor of exon 1A, which is, according to recent annotations, the 5′ part of exon 1. Our unpublished RNA sequencing (RNA-seq) data indicate that in agreement with the recent annotations, exon 1 consists of two small exons (exon 1A and exon 1B) that are spliced in cells from healthy individuals, as well as patients with Pompe disease. The c.-113+2T>C variant was also found in a homozygous state in the two siblings (patients 4 and 5) and in a heterozygous state in paternal DNA. This variant has not been described previously. DNA from the mother was unavailable.

To investigate the pathogenic nature of c.-113+2T>C, a previously published minigene model was modified to include the genomic region of exon 1A.[12] Two new constructs were generated, one wild-type and one carrying the c.-113+2T>C variant (Figure 3F). qRT-PCR of cDNA obtained from transfected HEK293T cells using primers located in exon 2 showed that c.-113+2T>C reduced *GAA* expression to 35% of the wild-type minigene (Figure 3G). Exon-flanking PCR for exons 1 to 3 showed that c.-113+2T>C lowered expression of mRNA containing exon 1B-exon 3 (Figure 3H, left panel). An additional exon-flanking RT-PCR was performed to detect aberrant splicing events between exons 1A and 1B (Figure 3H). The wild-type construct spliced both exons correctly, whereas the mutant completely abolished expression of the mRNA transcript containing the exon 1A-exon 1B splice junction (Figure 3H). We note that we were unable to test these splicing events in patient fibroblasts due to the low expression of GAA mentioned previously.

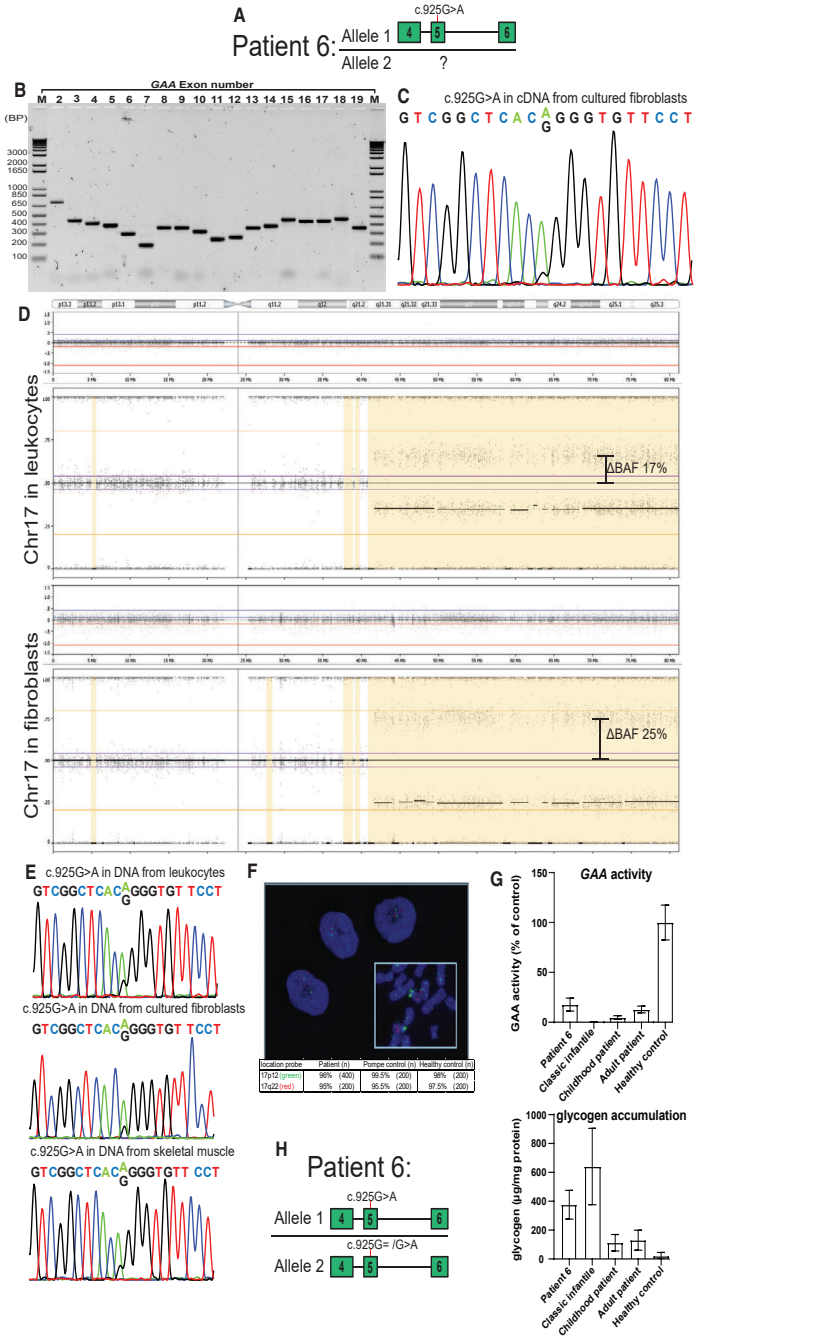
In summary, the c.-113+2T>C variant decreases expression of *GAA*. We note that the reduction in the minigene was less strong (65%) compared to the reduction in fibroblasts (92%). This difference might be due to promoter-dependent effects of the c.-113+2T>C variant, as in the patient's fibroblasts, mRNA expression is regulated by the endogenous *GAA* promoter, whereas expression from the minigene model is regulated by the cytomegalovirus (CMV) promoter. Additionally, all three patients carry both the c.-113+2T>C and c.2065G>A variants at a homozygous state (Figure 3I). The c.2065G>A missense variant is known to decrease activity to ~50%.[23] The combined action of both variants likely reduces endogenous GAA activity to pathogenic levels.

Patient 6

Indications for a superior vena cava syndrome were found during a routine prenatal ultrasound. Further cardiac screening was performed at the age of 2.5 months and revealed a left ventricular hypertrophy (LVH) with cardiac ultrasound, whereas electrocardiography showed a shortened PQ interval time without signs of pre-excitation (Table 2). At this time, the patient was normotonic with good psychomotor development and showed no signs of muscle weakness. Exome screening of cardiac genes was negative, and a metabolic examination of plasma and urine showed mildly elevated creatine kinase (CK) and transaminase levels. GAA activity was found to be only slightly

Figure 3. Identification of a RoH and a Novel Splicing Variant in Exon 1A in Patients 3–5

(A) Genotype of patient 3 at the start of analysis. (B) Splicing assay for *GAA* exons 2–19 performed on cDNA obtained from fibroblasts from patient 3. (C) qRT-PCR results for *GAA* performed on cDNA obtained from fibroblasts, normalized for *GAPDH*. Significance comparing healthy control and patient 3: $p < 0.001$. (D) SNP array result visualized for a segment of chromosome 17 showing absence of deletions or duplications (top) and a RoH from 17q25.3 to 17qter (bottom). (E) Sanger sequencing of exon 1, demonstrating the presence of the c.-113+2T>C variant. (F) Cartoon of the generated constructs ranging from *GAA* exon 1A to exon 3, with the c.-113+2T>C variant indicated. (G) qRT-PCR of *GAA* exon 2 performed on cDNA obtained from HEK293T cells after transfection with the minigene model. Results are normalized for neomycin. All comparisons were significant: $p < 0.001$. (H) RT-PCRs performed on cDNA obtained from transfected HEK293T. RT-PCR ranging from exon 1b to exon 3 (left) and RT-PCR for the exon1a–1b splice junction (right). With an accompanying cartoon for the RT-PCRs, the location of the primers are indicated using arrows. (I) Cartoon of the disease-associated variant c.-113+2T>C and the pseudodeficiency variant c.2065G>A, both at a homozygous state in patient 3. Error bars in C and G indicate standard deviation (S.D.)



(legend on next page)

Table 2. Clinical and Molecular Information Accompanying Patient 6

Patient Information		
Age at analysis	2.5 months to 4 years	
Gender	male	
	GAA Activity	Patient Range
Fibroblasts	8.1 nmol/h/mg	0–20 nmol/h/mg
Leukocytes	26 nmol/h/mg	0–10 nmol/h/mg
Clinical Description		
Left ventricular hypertrophy detected		
No signs of respiratory insufficiency		
No muscle weakness		
Cardiac Muscle Measurements		
Left ventricular mass	122.1 g (Z score +3.7)	
Shortening fraction	37.1% (Z score +0.14)	
CK levels	243 U/L (ref < 230 U/L)	
Disease-Associated Variants		
Allele 1	c.925G>A	
Allele 2	no variant found	
Parental Disease-Associated Variants		
Father (Asymptomatic)		
Allele 1	c.925G>A	
Allele 2	no variant found	
Mother (Asymptomatic)		
Allele 1	no variant found	
Allele 2	no variant found	
SNP Array Result		
Segmental abnormality range: Chr17:41.681.527–Chr17qter		

decreased in leukocytes, whereas in primary fibroblasts, enzymatic activity was within the range of late-onset Pompe disease (Table 2).[17] Histochemical analysis of a muscle biopsy showed normal morphology, but some enlarged lysosomes were visible using acid phosphatase staining (Figure S3). In addition, periodic acid-Schiff (PAS) staining showed minor glycogen accumulation (Figure S3). These findings suggested that the patient might have late-onset Pompe disease,[24] but there were two inconsistencies: the lack of GAA deficiency in lymphocytes and the LVH, a characteristic of the classic infantile phenotype but not late-onset Pompe disease. Sanger sequencing revealed only the disease-associated variant c.925G>A at a heterozygous state (Figure 4A). This variant is associated with the classic infantile phenotype when combined with a null allele.[25,26] Sanger sequence analysis of the parents (both showed no signs of Pompe disease) found the c.925G>A only in a heterozygous state in paternal DNA. No other disease-associated variants were identified in the patient nor in his parents (Table 2). Exon-flanking RT-PCR of cDNA from primary fibroblasts of the patient did not detect any splicing aberrations (Figure 4B). Sequence analysis of the *GAA* mRNA transcripts showed that the majority of transcripts contained the c.925G>A variant (Figure 4C). This indicated that the second *GAA* allele was present but was expressed at reduced levels. SNP array analysis identified a 39.5-Mb allelic imbalance, ranging from 17q21.31 to the 17qter (Figure 4D).[17] The imbalance showed no change in CNV but a change in chromosomal equilibrium (i.e., a change in the B allele frequency of heterozygous SNPs) for SNPs located in the affected segment.[17] Within the imbalanced segment, *GAA* was present, whereas all other genes present within this region were unlikely to be the cause of the presented phenotype (Table S1).[27] When analyzing the SNP array data, it became clear that the allelic imbalance was more profound in fibroblasts compared to leukocytes. We estimated a deviation in B allele frequency (ΔBAF) of ±25% in fibroblasts compared to ±17% in leukocytes (Figure 4D). This was consistent with the GAA activity, which showed that activity in fibroblasts was lower compared to leukocytes (Table 2).

Additional Sanger sequence analysis of the patient's tissues showed a skewed contribution in favor of the c.925A variant in primary fibroblasts, leukocytes, and also in skeletal muscle tissue (Figure 4E). Microsatellite analysis of chromosome 17 showed a skewed composition in favor of the paternal allele in the affected segment, whereas markers located in the unaffected segment were distributed evenly (Figure S4). Fluorescence *in situ* hybridization (FISH), using probes located on 17p12 and 17q22 to detect both ends of chromosome 17, showed a normal karyotype in the patient's fibroblasts (Figure 4F). Both ends of the chromosome were present at equal numbers, confirming that this imbalance is not the result of a change in genomic copy number.

The results are in line with a mosaic segmental UPD, in which affected cells contain two copies of the paternal *GAA* that harbor the c.925G>A variant. The patient's tissue is comprised of two distinct cell populations: one with the c.925G>A variant in a homozygous state, resulting in a classic infantile Pompe genotype, and one with

Figure 4. Identification of a Mosaic Segmental UPD in Patient 6

(A) Genotype of patient 6 at the start of analysis. (B) Splicing assay for *GAA* exons 2–19 performed on cDNA obtained from fibroblasts. (C) Sanger sequencing around the c.925G>A variant performed on cDNA obtained from fibroblasts. (D) SNP array analysis of chromosome 17 performed on DNA obtained from leukocytes (top) or fibroblasts (bottom). ΔBAF indicates the estimated difference in BAF between the affected and unaffected segment. (E) Sanger sequencing around the 925G>A variant performed on DNA obtained from leukocytes (top), fibroblasts (middle), or skeletal muscle tissue (bottom). (F) FISH analysis performed in fibroblasts showing probes located on 17p12 (green) and 17q22 (red) and 4′,6-diamidino-2-phenylindole (DAPI) staining to highlight the cell nucleus. A representative image is shown, and a quantification is given below the figure. (G, top) Biochemical quantification of GAA activity measured using 4MU as a substrate. Significance comparing patient 6 and all other samples: not significant. Significance comparing classic infantile and healthy control: $p = 0.01$. (G, bottom) Measurements of intracellular glycogen. Significance comparing classic infantile versus the childhood, adult, and healthy control samples: $p < 0.01$. Comparison of patient 6 and all other samples: not significant (patient 6 versus healthy control: $p = 0.052$). Both assays were performed on the same protein extracts obtained from fibroblasts after culturing to confluence for 3 weeks, followed by a 24h starvation to clear intracellular glycogen. GAA activity is shown as a percentage of the healthy control fibroblasts. (H) Cartoon of both disease-associated variants identified in patient 6. Error bars in G indicate S.D.

Table 3. Novel Variants Identified with Proof of Pathogenicity

	Disease-Associated Variant on Allele 1	Disease-Associated Variant/Event on Allele 2	Proof of Pathogenicity for Novel Variant
Patient 1	c.2331+2T>A	c.1327-61_1437+171del[a]	resulting deletion in the transcript is known to cause a deficiency of *GAA*
Patient 2	c.-32-13T>G	17 kb deletion, including promoter and transcription start site of *GAA*[a]	no RNA expression originating from the allele containing the deletion
Patients 3, 4, and 5	c.-113+2T>C[a] + c.2065G>A[b]	c.-113+2T>C[a] + c.2065G>A[b]	expression and splicing analysis of variant in minigene model
Patient 6	c.925G>A	mosaicism c.925G>A[a]	SNP arrays of different tissues, Sanger sequencing, glycogen accumulation assay

[a]Novel variants.
[b]Pseudodeficiency variant.

the c.925G>A variant in a heterozygous state, which results in a healthy cell population. To confirm this further, two protein-based assays were performed to examine the biochemical properties of the patient's cells. Figure 4G shows the results of the glycogen accumulation and GAA activity assays, which were performed on the same cellular extract. Patient fibroblasts accumulated intracellular glycogen, a characteristic that normally only occurs *in vitro* in fibroblasts derived from classic infantile Pompe patients but not in late-onset patients. However, GAA enzymatic activity (Figure 4G) was present and in the range of late-onset Pompe disease. This is in agreement with a genetic mosaicism with the sample consisting of two cell populations: one that accumulates glycogen *in vitro* and one that still possesses enzymatic GAA activity and thereby does not accumulate glycogen.

In summary, the atypical phenotype for Pompe disease presented by patient 6 can be explained by a genetic mosaicism for a segment of chromosome 17 that includes *GAA*. It is likely that in cardiac cells, the contribution of homozygous cells to heart tissue is relatively large, causing a LVH. In skeletal muscle cells, the mixture of homozygous and heterozygous nuclei in muscle fibers likely causes only mild pathology.

DISCUSSION

Conventional diagnostic procedures are able to identify and diagnose most individuals with Pompe disease; however, a number of cases remain unexplained.[4] Here, we emphasize the need to introduce new methods in the diagnostic validation of Pompe disease in cases for which conventional procedures prove insufficient. With the use of extended diagnostic analyses, we identified three new variants in the *GAA* gene. With the use of minor adjustments to already-existing assays, we proved the pathogenicity of the variants and completed the diagnosis for six cases of Pompe disease derived from four families (Table 3).

In previous work, the exon-flanking RT-PCR analysis proved useful to identify novel variants located outside the coding regions, for example, as shown by the identification of c.2190-345A>G.[9] Here, we showed this assay to be capable of detecting novel deletions and revealing differential gene expression as a result of *GAA* variants. However, the 17kb deletion in patient 2 and thereby other deletions spanning more than one exon remained undetected, and additional assays were required. To detect the deletion of exon 18, which is a common variant in Caucasian patients with Pompe disease, our diagnostic center performs a separate PCR reaction.[28] However, other large deletions are not routinely analyzed and will therefore not be identified. This is supported by the under-representation of gross deletions and other types of variants, such as promoter variants in *GAA*.[10,29,30] As described in the Human Gene Mutation Database (HGMD), gross deletions represent 7.5% of all disease-associated variants, whereas approximately 1% of all disease-associated variants are reported to be located in promoter regions genome wide.[29] For Pompe disease, gross deletions constitute only 1.5% of known variants, and the variants described here in patients 2 and 3 represent the first reported disease-associated variants located in the promoter and 5′ UTR regions of *GAA*.[10] These regions are critical for the regulation of gene expression, and we recommend these regions to be included in future diagnostic analyses. The detection of non-coding variants can be improved with the implementation of techniques already established in many diagnostic laboratories as a follow-up to DNA sequencing. SNP array analysis allows for the detection of large deletions or duplications, whereas aberrant splicing events or severe reductions in gene expression can be detected using whole transcriptome sequencing.[18,31] Other existing techniques can also be implemented for specific monogenic disorders and are capable of detecting smaller aberrations. Multiplex ligation-dependent probe amplification (MLPA) can detect deletions affecting single exons, whereas exon-flanking RT-PCR can be performed to detect deletions, splicing aberrations, and differential gene expression.[9,18] A more systematic approach, as described above, could lead to the identification of new disease-associated variants and provide a complete DNA diagnosis for patients. The implementation of the approaches outlined here should therefore be considered to improve the genetic analysis of cases with an unexplained genetic diagnosis.

Patients with an atypical phenotype often represent a challenge when diagnosing genetic disorders. Patient 6 presented a unique pathological condition: LVH, a characteristic usually observed only in the classic infantile phenotype, while lacking any sign of muscle weakness after 4 years of age. This phenotype was corroborated by the differences in

GAA activity in the patient's tissues. With the use of the SNP array analysis, we detected a chromosomal imbalance on chromosome 17, which was crucial to identify the genetic mosaicism of Pompe disease. Genetic mosaicisms are often associated with a milder symptomatology due to the presence of a healthy or less affected cell population.[32]

The mosaicism in patient 6 appears to be the result of a segmental UPD. UPDs have been associated with many inherited disorders and appropriately explain a subset of patients carrying homozygous disease-associated variants.[33] Whole-chromosome UPDs are reported to occur in 1/3,500 births,[33] whereas segmental UPDs occur less frequently and are the result of somatic recombination.[34–36] The mosaic nature of the segmental UPD and the high percentage of affected cells in several of the patient tissues suggest that the recombination event occurred early in the postzygotic stage of embryonal development.[34–36] Tissues of this patient were thereafter comprised of two cell populations, with the c.925G>A disease-associated variant being present at either a heterozygous (healthy) or homozygous (classic infantile) state.

The presence of a population of healthy cells in the patient did not prevent the development of LVH, which normally only occurs in classic infantile patients. However, at 4 years of age, the patient has not yet manifested any other symptoms related to Pompe disease. The lack of muscle weakness in this patient was reflected by mild pathology in his muscle biopsies. This is likely due to the multinucleation of muscle fibers, whereby the genetic mosaicism in this patient is offset by the presence of both affected and unaffected nuclei in individual muscle fibers, which could be sufficient to prevent muscle pathology. A comprehensive follow-up of this patient, including CNS magnetic resonance imaging (MRI) and neuropsychological investigations, will be required to rule out additional symptoms, such as white matter changes in the brain, which can manifest in classic infantile Pompe patients.[37,38]

This study shows the need for the implementation of additional diagnostic assays and research as a follow-up when conventional procedures prove insufficient or when a patient's symptoms or biochemical characteristics do not match to a disorder. For monogenic disorders, such as Pompe disease, more specific assays can be implemented for an improved analysis of a singular gene. The implementation of this follow-up will likely result in the identification of novel variants and provide a complete DNA diagnosis for patients carrying non-coding variants.

MATERIALS AND METHODS

Patients

The patients described here were selected for their clinical and/or biochemical diagnosis that could be attributed to a deficiency of GAA. Informed consents were obtained to perform extended diagnostic assays to identify *GAA* variants and explain the presented phenotypes. Analysis of the patients analyzed in this study was approved by the medical ethics committee of the Erasmus MC. Primary fibroblasts were not available for patient 4 and 5, assays were performed for patient 3, and findings were thereafter confirmed in DNA from patients 4 and 5.

DNA Analysis

GenBank: NM_001079803.2 and NM_0001079803.2 were used as reference sequences for *GAA* DNA and mRNA, respectively, where c.1 represents the first nucleotide of the translation start codon ATG. NP000143.2 was used as a reference for GAA protein. We note that in previous annotations, exon 1 comprised 334 nt, whereas in current annotations, this region is divided into two exons (exon 1A and 1B) and a 185-nt intron. All SNP arrays analyses were performed using the Illumina Infinium CytoSNP-850K BeadChip platform. Analysis of genes related to genetic disorders located within specified genomic regions was performed using Genomic Oligoarray and SNP Array Evaluation Tool v3.0.

Histology and Imaging

Acid phosphatase staining was performed on cryosections, as described.[39] PAS staining was performed on sections of glycolmethacrylate (GMA)-processed tissue.[24] Hematoxylin and eosin (H&E) stainings were performed on both GMA fixated and cryosections. A Hamamatsu NanoZoomer 2.0 (Hamamatsu Photonics) was used for imaging, and images were analyzed with NDP.view software.

RNA Isolation and cDNA Synthesis

Total RNA was isolated from cultured fibroblasts using the RNeasy minikit (QIAGEN), including a DNaseI (QIAGEN) treatment. RT-PCR was performed using the iScript cDNA synthesis kit (Bio-Rad) in reactions with 300–800 ng RNA input.

PCR and Sanger Sequencing

RT-PCR and qRT-PCR were performed as described.[9] Sanger sequencing of PCR products was performed using a ABI3730XL DNA analyzer (Thermo Fisher Scientific) and analyzed using ApE software.

Generation of the Minigene Constructs and Transfection

The minigene model described in Bergsma et al.[12] was modified as follows: with the use of a primer with a XhoI overhang and a SBF1 restriction site, the exon 1a region was amplified and cloned in the construct. We used the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies) to introduce the c.-113+2T>C variant. Transfection of the minigene constructs was performed in HEK293T cells using Lipofectamine 2000 (Invitrogen). RNA was harvested 48 h after transfection.

Enzyme Activity and Glycogen Assay

GAA activity was measured in leukocytes or fibroblasts using 4MU, as described.[9] All 4MU assays were performed in diagnostic settings, with the exception of Figure 4G. For this experiment, fibroblasts had been confluent for 3 weeks and starved in glucose-free medium, 24 h before harvest, after which, GAA activity was measured using the

4MU.[9] Intracellular glycogen was measured with a two-step protocol using amylase and glucose-oxidase.[9]

Fragment Analysis of Microsatellite Markers

15 probes were selected for their location on chromosome 17. Analysis of the VIC- or FAM-labeled products was performed using an ABI3730XL DNA analyzer (Thermo Fisher Scientific).

Fluorescence *In Situ* Hybridization Analysis

FISH was performed on cultured fibroblasts at a low passage number. Two probes located on 17p12 and 17q22 were used to visualize both ends of the chromosome. The cells were not confluent when fixed and classified as having a normal karyotype when two or four copies of each probe were visible.

Statistics

Data were analyzed using IBM SPSS statistics, version 26. For all experiments, normal distribution of data was determined based on calculated residuals. Significance between normally distributed data from two groups was tested using an unpaired two-tailed t test. For experiments with three or more groups, a one-way ANOVA of independent samples with Tukey honestly significant difference (HSD) or Games-Howell post hoc multiple correction (depending on homogeneity of the variance) was performed. Non-normal distributed data were statistically tested using the Kruskal-Wallis method of independent samples with Bonferroni multiple comparison for three or more groups. A p value of less than 0.05 was considered significant for all tests.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.omtm.2019.12.016.

AUTHOR CONTRIBUTIONS

Conceptualization, S.L.M.i.G., A.T.v.d.P., A.J.B., and W.W.M.P.P.; Methodology, S.L.M.i.G., D.O.S.d.F., and A.J.B.; Validation Verification, M.-A.B.R., and P.W.; Formal Analysis, S.L.M.i.G., A.I., and J.J.S.; Investigation, S.L.M.i.G., D.O.S.d.F., H.D., T.D., and G.M.S.-B.; Resources, T.D., P.W., and M.-A.B.R.; Data Curation, H.D., J.J.S., and A.d.K.; Writing – Original Draft, S.L.M.i.G. and A.I.; Writing – Review & Editing, all authors; Supervision Project Administration, A.J.B., J.M.P.v.d.H., L.H.H., and W.W.M.P.P.; Funding Acquisition, A.T.v.d.P. and W.W.M.P.P.

CONFLICTS OF INTEREST

A.T.v.d.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. All other authors declare no competing interests.

ACKNOWLEDGMENTS

We thank the members of the Molecular Stem Cell Biology Group for the critical discussions. This work was funded through Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; Brazil), the Sophia Foundation for Medical Research (SSWO; project number s17-32), and Metakids (project number 2016 - 063). This project has received funding from the Ministry of Economic Affairs under TKI allowance under the TKI-Programme Life Sciences & Health. The collaboration project is co-funded by the PPP Allowance made available by Health ~ Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships (project numbers LSHM16008 and LSHM19015). The collaboration project is co-initiated by the Prinses Beatrix Spierfonds.

REFERENCES

1. Güngör, D., and Reuser, A.J. (2013). How to describe the clinical spectrum in Pompe disease? Am. J. Med. Genet. A. *161A*, 399–400.
2. Prater, S.N., Patel, T.T., Buckley, A.F., Mandel, H., Vlodavski, E., Banugaria, S.G., Feeney, E.J., Raben, N., and Kishnani, P.S. (2013). Skeletal muscle pathology of infantile Pompe disease during long-term enzyme replacement therapy. Orphanet J. Rare Dis. *8*, 90.
3. 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.
4. van der Ploeg, A.T., Kruijshaar, M.E., Toscano, A., Laforêt, P., Angelini, C., Lachmann, R.H., Pascual Pascual, S.I., Roberts, M., Rösler, K., Stulnig, T., et al.; European Pompe Consortium (2017). European consensus for starting and stopping enzyme replacement therapy in adult patients with Pompe disease: a 10-year experience. Eur. J. Neurol. *24*, 768–e31.
5. Goldstein, J.L., Young, S.P., Changela, M., Dickerson, G.H., Zhang, H., Dai, J., Peterson, D., Millington, D.S., Kishnani, P.S., and Bali, D.S. (2009). Screening for Pompe disease using a rapid dried blood spot method: experience of a clinical diagnostic laboratory. Muscle Nerve *40*, 32–36.
6. Burton, B.K., Kronn, D.F., Hwu, W.L., and Kishnani, P.S.; Pompe Disease Newborn Screening Working Group (2017). The Initial Evaluation of Patients After Positive Newborn Screening: Recommended Algorithms Leading to a Confirmed Diagnosis of Pompe Disease. Pediatrics *140* (*Suppl 1*), S14–S23.
7. Labrousse, P., Chien, Y.H., Pomponio, R.J., Keutzer, J., Lee, N.C., Akmaev, V.R., Scholl, T., and Hwu, W.L. (2010). Genetic heterozygosity and pseudodeficiency in the Pompe disease newborn screening pilot program. Mol. Genet. Metab. *99*, 379–383.
8. Kishnani, P.S., Steiner, R.D., Bali, D., Berger, K., Byrne, B.J., Case, L.E., Crowley, J.F., Downs, S., Howell, R.R., Kravitz, R.M., et al. (2006). Pompe disease diagnosis and management guideline. Genet. Med. *8*, 267–288.
9. 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.
10. Niño, M.Y., In 't Groen, S.L.M., Bergsma, A.J., van der Beek, N.A.M.E., Kroos, M., Hoogeveen-Westerveld, M., van der Ploeg, A.T., and Pijnappel, W.W.M.P. (2019). Extension of the Pompe mutation database by linking disease-associated variants to clinical severity. Hum. Mutat. *40*, 1954–1967.
11. Bergsma, A.J., In 't Groen, S.L.M., van den Dorpel, J.J.A., van den Hout, H.J.M.P., van der Beek, N.A.M.E., Schoser, B., Toscano, A., Musumeci, O., Bembi, B., Dardis, A., et al. (2019). A genetic modifier of symptom onset in Pompe disease. EBioMedicine *43*, 553–561.
12. Bergsma, A.J., In 't Groen, S.L., Verheijen, F.W., van der Ploeg, A.T., and Pijnappel, W.W.M.P. (2016). From Cryptic Toward Canonical Pre-mRNA Splicing in Pompe Disease: a Pipeline for the Development of Antisense Oligonucleotides. Mol. Ther. Nucleic Acids *5*, e361.
13. van der Wal, E., Bergsma, A.J., van Gestel, T.J.M., In 't Groen, S.L.M., Zaehres, H., Araúzo-Bravo, M.J., Schöler, H.R., van der Ploeg, A.T., and Pijnappel, W.W.M.P. (2017). GAA Deficiency in Pompe Disease Is Alleviated by Exon Inclusion in iPSC-Derived Skeletal Muscle Cells. Mol. Ther. Nucleic Acids *7*, 101–115.

14. van der Wal, E., Bergsma, A.J., Pijnenburg, J.M., van der Ploeg, A.T., and Pijnappel, W.W.M.P. (2017). Antisense Oligonucleotides Promote Exon Inclusion and Correct the Common c.-32-13T>G GAA Splicing Variant in Pompe Disease. Mol. Ther. Nucleic Acids *7*, 90–100.
15. Bergsma, A.J., van der Wal, E., Broeders, M., van der Ploeg, A.T., and Pim Pijnappel, W.W.M. (2018). Alternative Splicing in Genetic Diseases: Improved Diagnosis and Novel Treatment Options. Int. Rev. Cell Mol. Biol. *335*, 85–141.
16. Eggermann, T., Soellner, L., Buiting, K., and Kotzot, D. (2015). Mosaicism and uniparental disomy in prenatal diagnosis. Trends Mol. Med. *21*, 77–87.
17. Labrijn-Marks, I., Somers-Bolman, G.M., In 't Groen, S.L.M., Hoogeveen-Westerveld, M., Kroos, M.A., Ala-Mello, S., Amaral, O., Miranda, C.S., Mavridou, I., Michelakakis, H., et al. (2019). Segmental and total uniparental isodisomy (UPiD) as a disease mechanism in autosomal recessive lysosomal disorders: evidence from SNP arrays. Eur. J. Hum. Genet. *27*, 919–927.
18. Silva, M., de Leeuw, N., Mann, K., Schuring-Blom, H., Morgan, S., Giardino, D., Rack, K., and Hastings, R. (2019). European guidelines for constitutional cytogenomic analysis. Eur. J. Hum. Genet. *27*, 1–16.
19. Bali, D.S., Goldstein, J.L., Banugaria, S., Dai, J., Mackey, J., Rehder, C., and Kishnani, P.S. (2012). Predicting cross-reactive immunological material (CRIM) status in Pompe disease using GAA mutations: lessons learned from 10 years of clinical laboratory testing experience. Am. J. Med. Genet. C. Semin. Med. Genet. *160C*, 40–49.
20. 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.
21. 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.
22. Genome of the Netherlands Consortium (2014). Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat. Genet. *46*, 818–825.
23. Kroos, M.A., Mullaart, R.A., Van Vliet, L., Pomponio, R.J., Amartino, H., Kolodny, E.H., Pastores, G.M., Wevers, R.A., Van der Ploeg, A.T., Halley, D.J., and Reuser, A.J. (2008). p.[G576S; E689K]: pathogenic combination or polymorphism in Pompe disease? Eur. J. Hum. Genet. *16*, 875–879.
24. 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.
25. Kroos, M.A., van Leenen, D., Verbiest, J., Reuser, A.J., and Hermans, M.M. (1998). Glycogen storage disease type II: identification of a dinucleotide deletion and a common missense mutation in the lysosomal alpha-glucosidase gene. Clin. Genet. *53*, 379–382.
26. Markic, J., Polic, B., Stricevic, L., Metlicic, V., Kuzmanic-Samija, R., Kovacevic, T., Ivkosic, I.E., and Mestrovic, J. (2014). Effects of immune modulation therapy in the first Croatian infant diagnosed with Pompe disease: a 3-year follow-up study. Wien. Klin. Wochenschr. *126*, 133–137.
27. Wierenga, K.J., Jiang, Z., Yang, A.C., Mulvihill, J.J., and Tsinoremas, N.F. (2013). A clinical evaluation tool for SNP arrays, especially for autosomal recessive conditions in offspring of consanguineous parents. Genet. Med. *15*, 354–360.
28. Van der Kraan, M., Kroos, M.A., Joosse, M., Bijvoet, A.G., Verbeet, M.P., Kleijer, W.J., and Reuser, A.J. (1994). Deletion of exon 18 is a frequent mutation in glycogen storage disease type II. Biochem. Biophys. Res. Commun. *203*, 1535–1541.
29. Stenson, P.D., Mort, M., Ball, E.V., Evans, K., Hayden, M., Heywood, S., Hussain, M., Phillips, A.D., and Cooper, D.N. (2017). The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum. Genet. *136*, 665–677.
30. Cooper, D.N. (2002). Human gene mutation in pathology and evolution. J. Inherit. Metab. Dis. *25*, 157–182.
31. Gonorazky, H.D., Naumenko, S., Ramani, A.K., Nelakuditi, V., Mashouri, P., Wang, P., Kao, D., Ohri, K., Viththiyapaskaran, S., Tarnopolsky, M.A., et al. (2019). Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease. Am. J. Hum. Genet. *104*, 1007.
32. Dipple, K.M., and McCabe, E.R. (2000). Phenotypes of patients with "simple" Mendelian disorders are complex traits: thresholds, modifiers, and systems dynamics. Am. J. Hum. Genet. *66*, 1729–1735.
33. Yamazawa, K., Ogata, T., and Ferguson-Smith, A.C. (2010). Uniparental disomy and human disease: an overview. Am. J. Med. Genet. C. Semin. Med. Genet. *154C*, 329–334.
34. Kotzot, D. (2008). Complex and segmental uniparental disomy updated. J. Med. Genet. *45*, 545–556.
35. Robinson, W.P. (2000). Mechanisms leading to uniparental disomy and their clinical consequences. BioEssays *22*, 452–459.
36. Kearney, H.M., Kearney, J.B., and Conlin, L.K. (2011). Diagnostic implications of excessive homozygosity detected by SNP-based microarrays: consanguinity, uniparental disomy, and recessive single-gene mutations. Clin. Lab. Med. *31*, 595–613, ix.
37. 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.
38. Ebbink, B.J., Poelman, E., Aarsen, F.K., Plug, I., Régal, L., Muentjes, C., van der Beek, N.A.M.E., Lequin, M.H., van der Ploeg, A.T., and van den Hout, J.M.P. (2018). Classic infantile Pompe patients approaching adulthood: a cohort study on consequences for the brain. Dev. Med. Child Neurol. *60*, 579–586.
39. van den Berg, L.E., de Vries, J.M., Verdijk, R.M., van der Ploeg, A.T., Reuser, A.J., and van Doorn, P.A. (2011). A case of adult Pompe disease presenting with severe fatigue and selective involvement of type 1 muscle fibers. Neuromuscul. Disord. *21*, 232–234.

OMTM, Volume *17*

Supplemental Information

Novel *GAA* Variants and Mosaicism in Pompe Disease Identified by Extended Analyses of Patients with an Incomplete DNA Diagnosis

Stijn L.M. in 't Groen, Douglas O.S. de Faria, Alessandro Iuliano, Johanna M.P. van den Hout, Hannie Douben, Trijnie Dijkhuizen, David Cassiman, Peter Witters, Miguel-Ángel Barba Romero, Annelies de Klein, Galhana M. Somers-Bolman, Jasper J. Saris, Lies H. Hoefsloot, Ans T. van der Ploeg, Atze J. Bergsma, and W.W.M. Pim Pijnappel

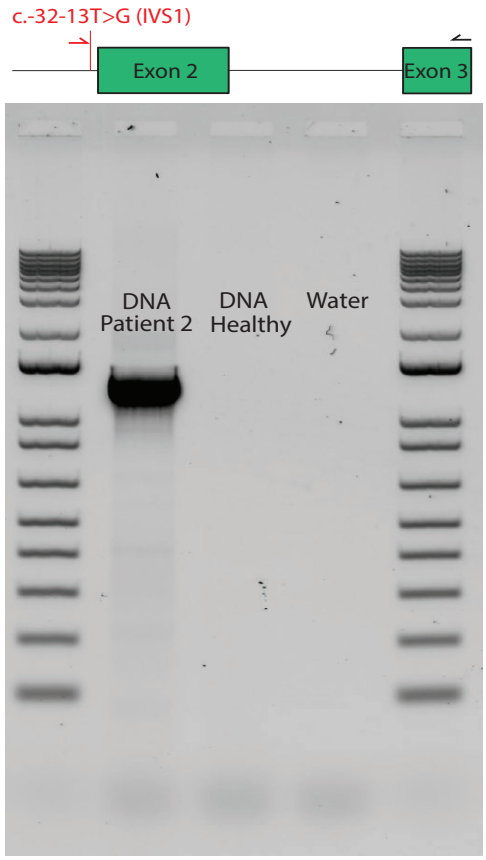


Figure S1, Chapter 3

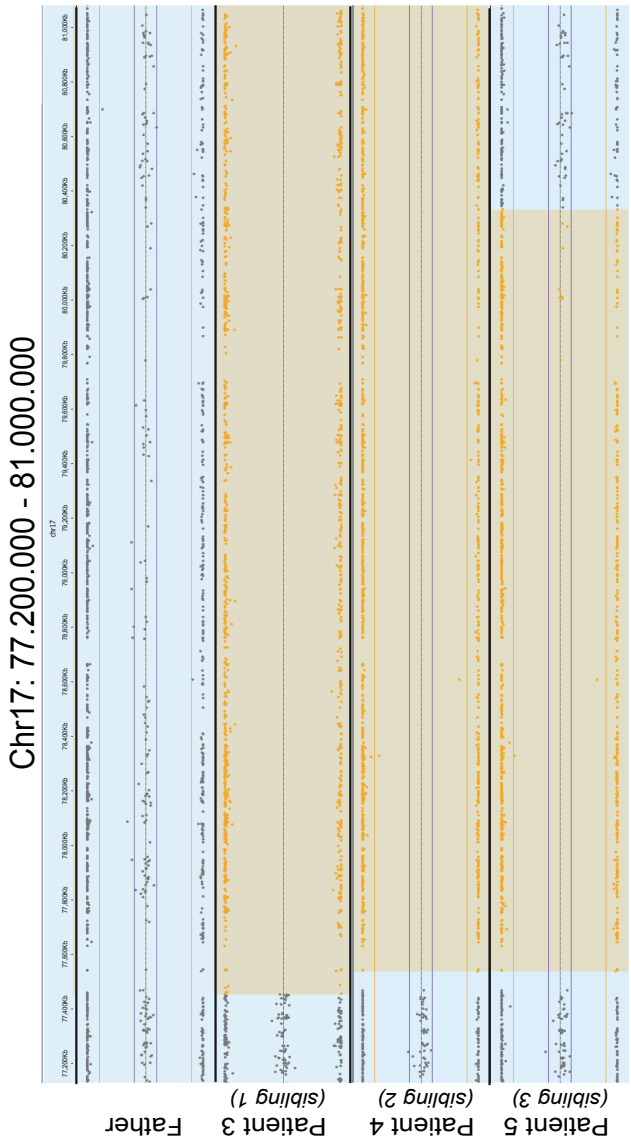


Figure S2, Chapter 3

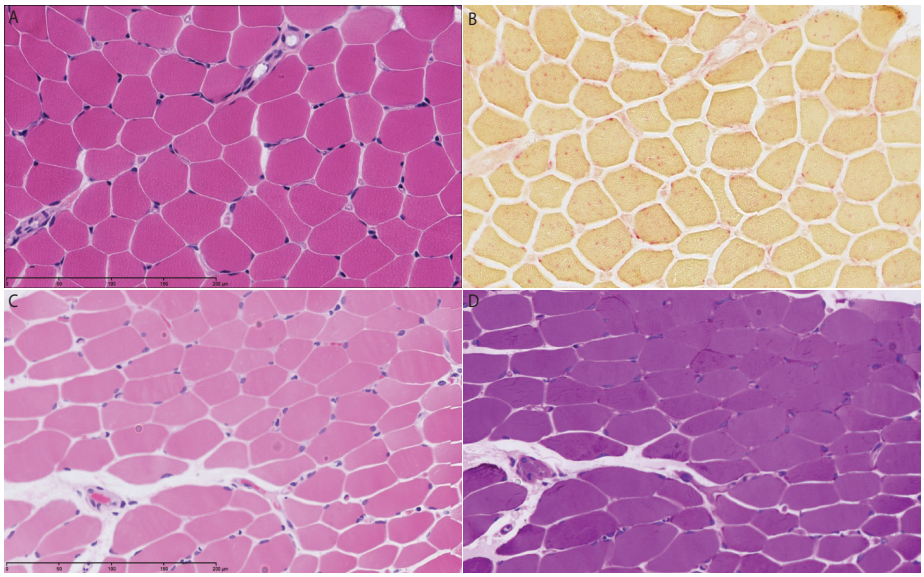


Figure S3, Chapter 3

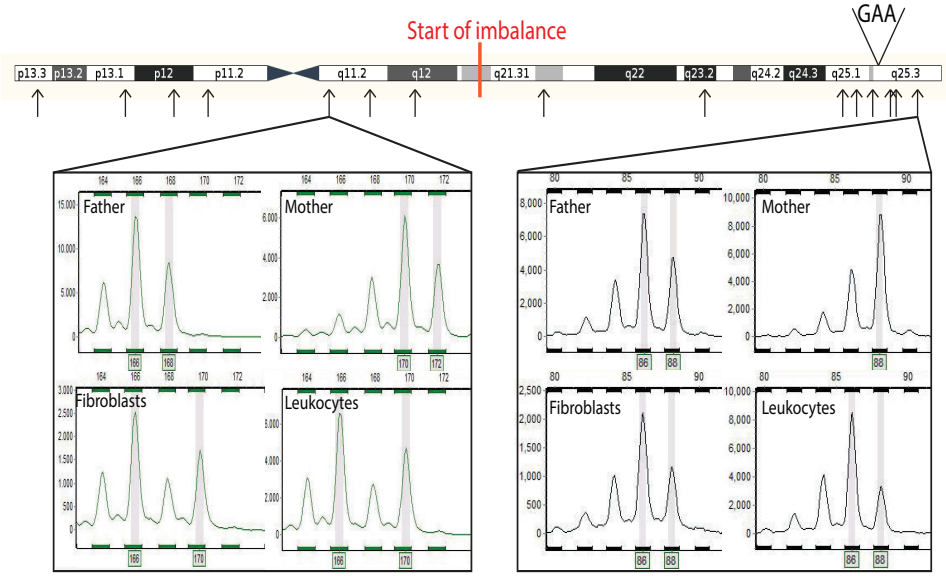


Figure S4, Chapter 3

Figure S1: IVS1-specific RT-PCR.

The forward primer was designed in intron 1b with the last base being c.-32-13G to only amplify the IVS1 allele. The reverse primer was located in exon 3. DNA was obtained from patient 2's leukocytes. A control DNA sample was obtained from a healthy individual that did not carry the IVS1 variant.

Figure S2: RoH in chromosome 17 of patients 3, 4 and 5

SNP array analysis visualized for Chr17:77.200.000-81.000.000. The parental DNA showed a normal composition, while patients 3 and 4 showed a RoH ranging from 17q25.3 to the 17qter. Patient 5 showed a RoH ranging 3Mb in 17q25.3.

Figure S3: Skeletal muscle biopsy of patient 6.

Representative images of histological stainings performed on a muscle biopsy tissue obtained from patient 6. **(A)** Heamatoxylin-eosin staining performed on a cryosection. **(B)** Acid phosphatase staining performed on a cryosection, lysosomes are visible as red dots located within the myofibers. **(C)** Heamatoxylin-eosin staining performed on a GMA fixed tissue section. **(D)** PAS staining performed on a GMA fixed tissue section. Minor glycogen accumulation is visible as dark purple patches in the tissue.

Figure S4: Microsatellite analysis of chromosome 17 in patient 6.

Microsatellite analysis for chromosome 17 of patient 6. The location of the 15 probes used are indicated with arrows. The start of the imbalance is highlighted in red and the details of two representative probes are shown: one probe is located in the unaffected segment (left) (17q11.2) and one in the imbalanced segment (right) (17q25.3).

Table S1: List of genes located in the chromosomal imbalance found in patient 6 and the recessively inherited disorders associated with these genes

#	Mapped genes	Gene Map Disorder
1 :	**GAA(17q25.2-q25.3)**	**Glycogen storage disease II**
2 :	**SGCA(17q21)**	**Muscular dystrophy, limb-girdle, type 2D**
3 :	MEOX1(17q21)	Klippel-Feil syndrome 2
4 :	SOST(17q11.2)	Craniodiaphyseal dysplasia, autosomal dominant
5 :		Van Buchem disease
6 :		Sclerosteosis 1
7 :	NAGS(17q21.31)	N-acetylglutamate synthase deficiency
8 :	G6PC3(17q21.31)	Neutropenia, severe congenital 4, autosomal recessive
9 :		Dursun syndrome
10 :	GRN(17q21.32)	Frontotemporal lobar degeneration with ubiquitin-positive inclusions
11 :		Aphasia, primary progressive
12 :		Ceroid lipofuscinosis, neuronal, 11
13 :	ITGA2B(17q21.32)ITGB3(17q21.32)	Bleeding disorder, platelet-type, 16, autosomal dominant
14 :		Glanzmann thrombasthenia
15 :		Thrombocytopenia, neonatal alloimmune, BAK antigen related
16 :	CCDC103(17q21.31)	Ciliary dyskinesia, primary, 17
17 :	MAPT(17q21.1)	Tauopathy and respiratory failure
18 :		Susceptibility to parkinson disease
19 :		Pick disease
20 :		Supranuclear palsy, progressive atypical
21 :		Dementia, frontotemporal, with or without parkinsonism
22 :		Supranuclear palsy, progressive
23 :	WNT3(17q21)	Tetra-amelia, autosomal recessive
24 :	GOSR2(17q21)	Epilepsy, progressive myoclonic 6
25 :	ITGB3(17q21.32)	Thrombocytopenia, neonatal alloimmune
26 :		Myocardial infarction, protection and/or susceptibility
27 :	TBX21(17q21.32)	Asthma and nasal polyps
28 :	PNPO(17q21.32)	Pyridoxamine 5'-phosphate oxidase deficiency
29 :	HOXB1(17q21.3)	Facial paresis, hereditary congenital, 3
30 :	XYLT2(17q21.33)	Pseudoxanthoma elasticum
31 :	DGKE(17q22)	Nephrotic syndrome, type 7
32 :	EPX(17q23.1)	Eosinophil peroxidase deficiency
33 :	MKS1(17q22)	Bardet-Biedl syndrome 1-12
34 :		Meckel syndrome 1
35 :	ACE(17q23.3) MPO(17q23.1)	Alzheimer disease
36 :	MPO(17q23.1)	Myeloperoxidase deficiency
37 :	RAD51C(17q25.1)	Fanconi anemia, complementation group O
38 :		Susceptibility to breast-ovarian cancer, familial

#	Mapped genes	Gene Map Disorder
39 :	TRIM37(17q23.2)	Mulibrey nanism
40 :	ACE(17q23.3)	Renal tubular dysgenesis
41 :		Microvascular complications of diabetes 3
42 :		Stroke, hemorrhagic
43 :	TACO1(17q23.3)	Mitochondrial complex IV deficiency
44 :	GH1(17q24.2)	Growth hormone deficiency
45 :		Kowarski syndrome
46 :	CD79B(17q23)	Agammaglobulinemia 6
47 :	PRKAR1A(17q23-q24)	Acrodysostosis 1
48 :		Carney complex, type 1
49 :		Thyroid papillary carcinoma
50 :		Adrenocortical tumor, somatic
51 :		Myxoma, intracardiac
52 :		Pigmented nodular adrenocortical disease, primary, 1
53 :	FAM20A(17q24.2)	Amelogenesis imperfecta and gingival fibromatosis syndrome
54 :	USH1G(17q25.1)	Usher syndrome, type 1G
55 :	SLC25A19(17q25.3)	Microcephaly, Amish type
56 :		Thiamine metabolism dysfunction syndrome 4
57 :	TSEN54(17q25.1)	Pontocerebellar hypoplasia
58 :	ITGB4(17q25)	Epidermolysis bullosa
59 :	GALK1(17q24)	Galactokinase deficiency with cataracts
60 :	ACOX1(17q24-q25|17q25.1)	Peroxisomal acyl-CoA oxidase deficiency
61 :	CANT1(17q25.3)	Desbuquois dysplasia,
62 :	EIF4A3(17q25.3)	Robin sequence with cleft mandible and limb anomalies
63 :	SGSH(17q25.3)	Mucopolysaccharidisis type IIIA (Sanfilippo A)
64 :	PDE6G(17q25)	Retinitis pigmentosa 57
65 :	ARHGDIA(17q25.3)	Nephrotic syndrome, type 8
66 :	PYCR1(17q25.3)	Cutis laxa, autosomal recessive, type IIB

CHAPTER 4

Update of the Pompe variant database for the prediction of clinical phenotypes: Novel disease-associated variants, common sequence variants, and results from newborn screening

Douglas O. S. de Faria*, Stijn L. M. in 't Groen*, Marianne Hoogeveen-Westerveld, Monica Y. Niño, Ans T. van der Ploeg, Atze J. Bergsma, W.W.M. Pim Pijnappel

*These authors contributed equally to this study
Hum Mutat. 2021 Feb;42(2):119-134.
doi: 10.1002/humu.24148. Epub 2020 Dec 21

Received: 21 September 2020 | Revised: 6 November 2020 | Accepted: 30 November 2020

DOI: 10.1002/humu.24148

DATABASES

Human Mutation HGVS HUMAN GENOME VARIATION SOCIETY WILEY

Update of the Pompe variant database for the prediction of clinical phenotypes: Novel disease-associated variants, common sequence variants, and results from newborn screening

Douglas O. S. de Faria[1,2,3] | Stijn L. M. in 't Groen[1,2,3] | Marianne Hoogeveen-Westerveld[2] | Monica Y. Niño[1,2,3] | Ans T. van der Ploeg[1,3] | Atze J. Bergsma[1,2,3] | W. W. M. Pim Pijnappel[1,2,3]

[1]Department of Pediatrics, Erasmus MC University Medical Center, Rotterdam, The Netherlands

[2]Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, The Netherlands

[3]Center for Lysosomal and Metabolic Diseases, Erasmus MC University Medical Center, Rotterdam, The Netherlands

Correspondence
Atze J. Bergsma and W. W. M. Pim Pijnappel, Department of Pediatrics, Erasmus MC University Medical Center, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands.
Email: a.bergsma@erasmusmc.nl (A. J. B.) and w.pijnappel@erasmusmc.nl (W. W. M. P. P.)

Abstract

Pompe disease is an inherited disorder caused by disease-associated variants in the acid α-glucosidase gene (*GAA*). The Pompe disease *GAA* variant database (http://www.pompevariantdatabase.nl) is a curated, open-source, disease-specific database, and lists disease-associated *GAA* variants, in silico predictions, and clinical phenotypes reported until 2016. Here, we provide an update to include 226 disease-associated variants that were published until 2020. We also listed 148 common *GAA* sequence variants that do not cause Pompe disease. *GAA* variants with unknown severity that were identified only in newborn screening programs were listed as a new feature to indicate the reason why phenotypes were still unknown. Expression studies were performed for common missense variants to predict their severity. The updated Pompe disease *GAA* variant database now includes 648 disease-associated variants, 26 variants from newborn screening, and 237 variants with unknown severity. Regular updates of the Pompe disease *GAA* variant database will be required to improve genetic counseling and the study of genotype–phenotype relationships.

KEYWORDS
database, disease-associated variants, *GAA*, NBS, Pompe disease, SNP

1 | INTRODUCTION

Pompe disease (glycogen storage disease type II; MIM #232300) is an autosomal recessive disorder caused by disease-associated variants in the acid α-glucosidase (*GAA*) gene, resulting in a deficiency of the GAA enzyme, accumulation of lysosomal glycogen, and progressive muscle weakness. The clinical spectrum of Pompe disease is broad (Güngör & Reuser, 2013). The most severe classic infantile phenotype presents shortly after birth with hypertrophic cardiomyopathy and generalized muscle weakness. These patients die in the first year of life due to cardiorespiratory insufficiency if left untreated. The slower progressing phenotype is characterized by

Douglas O. S. de Faria and Stijn L. M. in't Groen contributed equally to this study.

muscle weakness that can appear at any age from <1 year into adulthood. These patients are generally spared from cardiac symptoms (Kohler et al., 2018; van der Ploeg & Reuser, 2008). Enzyme replacement therapy (ERT) with intravenously applied recombinant human GAA is available since 2006. ERT normalizes hypertrophic cardiomyopathy, improves motor function, and extends survival.

The differences between phenotypes in Pompe disease can, in part, be attributed to the severity of the disease-associated variants present in the *GAA* gene. Classic infantile patients carry two disease-associated variants that completely disrupt the function of GAA (i.e., null alleles). This group of patients can be subdivided based on their cross-reactive immunological material (CRIM) status, which is defined by the disease-associated variants involved. When two *GAA* variants are present that do not result in GAA protein expression, the patient is classified as CRIM-negative. When at least one *GAA* variant gives rise to GAA protein expression (in which the GAA protein can be enzymatically inactive), the patient is classified as CRIM-positive. The clinical importance of CRIM status is highlighted by the fact that CRIM-negative classic infantile patients have a poorer prognosis compared with CRIM-positive classic infantile patients, possibly due to the formation of high sustained anti-GAA antibody titers upon treatment with ERT (Bali et al., 2012; van Gelder et al., 2015). Patients who do not have the classic infantile phenotype carry at least one disease-associated variant that allows some residual enzymatic activity. These patients are, by definition, CRIM-positive (Kroos et al., 2012b; Kulessa et al., 2020).

The "Pompe disease *GAA* variant database" (http://www.pompevariantdatabase.nl) is an open-access database that lists and classifies all reported variants in the *GAA* gene. We recently revised this database to include clinical data from patients collected from the literature, adapted the classification system for variant severity, and added (predicted) CRIM status for disease-associated variants. The database included literature up to May 2016, resulting in a total of 561 variants (Niño et al., 2019). In recent years, many new patients and *GAA* variants have been reported. These include findings from large patient populations, such as the French nationwide study (246 patients with late-onset Pompe disease) and the Pompe registry (1079 patients from 26 countries; Reuser et al., 2019; Semplicini et al., 2018).

In addition, various countries, including Taiwan, the United States, Italy, Brazil, and Japan, have implemented newborn screening (NBS) programs for Pompe disease, resulting in an increase of variants of unknown significance (VUS; Bravo et al., 2017; Burlina et al., 2018; Chien et al., 2019; Elliott et al., 2016; Momosaki et al., 2019; Yang et al., 2014). For variants associated with late onset, the associated phenotypes from NBS cases are still unknown as symptom onset could, in principle, be delayed until (late) adulthood. It will be important to monitor the onset and progress of symptoms in patients identified via NBS programs closely to determine the severity of the newly identified genetic variants.

Public databases, such as dbSNP (https://www.ncbi.nlm.nih.gov/snp) and gnomAD (https://gnomad.broadinstitute.org), provide a source of variants that have been detected in various genome-wide studies (Karczewski et al., 2020; Sherry et al., 2001). A large percentage of these variants represent common sequence variants that have a minor allele frequency (MAF) ≥ 1%. Several of these variants have already been reported for the *GAA* gene and have been ruled out to cause Pompe disease (Kroos et al., 2007; Labrousse et al., 2010; Turaça et al., 2015). However, most of the common sequence variants in these databases are listed as VUSs and may lead to misinterpretation during molecular diagnostics.

In this study, we provide an update of the Pompe disease *GAA* variant database with variants and patients described in the literature up to January 2020. We included information on novel *GAA* variants that were identified via NBS and for which no phenotype was yet known. Known common sequence variants in the *GAA* gene that do not cause Pompe disease have now also been added to prevent misdiagnosis. In addition, selected common missense variants were tested in expression studies and also this information was added to the updated database. The database provides a curated up-to-date reference source for the molecular diagnosis of Pompe disease.

2 | METHODS

The Pompe disease *GAA* variant database is publicly available at http://www.pompevariantdatabase.nl. The previous version of the database included literature until 2016; the update described here contains variants from publications up to January 2020. Additionally, NBS studies that screened for Pompe disease were now included if the authors provided the genotypes of the described cases. Novel variants were analyzed as described in Niño et al. (2019). Variants were annotated based on the reference sequences NM_000152.3 for *GAA* messenger RNA (mRNA), LRG_673 genomic sequence for describing variants in intronic sequences, and NP_000143.2 for GAA protein. Exon annotations were based on the human genomic build (GRCH37/hg19) for exons 2–20; however, changes were made to the annotation of exon 1 to reflect the findings of (GRCH38/hg38). Within this region, a new 195-bp intron was identified at positions c.−112 and c.−113. Therefore, the region that was previously annotated as exon 1 has been split between exons 1A and 1B, which are separated by intron 1A. Intron 1 has been renamed to intron 1B. This numbering was made to maintain the same numbering of subsequent exons compared with existing literature.

Common sequence variants in the *GAA* gene (hg38 Chr17:80,101,556-80,119,881) were extracted from gnomAD and were categorized as "not disease-associated." Combined Annotation-Dependent Depletion (CADD) in silico predictions were performed using the CADD (https://cadd.gs.washington.edu) platform, which compiles different tools for analysis of intronic insertion and deletion variants (Rentzsch et al., 2019). The MAF and CADD scores were obtained in April 2020. Predictions of effect on pre-mRNA splicing were performed using Alamut Visual v.2.15 (Interactive Biosoftware).

Functional studies were performed using site-directed mutagenesis (SDM) to generate complementary DNA (cDNA) expression

TABLE 1 Novel disease-associated variants added to the Pompe variant database

DNA nomenclature	Phenotype combined with a null allele	DNA nomenclature	Phenotype combined with a null allele
Ch37/hg19 chr17:78,059,821_78,076,592del	Unknown (disease-associated)	c.1057C>T	Unknown (disease-associated)
c.−113+2T>A	Unknown (disease-associated)	c.1057del	Unknown (disease-associated)
c.−32-17_−32-10delins(30)	Classic infantile	c.1099T>G	Unknown (disease-associated)
c.−32-1G>C	Unknown (disease-associated)	c.1106T>A	Unknown (disease-associated)
c.40_47del	Classic infantile	c.1109G>A	Unknown (disease-associated)
c.104T>C	Classic infantile	c.1114C>G	Unknown (disease-associated)
c.169C>T	Classic infantile	c.1114C>T	Unknown (disease-associated)
c.205C>T	Unknown (disease-associated)	c.1121G>A	Unknown (disease-associated)
c.258C>A	Unknown (disease-associated)	c.1127_1130del	Unknown (disease-associated)
c.265C>T	Unknown (disease-associated)	c.1129G>A	Unknown (disease-associated)
c.295_314del	Unknown (disease-associated)	c.1153del	Unknown (disease-associated)
c.323G>C	Unknown (disease-associated)	c.1192del	Unknown (disease-associated)
c.365del	Unknown (disease-associated)	c.1193del	Unknown (disease-associated)
c.380G>A	Unknown (disease-associated)	c.1201C>A	Unknown (disease-associated)
c.397T>G	Unknown (disease-associated)	c.1209C>A	Unknown (disease-associated)
c.437del	Classic infantile	c.1211A>C	Unknown (disease-associated)
c.445A>C	Unknown (disease-associated)	c.1211A>T	Classic infantile
c.484A>C	Classic infantile	c.1212C>G	Unknown (disease-associated)
c.502C>T	Unknown (disease-associated)	c.1216G>A	Childhood
c.505C>A	Unknown (disease-associated)	c.1219T>C	Unknown (disease-associated)
c.517_519del	Childhood	c.1221C>A	Classic infantile
c.541_545del	Classic infantile	c.1221del	Unknown (disease-associated)
c.547-1G>C	Unknown (disease-associated)	c.1226_1227insG	Classic infantile
c.568C>T	Unknown (disease-associated)	c.1231del	Unknown (disease-associated)
c.665T>G	Classic infantile	c.1240T>C	Unknown (disease-associated)
c.686G>C	Unknown (disease-associated)	c.1241del	Classic infantile
c.691C>T	Unknown (disease-associated)	c.1242C>A	Unknown (disease-associated)
c.692T>C	Unknown (disease-associated)	c.1249A>C	Unknown (disease-associated)
c.692+1G>T	Unknown (disease-associated)	c.1281G>T	Classic infantile
c.693-2A>C	Classic infantile	c.1292_1295dup	Classic infantile
c.693-1G>C	Unknown (disease-associated)	c.1293_1326+57del	Unknown (disease-associated)
c.715_716del	Unknown (disease-associated)	c.1298A>C	Classic infantile
c.730C>T	Classic infantile	c.1311_1312ins(26)	Classic infantile
c.736del	Unknown (disease-associated)	c.1320_1322del	Classic infantile
c.756_757insT	Unknown (disease-associated)	c.1327-54_1437+178del	Classic infantile
c.759del	Unknown (disease-associated)	c.1358_1361del	Classic infantile
c.766_784del	Unknown (disease-associated)	c.1378G>T	Unknown (disease-associated)
c.781G>A	Classic infantile	c.1388_1406del	Unknown (disease-associated)
c.784G>C	Unknown (disease-associated)	c.1396dup	Unknown (disease-associated)

(Continues)

TABLE 1 (Continued)

DNA nomenclature	Phenotype combined with a null allele	DNA nomenclature	Phenotype combined with a null allele
c.796C>A	Childhood	c.1402A>T	Unknown (disease-associated)
c.799_803delinsA	Unknown (disease-associated)	c.1409A>G	Unknown (disease-associated)
c.837G>C	Unknown (disease-associated)	c.1431del	Classic infantile
c.841C>T	Unknown (disease-associated)	c.1441del	Unknown (disease-associated)
c.876C>G	Classic infantile	c.1447G>T	Unknown (disease-associated)
c.878G>T	Unknown (disease-associated)	c.1456G>T	Unknown (disease-associated)
c.883C>A	Unknown (disease-associated)	c.1464dup	Classic infantile
c.930_932del	Classic infantile	c.1470C>A	Childhood
c.942C>A	Unknown (disease-associated)	c.1477C>T	Unknown (disease-associated)
c.947A>G	Classic infantile	c.1493G>A	Classic infantile
c.950C>T	Unknown (disease-associated)	c.1501_1515del	Unknown (disease-associated)
c.955+1G>A	Classic infantile	c.1507del	Classic infantile
c.971dup	Classic infantile	c.1526A>T	Unknown (disease-associated)
c.982_988del	Classic infantile	c.1531C>A	Unknown (disease-associated)
c.983T>C	Classic infantile	c.1537G>A	Unknown (disease-associated)
c.994_995insTT	Unknown (disease-associated)	c.1538A>G	Classic infantile
c.1000G>T	Classic infantile	c.1551+3A>T	Unknown (disease-associated)
c.1004_1005dup	Unknown (disease-associated)	c.1551+5G>A	Unknown (disease-associated)
c.1047del	Unknown (disease-associated)	c.1559A>G	Unknown (disease-associated)
c.1560C>G	Unknown (disease-associated)	c.2096T>C	Unknown (disease-associated)
c.1579_1580del	Classic infantile	c.2109del	Unknown (disease-associated)
c.1583G>C	Unknown (disease-associated)	c.2131A>C	Classic infantile
c.1594G>A	Adult	c.2146G>C	Unknown (disease-associated)
c.1597T>G	Classic infantile	c.2153_2156delinsACGCCG	Classic infantile
c.1602_1605delinsAGG	Classic infantile	c.2182_2183del	Unknown (disease-associated)
c.1610del	Unknown (disease-associated)	c.2190-345A>G	Unknown (disease-associated)
c.1627T>G	Unknown (disease-associated)	c.2205dup	Classic infantile
c.1629C>G	Unknown (disease-associated)	c.2213G>A	Classic infantile
c.1636G>C	Unknown (disease-associated)	c.2221G>A	Classic infantile
c.1636+5G>A	Classic infantile	c.2222A>T	Unknown (disease-associated)
c.1650del	Unknown (disease-associated)	c.2234T>C	Classic infantile
c.1657C>T	Classic infantile	c.2235dup	Classic infantile
c.1681_1699dup	Unknown (disease-associated)	c.2237G>T	Unknown (disease-associated)
c.1688A>T	Unknown (disease-associated)	c.2240G>A	Unknown (disease-associated)
c.1716C>A	Classic infantile	c.2261dup	Unknown (disease-associated)
c.1721T>C	Unknown (disease-associated)	c.2294G>A	Classic infantile
c.1753_2799del	Classic infantile	c.2296T>A	Classic infantile
c.1754+1dup	Unknown (disease-associated)	c.2297A>C	Classic infantile
c.1754+2T>C	Unknown (disease-associated)	c.2304del	Unknown (disease-associated)
c.1780C>T	Unknown (disease-associated)	c.2320G>A	Unknown (disease-associated)

TABLE 1 (Continued)

DNA nomenclature	Phenotype combined with a null allele	DNA nomenclature	Phenotype combined with a null allele
c.1784C>T	Unknown (disease-associated)	c.2331+5G>C	Classic infantile
c.1799G>C	Unknown (disease-associated)	c.2331+102del	Unknown (disease-associated)
c.1822del	Unknown (disease-associated)	c.2334_2335dup	Unknown (disease-associated)
c.1825T>G	Unknown (disease-associated)	c.2377_2378insAC	Classic infantile
c.1835A>C	Unknown (disease-associated)	c.2380dup	Unknown (disease-associated)
c.1835A>G	Unknown (disease-associated)	c.2395C>T	Unknown (disease-associated)
c.1837T>G	Unknown (disease-associated)	c.2407C>T	Unknown (disease-associated)
c.1839G>C	Unknown (disease-associated)	c.2411G>A	Classic infantile
c.1844_1846del	Unknown (disease-associated)	c.2459_2461del	Unknown (disease-associated)
c.1844G>T	Classic infantile	c.2460dup	Unknown (disease-associated)
c.1844G>A	Classic infantile	c.2474C>G	Unknown (disease-associated)
c.1847dup	Unknown (disease-associated)	c.2480A>G	Unknown (disease-associated)
c.1859C>A	Unknown (disease-associated)	c.2515C>T	Unknown (disease-associated)
c.1879_1881del	Classic infantile	c.2537C>A	Unknown (disease-associated)
c.1888+2_1888+15del	Classic infantile	c.2544del	Unknown (disease-associated)
c.1895T>C	Unknown (disease-associated)	c.2563G>C	Classic infantile
c.1895T>G	Classic infantile	c.2578G>A	Unknown (disease-associated)
c.1903A>G	Unknown (disease-associated)	c.2584G>A	Childhood
c.1913G>A	Classic infantile	c.2585del	Classic infantile
c.1944_1950del	Unknown (disease-associated)	c.2596del	Unknown (disease-associated)
c.1952dup	Unknown (disease-associated)	c.2619C>G	Unknown (disease-associated)
c.1961C>G	Unknown (disease-associated)	c.2636T>C	Classic infantile
c.2004C>A	Unknown (disease-associated)	c.2655_2656del	Unknown (disease-associated)
c.2015G>T	Unknown (disease-associated)	c.2716G>A	Unknown (disease-associated)
c.2020C>G	Unknown (disease-associated)	c.2720T>C	Unknown (disease-associated)
c.2020C>T	Unknown (disease-associated)	c.2725G>A	Unknown (disease-associated)
c.2024A>G	Classic infantile	c.2740dup	Unknown (disease-associated)
c.2040+2dup	Unknown (disease-associated)	c.2742dup	Classic infantile
c.2040+29_2190-270del	Classic infantile	c.2757del	Unknown (disease-associated)
c.2041-2A>G	Classic infantile	c.2799+5G>A	Unknown (disease-associated)
c.2051C>A	Unknown (disease-associated)	c.2800-1G>C	Classic infantile
c.2051C>G	Unknown (disease-associated)	c.2843dup	Classic infantile
c.2051C>T	Classic infantile	c.2845_2847del	Unknown (disease-associated)
c.2056_2057delinsCC	Unknown (disease-associated)		
c.2084dup	Unknown (disease-associated)		

constructs containing the missense variant of interest as described (in 't Groen et al., 2020). The activity of the GAA protein produced by the constructs was measured using 4-methylumbelliferyl-α-D-glucopyranoside (4-MU) as a substrate in transfected COS-7 cells, as described in Kroos et al. (2008). Statistical analysis was performed using one-way analysis of variance with Tukey honestly significant difference post hoc multiple testing corrections. $p < .05$ was considered significant.

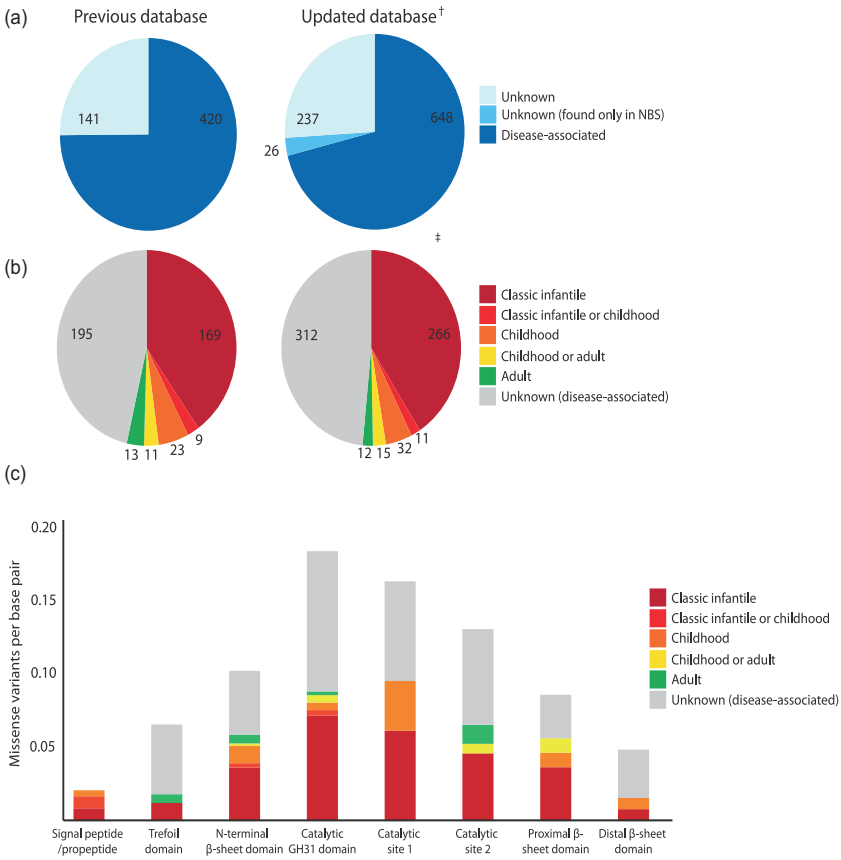


FIGURE 1 Overview of variants, comparing the previous (Niño et al., 2019) and updated version of the Pompe disease GAA variant database (http://www.pompevariantdatabase.nl). (a) Number of disease-associated and unknown variants in the previous database (left) and the updated version of the database (right). (b) Number of disease-associated variants classified based on the predicted clinical phenotype when combined with a null allele in the previous database (left) and in the updated version of the database (right). (c) Distribution of disease-associated missense variants listed in the updated database, based on the protein domains of GAA and the predicted clinical phenotype when combined with a null allele. Numbers are corrected for the length of each domain. †Two entries in the previous version of the database were removed as the variants were described twice using different nomenclatures. ‡For 36 variants listed in the previous version of the database, a reclassification of the phenotypic severity was performed due to the addition of novel patients included in this update

3 | RESULTS AND DISCUSSION

Table 1 provides an overview of the novel variants. We performed a literature search covering the past 4 years and identified 80 publications (listed in the updated database and Table S1) that described 350 novel variants, of which 226 were considered to be disease-associated (Table 1 and Figure 1a). Seventy-six novel variants (33%) were present in combination with a null allele, which allowed prediction of the clinical severity of these variants (Table 1 and Figure 1b). In addition, the inclusion of new patient information allowed us to classify the severity of 55 variants that were already present in the database. This resulted in a new total of 911 *GAA* variants, of which 648 were disease-associated (71%). In total, 336 out of 648 disease-associated variants (52%) could be associated with a clinical phenotype. The geographical or ethnical distribution of reported patients remained similar to what was described previously. The majority of patients had a Caucasian background or were of Caucasian descent (data not shown). This introduces a bias in the current version of the database and indicates the necessity of extending the database to patients of other descent. Mapping of missense variants to GAA protein domains revealed an even

TABLE 2 List of common sequence variants located within the boundaries of the *GAA* gene

Location	Variant	Variant ID	Global allele frequency (GnomAD)	Predictions of pre-mRNA splicing	CADD score PHRED
Exon 1A, 5′ UTR	c.−338C>G	rs144639114	2%	No effect on splicing	6.524
Exon 1A, 5′ UTR	c.−260G>C	rs2304849	16%	No effect on splicing	8.996
Exon 1A, 5′ UTR	c.−178G>A	rs77514632	2%	No effect on splicing	9.948
Exon 1B, 5′ UTR	c.−75C>G	rs80020206	0.9% (3% in African population)	No effect on splicing	9.989
Intron 1B	c.−33+219G>C	rs4889961	75%	No effect on splicing	0.866
Intron 1B	c.−33+316C>A	rs8077055	20%	No effect on splicing	9.079
Intron 1B	c.−33+317C>T	rs8077056	20%	No effect on splicing	8.579
Intron 1B	c.−33+671A>C	rs55751636	31%	No effect on splicing	1.456
Intron 1B	c.−33+757G>A	rs28413147	5%	No effect on splicing	4.974
Intron 1B	c.−33+903A>C	rs12450199	34%	No effect on splicing	8.196
Intron 1B	c.−33+1104A>G	rs11150841	75%	No effect on splicing	6.976
Intron 1B	c.−33+1172G>A	rs1442315	5%	No effect on splicing	0.064
Intron 1B	c.−33+1190G>T	rs12602593	10%	No effect on splicing	1.784
Intron 1B	c.−33+1309T>C	rs1442314	76%	No effect on splicing	1.752
Intron 1B	c.−32-1298G>C	rs12602610	33%	No effect on splicing	2.604
Intron 1B	c.−32-1124C>T	rs58959690	20%	No effect on splicing	5.825
Intron 1B	c.−32-884T>C	rs145362066	0.9% (3% in African population)	No effect on splicing	3.993
Intron 1B	c.−32-793C>G	rs55666739	2%	No effect on splicing	4.041
Intron 1B	c.−32-721G>C	rs75754966	2%	Generates a new cryptic splice accepter site	1.008
Intron 1B	c.−32-686A>G	rs147264695	0.3% (1% in Finnish population)	No effect on splicing	4.349
Intron 1B	c.−32-640C>T	rs12600845	51%	No effect on splicing	0.136
Intron 1B	c.−32-521G>T	rs115060925	1%	Generates a new cryptic splice donor site	0.639
Intron 1B	c.−32-494C>G	rs140325572	2%	No effect on splicing	0.036
Intron 1B	c.−32-462G>A	rs74003606	5%	No effect on splicing	0.226
Exon 2	c.271G>A	rs1800299	2%	No effect on splicing	0.256
Exon 2	c.324T>C	rs1800300	72%	No effect on splicing	8.391
Exon 2	c.447G>A	rs2289536	0.5% (3% in East Asian population)	No effect on splicing	1.252
Intron 2	c.546+293G>A	rs34746710	20%	No effect on splicing	1.899
Intron 2	c.547-243C>G	rs8065426	67%	No effect on splicing	2.529
Intron 2	c.547-238T>C	rs12452263	20%	No effect on splicing	5.667
Intron 2	c.547-67C>G	rs8069491	67%	No effect on splicing	1.337
Intron 2	c.547-39T>G	rs12452721	67%	Loss of cryptic splice donor site	2.78
Intron 2	c.547-4C>G	rs3816256	67%	No effect on splicing	4.721
Exon 3	c.596A>G	rs1042393	67%	No effect on splicing	0.548
Exon 3	c.642C>T	rs1800301	18%	No effect on splicing	1.805
Exon 3	c.668G>A	rs1042395	67%	No effect on splicing	1.46
Intron 3	c.692+38C>T	rs2304848	3%		5.574

(Continues)

TABLE 2 (Continued)

Location	Variant	Variant ID	Global allele frequency (GnomAD)	Predictions of pre-mRNA splicing	CADD score PHRED
				Generates a new cryptic splice donor site	
Intron 3	c.692+144A>G	rs2304847	67%	No effect on splicing	3.653
Intron 3	c.692+509T>C	rs8082405	66%	No effect on splicing	3.271
Intron 3	c.692+674G>C	rs8078350	67%	No effect on splicing	4.501
Intron 3	c.692+751T>C	rs8068051	67%	No effect on splicing	2.363
Intron 3	c.693-586G>A	rs112308142	3%	No effect on splicing	2.71
Intron 3	c.693-585T>C	rs8068555	67%	No effect on splicing	4.133
Intron 3	c.693-559C>T	rs12602422	67%	No effect on splicing	1.879
Intron 3	c.693-491G>A	rs12948631	67%	No effect on splicing	3.629
Intron 3	c.693-441C>G	rs12602440	67%	Loss of a cryptic splice acceptor site	7.559
Intron 3	c.693-434C>A	rs12941269	66%	No effect on splicing	4.416
Intron 3	c.693-414C>G	rs12941289	66%	Loss of a cryptic splice acceptor site	0.077
Intron 3	c.693-413A>G	rs12937590	67%	Loss of a cryptic splice acceptor site	1.544
Intron 3	c.693-216T>A	rs11150844	67%	No effect on splicing	4.13
Intron 3	c.693-94C>T	rs79849256	0.2% (3% in East Asian population)	No effect on splicing	9.666
Intron 3	c.693-78C>T	rs74003611	6%	No effect on splicing	0.06
Intron 3	c.693-49C>T	rs78855075	7%	No effect on splicing	2.374
Exon 4	c.852G>A	rs142626724	0.6% (1% in European population)	No effect on splicing	1.095
Intron 4	c.858+30T>C	rs2304845	66%	No effect on splicing	0.067
Exon 5	c.921A>T	rs1800303	8%	No effect on splicing	9.101
Intron 5	c.955+12G>A	rs2252455	69%	No effect on splicing	0.981
Intron 5	c.955+155C>A	rs9901190	5%	No effect on splicing	7.196
Intron 5	c.955+167C>T	rs77717164	0.7% (6% in East Asian population)	No effect on splicing	6.348
Intron 5	c.956-107G>A	rs2241888	73%	No effect on splicing	5.835
Intron 5	c.956-84C>T	rs2241887	67%	No effect on splicing	0.061
Intron 6	c.1075+13C>T	rs41292402	1%	No effect on splicing	7.496
Exon 8	c.1203G>A	rs1800304	67%	No effect on splicing	5.972
Exon 8	c.1286A>G	rs200294882	0.07% (1% in East Asian population)	Loss of cryptic splice acceptor site and generates a new cryptic splice donor site	0.068
Intron 8	c.1326+132G>A	rs894306	67%	No effect on splicing	1.999
Intron 8	c.1326+459C>T	rs74679377	0.7% (6% in East Asian population)	No effect on splicing	0.435
Intron 8	c.1326+460G>A	rs12150323	2%	No effect on splicing	0.322
Intron 8	c.1327-514G>A	rs72850826	5%	No effect on splicing	1.914
Intron 8	c.1327-356G>T	rs6565640	73%	No effect on splicing	0.258

TABLE 2 (Continued)

Location	Variant	Variant ID	Global allele frequency (GnomAD)	Predictions of pre-mRNA splicing	CADD score PHRED
Intron 8	c.1327-321del	rs140385114	7%	No effect on splicing	0.888
Intron 8	c.1327-269A>G	rs6565641	67%	No effect on splicing	4.207
Intron 8	c.1327-209C>T	rs76604157	0.3% (6% in East Asian population)	No effect on splicing	0.471
Intron 8	c.1327-179G>A	rs2278620	20%	No effect on splicing	0.643
Intron 8	c.1327-118A>G	rs74003628	7%	No effect on splicing	0.184
Intron 8	c.1327-18A>G	rs2278619	72%	No effect on splicing	0.124
Exon 9	c.1374C>T	rs1800305	7%	No effect on splicing	0.206
Intron 9	c.1438-220A>G	rs2278618	67%	No effect on splicing	6.607
Intron 9	c.1438-108G>A	rs12944802	67%	No effect on splicing	0.013
Intron 9	c.1438-19G>C	rs2304844	67%	No effect on splicing	3.529
Intron 10	c.1551+42G>A	rs115427918	0.9% (3% in African population)	No effect on splicing	5.792
Intron 10	c.1551+49C>A	rs2304843	67%	No effect on splicing	7.131
Exon 11	c.1581G>A	rs1042396	23%	No effect on splicing	6.758
Intron 11	c.1636+43G>T	rs2304842	5%	Generates a new cryptic splice accepter site	6.859
Intron 11	c.1636+117del	rs199788201	59%	No effect on splicing	0.045
Intron 11	c.1636+117C>T	rs12945868	11%	No effect on splicing	0.181
Intron 11	c.1636+118G>T	rs4889817	59%	No effect on splicing	3.161
Intron 11	c.1636+205C>T	rs79673008	3%	No effect on splicing	0.013
Intron 11	c.1636+210G>A	rs79487884	5%	No effect on splicing	1.463
Intron 11	c.1636+269C>T	rs111625854	2%	No effect on splicing	3.828
Intron 11	c.1636+284G>C	rs111551014	2%	No effect on splicing	1.81
Intron 11	c.1636+389C>G	rs7221675	63%	No effect on splicing	0.573
Intron 11	c.1636+390A>G	rs7209921	63%	No effect on splicing	1.829
Intron 11	c.1636+404A>G	rs4889818	74%	No effect on splicing	1.902
Intron 11	c.1637-185A>G	rs12951255	55%	No effect on splicing	0.576
Exon 12	c.1726G>A	rs1800307	2%	Generates a new cryptic splice acceptor	0.268
Intron 12	c.1754+12G>A	rs2304840	6%	No effect on splicing	4.325
Intron 12	c.1754+100C>T	rs113688685	0.9% (3% in African population)	No effect on splicing	8.142
Intron 12	c.1754+104C>G	rs2304839	5%	No effect on splicing	0.763
Intron 12	c.1754+144C>T	rs2304838	61%	No effect on splicing	1.787
Intron 12	c.1755-186A>G	rs62075593	2%	No effect on splicing	2.032
Intron 13	c.1888+21G>A	rs2304837	6%	No effect on splicing	3.378
Intron 14	c.2040+20A>G	rs2304836	72%	No effect on splicing	2.163
Intron 14	c.2040+66C>T	rs2304835	7%	No effect on splicing	3.54
Intron 14	c.2040+69A>G	rs2304834	6%	No effect on splicing	0.027
Intron 14	c.2041-64G>A	rs2304833	27%	No effect on splicing	0.371
Exon 15	c.2065G>A	rs1800309	6%	No effect on splicing	1.783

(Continues)

TABLE 2 (Continued)

Location	Variant	Variant ID	Global allele frequency (GnomAD)	Predictions of pre-mRNA splicing	CADD score PHRED
Exon 15	c.2133A>G	rs1800310	27%	No effect on splicing	1.134
Intron 15	c.2189+95C>T	rs72850840	5%	No effect on splicing	3,771
Intron 15	c.2189+263G>A	rs7221604	66%	Generates a new cryptic splice donor site	0.563
Intron 15	c.2189+510T>G	rs4889963	5%	No effect on splicing	1.444
Intron 15	c.2189+607G>A	rs112710614	7%	No effect on splicing	0.189
Intron 15	c.2189+616T>C	rs139307163	5%	No effect on splicing	1.94
Intron 15	c.2189+723G>A	rs4889819	20%	No effect on splicing	0.367
Intron 15	c.2189+729A>G	rs74737410	5%	No effect on splicing	0.498
Intron 15	c.2189+859A>G	rs4889964	5%	No effect on splicing	1.503
Intron 15	c.2189+884G>A	rs4889965	5%	No effect on splicing	0.355
Intron 15	c.2189+1153A>G	rs72850844	5%	No effect on splicing	3.687
Intron 15	c.2189+1201C>A	rs72850846	5%	No effect on splicing	2.352
Intron 15	c.2189+1208A>G	rs72850847	5%	No effect on splicing	0.367
Intron 15	c.2189+1263A>G	rs74700450	5%	No effect on splicing	2.97
Intron 15	c.2189+1290A>G	rs74003630	5%	No effect on splicing	6.015
Intron 15	c.2189+1600C>T	rs60668271	5%	No effect on splicing	0.481
Intron 15	c.2190-1531G>A	rs74702528	0.9% (3% in African population)	No effect on splicing	0.489
Intron 15	c.2190-1463G>A	rs116416508	0.9% (3% in African population)	No effect on splicing	0.328
Intron 15	c.2190-1139A>G	rs184803352	0.7% (2% in African population	No effect on splicing	0.095
Intron 15	c.2190-1005A>G	rs4889820	5%	No effect on splicing	2.452
Intron 15	c.2190-686G>A	rs12452616	19%	No effect on splicing	2.725
Intron 15	c.2190-647G>A	rs59362713	10%	No effect on splicing	0.227
Intron 15	c.2190-536G>A	rs60429724	10%	No effect on splicing	0.454
Intron 15	c.2190-490G>A	rs111477580	1%	No effect on splicing	3.101
Intron 15	c.2190-444A>G	rs4889967	73%	No effect on splicing	1.059
Intron 15	c.2190-336C>T	rs76178719	3%	No effect on splicing	1.566
Intron 16	c.2331+20G>A	rs2304832	75%	No effect on splicing	5.346
Intron 16	c.2331+24T>C	rs2304831	15%	No effect on splicing	0.204
Intron 16	c.2331+151C>T	rs111537160	2%	No effect on splicing	0.608
Intron 16	c.2332-198A>T	rs2304830	73%	No effect on splicing	3.363
Exon 17	c.2338G>A	rs1126690	72%	No effect on splicing	2.675
Exon 17	c.2446G>A	rs1800314	5%	No effect on splicing	5.793
Intron 17	c.2482-132C>T	rs113824706	0.9% (3% in African population)	No effect on splicing	0.066
Exon 18	c.2553G>A	rs1042397	57%	Weakens a cryptic splice donor site	1.241
Intron 18	c.2647-71G>C	rs4889821	5%	No effect on splicing	3.473
Exon 19	c.2780C>T	rs1800315	2%	No effect on splicing	0.222
Intron 19	c.2800-227C>G	rs9890469	66%	No effect on splicing	0.661
Intron 19	c.2800-60G>A	rs55662462	0.7% (11% in Latino population)	No effect on splicing	2.209
Exon 20, 3′ UTR	c.*3G>A	rs1800317	5%	No effect on splicing	0.03

TABLE 2 (Continued)

Location	Variant	Variant ID	Global allele frequency (GnomAD)	Predictions of pre-mRNA splicing	CADD score PHRED
Exon 20, 3′ UTR	c.*91G>A	rs2229221	12%	No effect on splicing	6.887
Exon 20, 3′ UTR	c.*223C>T	rs8132	22%	No effect on splicing	3.025
Exon 20, 3′ UTR	c.*419G>T	rs7567	19%	No effect on splicing	4.17

Abbreviations: CADD, Combined Annotation-Dependent Depletion; mRNA, messenger RNA; UTR, untranslated region.

stronger enrichment in the catalytic core compared with the mapping we performed previously (Niño et al., 2019; Figure 1c).

We included in the current version of the database common sequence variants that have a MAF ≥ 1% and do not cause Pompe disease. This resulted in a relative increase in the number of nondisease-associated variants (Table 2). We decided to include common sequence variants in response to the misreporting of these variants as the principal cause of disease in several patients. Examples of this are the c.547-67C>G (rs8069491) and 547-39T>G (rs12452721) variants, which were reported as the cause of disease while having an allele frequency of 67% in the global population (Bekircan-Kurt et al., 2017; Guevara-Campos et al., 2019). In total, the database now includes 148 variants with a MAF ≥ 1%. All variants had a low CADD score (<10; Table 2) and were classified as "unknown." We note that while these common sequence

(a)

Variant	Protein change	phenotype combined with a null allele	reported patients	Predictions on pre-mRNA splicing	CADD score PHRED
GAA + c.1597T>C	p.(Cys533Arg)	Classic infantile	1	no effect on splicing	25.5
GAA + c.307T>G	p.(Cys103Gly)	Classic infantile	11	loss of a cryptic splice donor site	25.1
GAA + c.309C>G	p.(Cys103Trp)	Unknown	1	no effect on splicing	5.6
GAA + c.655G>A	p.(Gly219Arg)	Classic infantile	14	no effect on splicing	28.2
GAA + c.670C>T	p.(Arg224Trp)	Classic infantile or Childhood	7	no effect on splicing	22.8
GAA + c.1655T>C	p.(Leu552Pro)	Classic infantile	41	no effect on splicing	29.9
GAA + c.1798C>T	p.(Arg600Cys)	Classic infantile	18	no effect on splicing	27.0

(b)

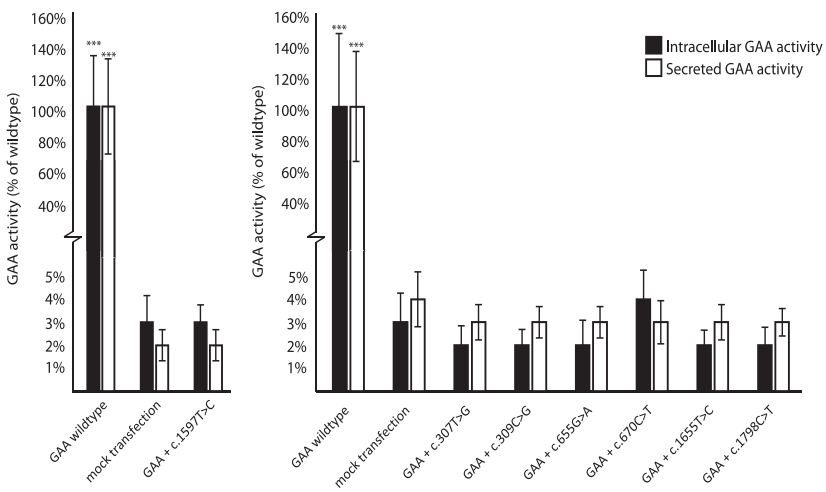


FIGURE 2 Expression study of seven disease-associated missense variants in the *GAA* gene. (a) Overview of basic information regarding the pathogenicity of selected variants. (b) Measured *GAA* activity in both cells and medium of COS-7 cultures after transfection with the generated constructs. Findings for the c.1597T>C variants are plotted separately as this was performed in a separate experiment. Data represent means, error bars represent *SD* ($n = 3$ biological replicates), ***$p < .001$. CADD, Combined Annotation-Dependent Depletion; mRNA, messenger RNA

TABLE 3 Variants of unknown significance that were found only through newborn screening programs

Variant	Protein change	Location	Type of variant (protein)	MAF	Predictions on splicing–Align GVGD–SIFT–Mutation taster–[CADD score]	Experimental data	Country and reference
c.317G>A*	p.(Arg106His)	Exon 2	Missense	MAF <1%	No effect on splicing–Class C0–Deleterious–Disease causing–[25.9]		Japan; Momosaki et al. (2019)
c.365T>A	p.(Met122Lys)	Exon 2	Missense	MAF not reported	No effect on splicing–Class C0–Tolerated–Polymorphism–[14.17]		USA; Scott et al. (2013)
c.424_440del	p.(Ser142Leufs*29)	Exon 2	Frameshift	MAF not reported	No effect on splicing–Results in an out of frame product–[32]		Taiwan; Chien et al. (2011)
c.533G>A*	p.(Arg178His)	Exon 2	Missense	MAF <1%	No effect on splicing–Class C0–Tolerated–Disease causing–[31]	No effect on splicing of exon 2 in minigene construct (Goina, et al., 2019)	Taiwan; Chien et al. (2011)
c.546+5G>T*	p.?	Intron 2	No category (splicing)	MAF <1%	Weakens exon 2 splice donor and generates a cryptic splice donor–[23.7]	Affects splicing of exon 2 in minigene construct (Goina, et al., 2019)	Taiwan; Labrousse et al. (2010)
c.705G>A	p.(=)	Exon 4	Silent	MAF <1%	No effect on splicing–[0.534]		Japan; Momosaki et al. (2019)
c.811A>G*	p.(Thr271Ala)	Exon 4	Missense	MAF not reported	No effect on splicing–Class C0–Tolerated–Polymorphism–[16.93]	71% residual activity of GAA in expression study (Kroos, et al., 2012a)	Taiwan; Labrousse et al. (2010)
c.1054C>T	p.(Gln352*)	Exon 6	Nonsense	MAF not reported	No effect on splicing–Introduces an early stop codon–[43]		Taiwan; Liao et al. (2014)
c.1080C>G	p.(Tyr360*)	Exon 7	Nonsense	MAF not reported	No effect on splicing–Introduces an early stop codon–[39]		Taiwan; Chien et al. (2011)
c.1082C>A	p.(Pro361Arg)	Exon 7	Missense	MAF <1%	No effect on splicing–Class C65–Deleterious–Disease causing–[25.5]		Japan; Momosaki et al. (2019)
c.1220A>G	p.(Tyr407Cys)	Exon 8	Missense	MAF <1%	No effect on splicing–Class C65–Deleterious–Disease causing–[25.9]		Mexico; Navarrete-Martínez et al. (2017)
c.1244C>T	p.(Thr415Met)	Exon 8	Missense	MAF <1%	No effect on splicing–Class C15–Deleterious–Disease causing–[24.6]		Japan; Momosaki et al. (2019)
c.1324G>A*	p.(Val442Met)	Exon 8	Missense	MAF <1%	No effect on splicing–Class C0–Deleterious–Disease causing–[23.8]		Taiwan; Chien et al. (2011)
c.1409A>C	p.(Asn470Thr)	Exon 9	Missense	MAF <1%	No effect on splicing–Class C25–Deleterious–Disease causing–[23.2]		Hungary; Witmann et al. (2012)

TABLE 3 (Continued)

Variant	Protein change	Location	Type of variant (protein)	MAF	Predictions on splicing–Align GVGD–SIFT–Mutation taster–[CADD score]	Experimental data	Country and reference
c.1574T>A	p.(Phe525Tyr)	Exon 11	Missense	MAF not reported	No effect on splicing–Class C15–Deleterious–Disease causing–[28.8]	10% residual activity of GAA in expression study (Kroos, et al., 2012a)	Taiwan; Chien et al. (2011)
c.1805C>T	p.(Thr602Ile)	Exon 13	Missense	MAF not reported	No effect on splicing–Class C0–Tolerated–Disease causing–[24.1]		USA; Elliott et al. (2016)
c.1840A>G	p.(Thr614Ala)	Exon 13	Missense	MAF not reported	No effect on splicing–Class C55–Deleterious–Disease causing–[24.3]		Taiwan; Liao et al. (2014)
c.1925T>A	p.(Val642Asp)	Exon 14	Missense	MAF not reported	No effect on splicing–Class C45–Deleterious–Disease causing–[29.2]		USA; Scott et al. (2013)
c.1958C>A	p.(Thr653Asn)	Exon 14	Missense	MAF <1%	No effect on splicing–Class C15–Tolerated–Disease causing–[25.4]		Taiwan; Chien et al. (2011)
c.2003A>G*	p.(Tyr668Cys)	Exon 14	Missense	MAF not reported	No effect on splicing–Class C65–Deleterious–Disease causing–[31]		Japan; Momosaki et al. (2019)
c.2055C>G	p.(Tyr685*)	Exon 15	Nonsense	MAF not reported	No effect on splicing–Introduces an early stop codon–[36]		Japan; Momosaki et al. (2019)
c.2174G>A	p.(Arg725Gln)	Exon 15	Missense	MAF <1%	No effect on splicing–Class C0–Tolerated–Disease causing–[32]		Hungary; Witmann et al. (2012)
c.2482-5T>C*	p.?	Intron 17	No category (splicing)	MAF not reported	No effect on splicing–[8.409]		Taiwan; Liao et al. (2014)
c.2482-2A>G	p.?	Intron 17	No category (splicing)	MAF <1%	Loss of exon 18 splice acceptor site–[35]		Hungary; Witmann et al. (2012)
c.2647-23del	p.?	Intron 18	No category (intron variant)	MAF <1%	No effect on splicing–[0.451]		Taiwan; Liao et al. (2014)
c.2843dup	p.(Val949Argfs*69)	Exon 20	Frameshift	MAF not reported	No effect on splicing–Results in an out of frame product–[23.1]		Taiwan; Liao et al. (2014)

Abbreviations: CADD, Combined Annotation-Dependent Depletion; MAF, minor allele frequency.

*Variants found in *cis* with the Asian pseudodeficiency allele c.[1726G>A; 2065G>A].

variants do not result in clinical manifestation of Pompe disease, it remains possible that they might modify disease progression when present in *cis* with a disease-associated variant. In Pompe disease, this is the case for the Asian pseudodeficiency allele (c.[1726G>A (p.Gly576Ser);2065G>A (p.Glu689Lys)]) and GAA2 (c.271G>A, (p.Asp91Asn)), which have a MAF of 14% for c.1726G>A, 23.5% for c.2065G>A (both East Asian), and 3.2% for GAA2 (European), and can be present in *cis* with known disease-associated variants (Kroos et al., 2006; Labrousse et al., 2010). Also, a variant with a low MAF in the general population, c.510C>T (p.=) (rs564758226), is known to be linked to the late-onset variant c.−32-13T>G (p.[=,0]) (IVS1). c.510C>T has a global MAF of 0.005%, but a MAF of 27.3% in compound heterozygous IVS1 patients with symptom onset at childhood. It worsens aberrant splicing caused by IVS1 and causes lower levels of leaky wild-type splicing and lower GAA enzyme activity, resulting in accelerated disease onset (Bergsma et al., 2019).

Figure 2a,b shows the results on the *GAA* variants we subjected to a more in-depth investigation. We selected the common missense variants c.307T>G (p.Cys103Gly), c.655G>A (p.Gly219Arg), c.670C>T (p.Arg224Trp), c.1655T>C (p.Leu552Pro), and c.1798C>T (p.Arg600Cys) and performed in vitro analysis of their severity using SDM of *GAA* cDNA expression constructs. In addition, c.1597T>C (p.Cys533Arg) and c.309C>G (p.Cys103Trp) were tested due to a request for diagnostic purposes. All of these variants fully abrogated GAA enzymatic activity following transfection in COS-7 cells (Figure 2, compare mutant GAA with mock transfections). The c.309C>G variant was included because the patient that harbored this variant in combination with c.525del p.(Glu176Argfs*45) showed an atypical Pompe disease phenotype (Mori et al., 2017). This case report described an adult patient with cardiomyopathy. Molecular analysis of primary skin fibroblasts identified a reduction in GAA activity, although not at pathogenic levels, and GAA activity was in the normal range for skeletal muscle tissue (Mori et al., 2017). We note that the c.309C>G variant was not detected in DNA from either parent and was described as a de novo variant (Mori et al., 2017). This variant might have been introduced during embryonic development, resulting in mosaicism similar to, as described previously in Labrijn-Marks et al. (2019) and in 't Groen et al. (2020). This might explain the "uneven pattern" of glycogen accumulation in histological sections derived from cardiac tissue (Mori et al., 2017). The in vitro analysis indicated that the c.309C>G variant is fully deleterious and has a predicted classic infantile phenotype in combination with a null allele. A comprehensive genetic analysis would be necessary to confirm this hypothesis.

Novel variants that have been reported only through NBS studies, but for which no clinical phenotype has been provided, were classified as "Unknown (found only in NBS)". In the current version of the database, 26 variants have been classified as such (Table 3). Seven out of 26 variants were also present in *cis* with the Asian pseudodeficiency allele, indicating that additional testing is required because the Asian pseudodeficiency is known to result in false-positive outcomes in dried blood spot-based assays (Liao et al., 2014; Momosaki et al., 2019). It is currently unknown at what age symptoms will develop in neonates diagnosed with disease-associated variants that are potentially associated with a late-onset phenotype. Symptoms might be delayed until late adulthood or, for some genetic variants, might not even lead to disease. In these cases, further research on the effect of the genetic variants is essential to better inform patients, families, and doctors. As reported, in these cases, the uncertainty of the diagnosis, the possibility of an emerging disease, and the doubt on when to start treatment with ERT could lead to emotional stress (Bodamer et al., 2017). This underscores the importance of phenotype prediction for disease-associated variants, especially in the case of asymptomatic patients identified through NBS programs.

The sharp increase in reports on patients with Pompe disease and *GAA* disease-associated variants highlights the need for regular updates of the Pompe disease *GAA* variant database. Increased awareness and improved diagnostic technology with exome and genome sequencing and NBS programs are expected to further increase the number of entries in the database in the coming years. It will be important to link variants to clinical information and to test their deleterious effect in vitro using expression and splicing assays. Curated disease-specific databases such as the Pompe disease *GAA* variant database will be important to provide guidance to clinicians and clinical geneticists to establish an accurate molecular diagnosis.

ACKNOWLEDGMENTS

We thank the members of the Molecular Stem Cell Biology group for the critical discussions. This study was funded through Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brasil; grant number: 234407/2014-0), Sophia Foundation for Medical Research (SSWO; project number: s17-32); Metakids (project number: 2016-063), and Departamento Administrativo de Ciencia, Tecnología e Innovación (Colciencias). The collaboration project is cofunded by the PPP Allowance made available by Health-Holland, Top Sector Life Sciences & Health, to the Prinses Beatrix Spier fonds to stimulate public–private partnerships (project number: LSHM19015).

CONFLICT OF INTERESTS

Ans T. van der Ploeg 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 remaining authors declare that there are no conflict of interests.

WEB RESOURCES

Pompe disease *GAA* variant database: http://www.pompevariantdatabase.nl/

LOVD: http://gaa.lovd.nl/

GnomAD: https://gnomad.broadinstitute.org/

dbSNP: https://www.ncbi.nlm.nih.gov/snp/

CADD score: https://cadd.gs.washington.edu/

DATA AVAILABILITY STATEMENT

The data described in this study is available upon request from the corresponding authors, and new variants have been added to the Pompe disease *GAA* variant database (http://www.pompevariantdatabase.nl/) and LOVD (http://gaa.lovd.nl/).

ORCID

Douglas O. S. de Faria https://orcid.org/0000-0001-9525-7260

W. W. M. Pim Pijnappel https://orcid.org/0000-0002-7042-2482

REFERENCES

Bali, D. S., Goldstein, J. L., Banugaria, S., Dai, J., Mackey, J., Rehder, C., & Kishnani, P. S. (2012). Predicting cross-reactive immunological material (CRIM) status in Pompe disease using GAA mutations: Lessons learned from 10 years of clinical laboratory testing experience. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, *160C*(1), 40–49. https://doi.org/10.1002/ajmg.c.31319

Bekircan-Kurt, C. E., Güneş, H. N., Yildiz, F. G., Saka, E., Tan, E., & Erdem-Özdamar, S. (2017). New mutations and genotype–phenotype correlation in late-onset Pompe patients. *Acta Neurologica Belgica*, *117*, 269–275. https://doi.org/10.1007/s13760-016-0738-7

Bergsma, A. J., in 't Groen, S. L. M., van den Dorpel, J. J. A., van den Hout, H. J. M. P., van der Beek, N. A. M. E., Schoser, B., Toscano, A., Musumeci, O., Bembi, B., Dardis, A., Morrone, A., Tummolo, A., Pasquini, E., van der Ploeg, A. T., & Pijnappel, W. W. M. P. (2019). A genetic modifier of symptom onset in Pompe disease. *EBioMedicine*, *43*, 553–561. https://doi.org/10.1016/j.ebiom.2019.03.048

Bodamer, O. A., Ronald Scott, C., & Giugliani, R. (2017). Newborn screening for Pompe disease. *Pediatrics*, *140*, S4–S13. https://doi.org/10.1542/peds.2016-0280C

Bravo, H., Neto, E. C., Schulte, J., Pereira, J., Filho, C. S., Bittencourt, F., Sebastião, F., Bender, F., de Magalhães, A. P. S., Guidobono, R., Trapp, F. B., Michelin-Tirelli, K., Souza, C. F. M., Rojas Málaga, D., Pasqualim, G., Brusius-Facchin, A. C., & Giugliani, R. (2017). Investigation of newborns with abnormal results in a newborn screening program for four lysosomal storage diseases in Brazil. *Molecular Genetics and Metabolism Reports*, *12*, 92–97. https://doi.org/10.1016/j.ymgmr.2017.06.006

Burlina, A. B., Polo, G., Salviati, L., Duro, G., Zizzo, C., Dardis, A., Bembi, B., Cazzorla, C., Rubert, L., Zordan, R., Desnick, R. J., & Burlina, A. P. (2018). Newborn screening for lysosomal storage disorders by tandem mass spectrometry in North East Italy. *Journal of Inherited Metabolic Disease*, *41*, 209–219. https://doi.org/10.1007/s10545-017-0098-3

Chien, Y.-H., Hwu, W.-L., & Lee, N.-C. (2019). Newborn screening: Taiwanese experience. *Annals of Translational Medicine*, *7*, 281. https://doi.org/10.21037/atm.2019.05.47

Chien, Y.-H., Lee, N.-C., Huang, H.-J., Thurberg, B.-L., Tsai, F.-J., & Hwu, W.-L. (2011). Later-onset Pompe disease: Early detection and early treatment initiation enabled by newborn screening. *Journal of Pediatrics*, *158*(6), 1023–1027. https://doi.org/10.1016/j.jpeds.2010.11.053

Elliott, S., Buroker, N., Cournoyer, J. J., Potier, A. M., Trometer, J. D., Elbin, C., Schermer, M. J., Kantola, J., Boyce, A., Turecek, F., Gelb, M. H., & Scott, C. R. (2016). Pilot study of newborn screening for six lysosomal storage diseases using tandem mass spectrometry. *Molecular Genetics and Metabolism*, *118*, 304–309. https://doi.org/10.1016/j.ymgme.2016.05.015

Goina, E., Musco, L., Dardis, A., & Buratti, E. (2019). Assessment of the functional impact on the pre-mRNA splicing process of 28 nucleotide variants associated with Pompe disease in GAA exon 2 and their recovery using antisense technology. *Human Mutation*, *40*(11), 2121–2130. https://doi.org/10.1002/humu.23867

Guevara-Campos, J., González-Guevara, L., & Cauli, O. (2019). Skeletal alterations, developmental delay and new mutations in juvenile-onset Pompe disease. *Neuromuscular Disorders*, *29*, 192–197. https://doi.org/10.1016/j.nmd.2018.11.013

Güngör, D., & Reuser, A. J. (2013). How to describe the clinical spectrum in Pompe disease? *American Journal of Medical Genetics, Part A*, *161*, 399–400. https://doi.org/10.1002/ajmg.a.35662

in't Groen S. L. M., de Faria, D. O. S., Iuliano, A., van den Hout, J. M. P., Douben, H., Dijkhuizen, T., Cassiman, D., Witters, P., Barba Romero, M.-Á., de Klein, A., Somers-Bolman, G. M., Saris, J. J., Hoefsloot, L. H., van der Ploeg, A. T., Bergsma, A. J., & Pijnappel, W. W. M. P. (2020). Novel GAA variants and mosaicism in Pompe disease identified by extended analyses of patients with an incomplete DNA diagnosis. *Molecular Therapy—Methods & Clinical Development*, *17*, 337–348. https://doi.org/10.1016/j.omtm.2019.12.016

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., … MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7

Kohler, L., Puertollano, R., & Raben, N. (2018). Pompe disease: From basic science to therapy. *Neurotherapeutics*, *15*, 928–942. https://doi.org/10.1007/s13311-018-0655-y

Kroos, M., Hoogeveen-Westerveld, M., Michelakakis, H., Pomponio, R., van der Ploeg, A., Halley, D., & Reuser, A. (2012a). GAA Database Consortium. 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. *Human Mutation*, *33*(8), 1161–1165. https://doi.org/10.1002/humu.22108

Kroos, M., Hoogeveen-Westerveld, M., van der Ploeg, A., & Reuser, A. J. (2012b). The genotype-phenotype correlation in Pompe disease. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, *160C*, 59–68. https://doi.org/10.1002/ajmg.c.31318

Kroos, M., Manta, P., Mavridou, I., Muntoni, F., Halley, D., Van der Helm, R., Zaifeiriou, D., Van der Ploeg, A., Reuser, A. J., & Michelakakis, H. (2006). Seven cases of Pompe disease from Greece. *Journal of Inherited Metabolic Disease*, *29*, 556–563. https://doi.org/10.1007/s10545-006-0280-5

Kroos, M., Pomponio, R. J., van Vliet, L., Palmer, R. E., Phipps, M., Van der Helm, R., Halley, D., Reuser, A. J., & GAA Database Consortium. (2008). Update of the Pompe disease mutation database with 107 sequence variants and a format for severity rating. *Human Mutation*, *29*, E13–E26. https://doi.org/10.1002/humu.20745

Kroos, M. A., Pomponio, R. J., Hagemans, M. L., Keulemans, J. L. M., Phipps, M., DeRiso, M., Palmer, R. E., Ausems, M. G. E. M., Van der Beek, N. A. M. E., Van Diggelen, O. P., Halley, D. J. J., Van der Ploeg, A. T., & Reuser, A. J. J. (2007). Broad spectrum of Pompe disease in patients with the same c.−32-13T→G haplotype. *Neurology*, *68*, 110–115. https://doi.org/10.1212/01.wnl.0000252798.25690.76

Kulessa, M., Weyer-Menkhoff, I., Viergutz, L., Kornblum, C., Claeys, K. G., Schneider, I., Plöckinger, U., Young, P., Boentert, M., Vielhaber, S., Mawrin, C., Bergmann, M., Weis, J., Ziagaki, A., Stenzel, W., Deschauer, M., Nolte, D., Hahn, A., Schoser, B., & Schänzer, A. (2020). An integrative correlation of myopathology, phenotype and genotype in late onset Pompe disease. *Neuropathology and Applied Neurobiology*, *46*, 359–374. https://doi.org/10.1111/nan.12580

Labrijn-Marks, I., Somers-Bolman, G. M., In't Groen, S. L. M., Hoogeveen-Westerveld, M., Kroos, M. A., Ala-Mello, S., Amaral, O., Miranda, C., Mavridou, I., Michelakakis, H., Naess, K., Verheijen, F. W., Hoefsloot, L. H., Dijkhuizen, T., Benjamins, M., van den Hout, H. J. M., van der Ploeg, A. T., Pijnappel, W. W. M. P., Saris, J. J., & Halley, D. J. (2019). Segmental and total uniparental isodisomy (UPiD) as a disease mechanism in autosomal recessive lysosomal disorders:

Evidence from SNP arrays. *European Journal of Human Genetics*, *27*, 919–927. https://doi.org/10.1038/s41431-019-0348-y

Labrousse, P., Chien, Y. H., Pomponio, R. J., Keutzer, J., Lee, N. C., Akmaev, V. R., Scholl, T., & Hwu, W. L. (2010). Genetic heterozygosity and pseudodeficiency in the Pompe disease newborn screening pilot program. *Molecular Genetics and Metabolism*, *99*, 379–383. https://doi.org/10.1016/j.ymgme.2009.12.014

Liao, H. C., Chiang, C. C., Niu, D. M., Wang, C. H., Kao, S. M., Tsai, F. J., Huang, Y. H., Liu, H. C., Huang, C. K., Gao, H. J., Yang, C. F., Chan, M. J., Lin, W. D., & Chen, Y. J. (2014). Detecting multiple lysosomal storage diseases by tandem mass spectrometry—A national newborn screening program in Taiwan. *Clinica Chimica Acta*, *431*, 80–86. https://doi.org/10.1016/j.cca.2014.01.030

Momosaki, K., Kido, J., Yoshida, S., Sugawara, K., Miyamoto, T., Inoue, T., Okumiya, T., Matsumoto, S., Endo, F., Hirose, S., & Nakamura, K. (2019). Newborn screening for Pompe disease in Japan: Report and literature review of mutations in the GAA gene in Japanese and Asian patients. *Journal of Human Genetics*, *64*, 741–755. https://doi.org/10.1038/s10038-019-0603-7

Mori, M., Bailey, L. A., Estrada, J., Rehder, C. W., Li, J. S., Rogers, J. G., Bali, D. S., Buckley, A. F., & Kishnani, P. S. (2017). Severe cardiomyopathy as the isolated presenting feature in an adult with late-onset Pompe disease: A case report. *JIMD Reports*, *31*, 79–83.

Navarrete-Martínez, J. I., Limón-Rojas, A. E., Gaytán-García, M. J., Reyna-Figueroa, J., Wakida-Kusunoki, G., Delgado-Calvillo, M. D. R., Cantú-Reyna, C., Cruz-Camino, H., & Cervantes-Barragán, D. E. (2017). Newborn screening for six lysosomal storage disorders in a cohort of Mexican patients: Three-year findings from a screening program in a closed Mexican health system. *Molecular Genetics and Metabolism*, *121*(1), 16–21. https://doi.org/10.1016/j.ymgme.2017.03.001

Niño, M. Y., Groen, S. L. M., in't Bergsma, A. J., van der Beek, N. A. M. E., Kroos, M., Hoogeveen-Westerveld, M., van der Ploeg, A. T., Pijnappel, W. W. M. P., & (2019). Extension of the Pompe mutation database by linking disease-associated variants to clinical severity. *Human Mutation*, *40*, 1954–1967. https://doi.org/10.1002/humu.23854

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, *47*, D886–D894. https://doi.org/10.1093/nar/gky1016

Reuser, A. J. J., Ploeg, A. T., Chien, Y.-H., Llerena, J., Abbott, M.-A., Clemens, P. R., Kimonis, V. E., Leslie, N., Maruti, S. S., Sanson, B.-J., Araujo, R., Periquet, M., Toscano, A., Kishnani, P. S., & on behalf of the Pompe Registry Sites (2019). GAA variants and phenotypes among 1,079 patients with Pompe disease: Data from the Pompe Registry. *Human Mutation*, *40*, 2146–2164. https://doi.org/10.1002/humu.23878

Scott, C. R., Elliott, S., Buroker, N., Thomas, L. I., Keutzer, J., Glass, M., Gelb, M. H., & Turecek, F. (2013). Identification of infants at risk for developing Fabry, Pompe, or mucopolysaccharidosis-I from newborn blood spots by tandem mass spectrometry. *Journal of Pediatrics*, *163*(2), 498–503. https://doi.org/10.1016/j.jpeds.2013.01.031

Semplicini, C., Letard, P., De Antonio, M., Taouagh, N., Perniconi, B., Bouhour, F., Echaniz-Laguna, A., Orlikowski, D., Sacconi, S., Salort-Campana, E., Solé, G., Zagnoli, F., Hamroun, D., Froissart, R., Caillaud, C., & Laforêt, P. (2018). Late-onset Pompe disease in France: Molecular features and epidemiology from a nationwide study. *Journal of Inherited Metabolic Disease*, *41*, 937–946. https://doi.org/10.1007/s10545-018-0243-7

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). DbSNP: The NCBI database of genetic variation. *Nucleic Acids Research*, *29*, 308–311. https://doi.org/10.1093/nar/29.1.308

Turaça, L. T., Faria, D. O. S., de Kyosen, S. O., Teixeira, V. D., Motta, F. L., Pessoa, J. G., Rodrigues e Silva, M., Almeida, S. S., de D'Almeida, V., Munoz Rojas, M. V., Martins, A. M., & Pesquero, J. B. (2015). Novel GAA mutations in patients with Pompe disease. *Gene*, *561*, 124–131. https://doi.org/10.1016/j.gene.2015.02.023

van der Ploeg, A. T., & Reuser, A. J. (2008). Pompe's disease. *Lancet*, *372*, 1342–1353. https://doi.org/10.1016/S0140-6736(08)61555-X

van Gelder, C.M., Hoogeveen-Westerveld, M., Kroos, M. A., Plug, I., van der Ploeg, A. T., & Reuser, A. J. (2015). Enzyme therapy and immune response in relation to CRIM status: The Dutch experience in classic infantile Pompe disease. *Journal of Inherited Metabolic Disease*, *8*(2), 305–314. https://doi.org/10.1007/s10545-014-9707-6

Wittmann, J., Karg, E., Turi, S., Legnini, E., Wittmann, G., Giese, A. K., Lukas, J., Gölnitz, U., Klingenhäger, M., Bodamer, O., Mühl, A., & Rolfs, A. (2012). Newborn screening for lysosomal storage disorders in Hungary. *JIMD Reports*, *6*, 117–125. https://doi.org/10.1007/8904_2012_130

Yang, C. F., Liu, H. C., Hsu, T. R., Tsai, F. C., Chiang, S. F., Chiang, C. C., Ho, H. C., Lai, C. J., Yang, T. F., Chuang, S. Y., Lin, C. Y., & Niu, D. M. (2014). A large-scale nationwide newborn screening program for Pompe disease in Taiwan: Towards effective diagnosis and treatment. *American Journal of Medical Genetics, Part A*, *164*, 54–61. https://doi.org/10.1002/ajmg.a.36197

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: de Faria DOS, in 't Groen SLM, Hoogeveen-Westerveld M, et al. Update of the Pompe variant database for the prediction of clinical phenotypes: Novel disease-associated variants, common sequence variants, and results from newborn screening. *Human Mutation*. 2021;42:119–134. https://doi.org/10.1002/humu.24148

Supplementary Table S1: List of publications added to the Pompe disease *GAA* variant database.

PMID	Author	Title	year	Country
19948615	Chien YH, Lee NC, Thurberg BL, et al.	Pompe disease in infants: improving the prognosis by newborn screening and early treatment	2009	Taiwan
20080426	Labrousse P, Chien YH, Pomponio RJ, et al.	Genetic heterozygosity and pseudodeficiency in the Pompe disease newborn screening pilot program	2010	Taiwan
21232767	Chien YH, Lee NC, Huang HJ, Thurberg BL, Tsai FJ, Hwu WL.	Later-onset Pompe disease: early detection and early treatment initiation enabled by newborn screening	2011	Taiwan
22133539	Mechtler TP, Stary S, Metz TF, et al.	Neonatal screening for lysosomal storage disorders: feasibility and incidence from a nationwide study in Austria	2012	Austria
23430949	Wittmann J, Karg E, Turi S, et al.	Newborn screening for lysosomal storage disorders in hungary	2012	Hungary
23465405	Scott CR, Elliott S, Buroker N, et al.	Identification of infants at risk for developing Fabry, Pompe, or mucopolysaccharidosis-I from newborn blood spots by tandem mass spectrometry	2013	USA
24243590	Yang CF, Liu HC, Hsu TR, et al.	A large-scale nationwide newborn screening program for Pompe disease in Taiwan: towards effective diagnosis and treatment	2014	Taiwan
24513544	Liao HC, Chiang CC, Niu DM, et al.	Detecting multiple lysosomal storage diseases by tandem mass spectrometry--a national newborn screening program in Taiwan	2014	Taiwan
27692865	Chu YP, Sheng B, Lau KK, et al.	Clinical manifestation of late onset Pompe disease patients in Hong Kong	2016	Hong Kong
27896132	Matsuoka T, Miwa Y, Tajika M, et al.	Divergent clinical outcomes of alpha-glucosidase enzyme replacement therapy in two siblings with infantile-onset Pompe disease treated in the symptomatic or pre-symptomatic state	2016	Japan
27666774	Ünver O, Hacıfazlıoğlu NE, Karatoprak E, et al.	The frequency of late-onset Pompe disease in pediatric patients with limb-girdle muscle weakness and nonspecific hyperCKemia: A multicenter study	2016	Turkey
26946079	Arslan A, Poyrazoğlu HG, Kiraz A, et al.	Combination of two different homozygote mutations in Pompe disease	2016	Turkey
26800218	Angelini C, Savarese M, Fanin M, Nigro V.	Next generation sequencing detection of late onset Pompe disease	2016	Italy
26474166	Pichiecchio A, Berardinelli A, Moggio M, et al.	Asymptomatic Pompe disease: Can muscle magnetic resonance imaging facilitate diagnosis?	2016	Italy
26809617	Lévesque S, Auray-Blais C, Gravel E, et al.	Diagnosis of late-onset Pompe disease and other muscle disorders by next-generation sequencing	2016	USA
27238910	Elliott S, Buroker N, Cournoyer JJ, et al.	Pilot study of newborn screening for six lysosomal storage diseases using Tandem Mass Spectrometry	2016	USA
26685070	Yang CF, Yang CC, Liao HC, et al.	Very Early Treatment for Infantile-Onset Pompe Disease Contributes to Better Outcomes	2016	Taiwan
28458930	Ceyhan D, Gucyetmez Topal B	An 18-Month-Old Child with Infantile Pompe Disease: Oral Signs	2017	Turkey
2843375	Lee JH, Shin JH, Park HJ, et al.	Targeted population screening of late onset Pompe disease in unspecified myopathy patients for Korean population	2017	South Korea
28380188	Sixel BS, Silva LD, Cavalcanti NC, et al.	Respiratory manifestations in late-onset Pompe disease: a case series conducted in Brazil	2017	Brazil
28554557	Almeida V, Conceição I, Fineza I, et al.	Screening for Pompe disease in a Portuguese high risk population	2017	Portugal
28265479	Sifi Y, Medjroubi M, Froissart R, et al.	Clinical Analysis of Algerian Patients with Pompe Disease	2017	Algeria
28394184	Chen X, Liu T, Huang M, et al.	Clinical and Molecular Characterization of Infantile-Onset Pompe Disease in Mainland Chinese Patients: Identification of Two Common Mutations	2017	China

29124014	Fukuhara Y, Fuji N, Yamazaki N, et al.	A molecular analysis of the GAA gene and clinical spectrum in 38 patients with Pompe disease in Japan	2017	Japan
29390460	Martínez M, Romero MG, Guereta LG, et al.	Infantile-onset Pompe disease with neonatal debut: A case report and literature review	2017	Nigeria
28694071	Torrealba-Acosta G, Rodríguez-Roblero MC, Bogantes-Ledezma S, Carazo-Céspedes K, Desnuelle C.	First clinical and genetic description of a family diagnosed with late-onset Pompe disease from Costa Rica	2017	Costa Rica
28657663	Tsai AC, Hung YW, Harding C, Koeller DM, Wang J, Wong LC.	Next generation deep sequencing corrects diagnostic pitfalls of traditional molecular approach in a patient with prenatal onset of Pompe disease	2017	USA
28032299	Bekircan-Kurt CE, Güneş HN, Yildiz FG, Saka E, Tan E, Erdem-Özdamar S.	New mutations and genotype-phenotype correlation in late-onset Pompe patients	2017	Turkey
28937052	Quan JJ, Liu LJ, Lyu TW, Huang XP, Tian J.	Atypical Infantile-onset Pompe Disease with Hypertrophic Cardiomyopathy	2017	China
28302345	Navarrete-Martínez JI, Limón-Rojas AE, Gaytán-García MJ, et al.	Newborn screening for six lysosomal storage disorders in a cohort of Mexican patients: Three-year findings from a screening program in a closed Mexican health system	2017	Mexico
27862019	Taisne N, Desnuelle C, Juntas Morales R, et al.	Bent spine syndrome as the initial symptom of late-onset Pompe disease	2017	France
28900456	Ebrahimi M, Behnam M, Behranvand-Jazi N, et al.	Identification a novel mononucleotide deletion mutation in GAA in pompe disease patients	2017	Iran
28856460	Pichiecchio A, Sacco S, De Filippi P, et al	Late-onset Pompe disease: a genetic-radiological correlation on cerebral vascular anomalies	2017	Italy
28763149	McIntosh P, Austin S, Sullivan J, et al.	Three cases of multi-generational Pompe disease: Are current practices missing diagnostic and treatment opportunities?	2017	USA
27142047	Mori M, Bailey LA, Estrada J, et al.	Severe Cardiomyopathy as the Isolated Presenting Feature in an Adult with Late-Onset Pompe Disease: A Case Report	2017	USA
27649523	Nazari F, Sinaei F, Nilipour Y, et al.	Late-onset pompe disease in Iran: A clinical and genetic report	2017	Iran
28721335	Bravo H, Neto EC, Schulte J, et al.	Investigation of newborns with abnormal results in a newborn screening program for four lysosomal storage diseases in Brazil	2017	Brazil
27927596	Schänzer A, Kaiser AK, Mühlfeld C, et al.	Quantification of muscle pathology in infantile Pompe disease	2017	Germany
29122469	Mori M, Haskell G, Kazi Z, et al.	Sensitivity of whole exome sequencing in detecting infantile- and late-onset Pompe disease	2017	USA
28838325	Byrne BJ, Geberhiwot T, Barshop BA, et al.	A study on the safety and efficacy of reveglucosidase alfa in patients with late-onset Pompe disease	2017	USA
27708273	Reddy HM, Cho KA, Lek M, et al.	The sensitivity of exome sequencing in identifying pathogenic mutations for LGMD in the United States	2017	USA
29149851	Johnson K, Töpf A, Bertoli M, et al.	Identification of GAA variants through whole exome sequencing targeted to a cohort of 606 patients with unexplained limb-girdle muscle weakness	2017	UK
28951071	Rairikar MV, Case LE, Bailey LA, et al.	Insight into the phenotype of infants with Pompe disease identified by newborn screening with the common c.-32-13T>G "late-onset" GAA variant	2017	USA
31966564	Xu L, Zhang L, Zhong L, et al.	A novel compound heterozygous GAA mutation in a Chinese family with juvenile onset form of Pompe disease with cardiomyopathy	2017	China
29326002	Golsari A, Nasimzadah A, Thomalla G, Keller S, Gerloff C, Magnus T.	Prevalence of adult Pompe disease in patients with proximal myopathic syndrome and undiagnosed muscle biopsy	2018	Germany

29451150	Liu HX, Pu CQ, Shi Q, Zhang YT, Ban R.	Identification of Seven Novel Mutations in the Acid Alpha-glucosidase Gene in Five Chinese Patients with Late-onset Pompe Disease	2018	China
29181627	Löscher WN, Huemer M, Stulnig TM, et al.	Pompe disease in Austria: clinical, genetic and epidemiological aspects	2018	Austria
29880332	Savarese M, Torella A, Musumeci O, et al.	Targeted gene panel screening is an effective tool to identify undiagnosed late onset Pompe disease	2018	Italy
29749992	Moravej H, Amirhakimi A, Showraki A, Amoozgar H, Hadipour Z, Nikfar G.	A New Mutation Causing Severe Infantile-Onset Pompe Disease Responsive to Enzyme Replacement Therapy	2018	Iran
28972689	Eleftheriadis T, Makri P, Karakosta P, et al.	Late-onset Pompe's disease in a hemodialysis patient: A first case report	2018	Greece
30023291	Al-Hassnan ZN, Khalifa OA, Bubshait DK, et al.	The phenotype, genotype, and outcome of infantile-onset Pompe disease in 18 Saudi patients	2018	Saudi Arabia
28431840	Roche Bueno JC, Arcos Sánchez C, Salgado Álvarez de Sotomayor F, Izquierdo-Álvarez S, Miramar Gallart MD, Solera García J.	Novel probable pathological variant c.1249A>C in exon 7 of the GAA gene associated with Pompe disease in adults	2018	Spain
29143201	Burlina AB, Polo G, Salviati L, et al.	Newborn screening for lysosomal storage disorders by tandem mass spectrometry in North East Italy	2018	Italy
29637184	Elenga N, Verloes A, Mrsic Y, et al.	Incidence of infantile Pompe disease in the Maroon population of French Guiana	2018	French Guiana
30093193	Hossain MA, Miyajima T, Akiyama K, Eto Y. A	Case of Adult-onset Pompe Disease with Cerebral Stroke and Left Ventricular Hypertrophy	2018	Japan
30371346	Adadi N, Sahli M, Egéa G, et al.	Post-mortem diagnosis of Pompe disease by exome sequencing in a Moroccan family: a case report	2018	Marocco
29742245	Lorenzoni PJ, Kay CSK, Higashi NS, D'Almeida V, Werneck LC, Scola RH.	Late-onset Pompe disease: what is the prevalence of limb-girdle muscular weakness presentation?	2018	Brazil
30155607	Semplicini C, Letard P, De Antonio M, et al.	Late-onset Pompe disease in France: molecular features and epidemiology from a nationwide study	2018	France
30564623	Nallamilli BRR, Chakravorty S, Kesari A, et al.	Genetic landscape and novel disease mechanisms from a large LGMD cohort of 4656 patients	2018	USA
30367637	Szklanny K, Tylki-Szymańska A.	Follow-up analysis of voice quality in patients with late-onset Pompe disease	2018	Poland
29422078	Parini R, De Lorenzo P, Dardis A, et al.	Long term clinical history of an Italian cohort of infantile onset Pompe disease treated with enzyme replacement therapy	2018	Italy
30175981	Remiche G, Lukacs Z, Kasper DC, Abramowicz M, Pandolfo M.	Low Prevalence Estimates of Late-Onset Glycogen Storage Disease Type II in French-Speaking Belgium are not Due to Missed Diagnoses	2018	Belgium
30105547	Rupp S, Felimban M, Schänzer A, et al.	Genetic basis of hypertrophic cardiomyopathy in children	2019	Germany
30770309	Amiñoso C, Gordillo-Marañón M, Hernández J, Solera J.	Reevaluating the pathogenicity of the mutation c.1194 +5 G>A in GAA gene by functional analysis of RNA in a 61-year-old woman diagnosed with Pompe disease by muscle biopsy	2019	Spain
30778879	Gragnaniello V, Fecarotta S, Pecoraro A, et al.	Desensitization of two young patients with infantile-onset Pompe disease and severe reactions to alglucosidase alfa	2019	Italy
30678746	Yeşilbaş O, Epçaçan S.	Occurrence of nutritional hypocalcaemic rickets-related dilated cardiomyopathy in a child with concomitant rickets and infantile-onset Pompe disease	2019	Turkey
31076647	Momosaki K, Kido J, Yoshida S, et al.	Newborn screening for Pompe disease in Japan: report and literature review of mutations in the GAA gene in Japanese and Asian patients	2019	Japan
31101460	Schänzer A, Görlach J, Claudi K, Hahn A.	Severe distal muscle involvement and mild sensory neuropathy in a boy with infantile onset Pompe disease treated with enzyme replacement therapy for 6 years	2019	Germany

30697769	Ebert SE, Brenzy K, Cartwright MS.	Neuromuscular ultrasound as an initial evaluation for suspected myopathy: A case report	2019	USA
30093709	Wasserstein MP, Caggana M, Bailey SM, et al.	The New York pilot newborn screening program for lysosomal storage diseases: Report of the First 65,000 Infants	2019	USA
31125121	Jastrzębska A, Potulska-Chromik A, Łusakowska A, et al.	Screening for late-onset Pompe disease in Poland	2019	Poland
31510962	Ngiwsara L, Wattanasirichaigoon D, Tim-Aroon T, et al.	Clinical course, mutations and its functional characteristics of infantile-onset Pompe disease in Thailand	2019	Thailand
31342611	Reuser AJJ, van der Ploeg AT, Chien YH, et al.	GAA variants and phenotypes among 1,079 patients with Pompe disease: Data from the Pompe Registry	2019	#
30902109	Poelman E, Hoogeveen-Westerveld M, van den Hout JMP, et al.	Effects of immunomodulation in classic infantile Pompe patients with high antibody titers	2019	The Netherlands
30360039	Kim MS, Song A, Im M, et al.	Clinical and molecular characterization of Korean children with infantile and late-onset Pompe disease: 10 years of experience with enzyme replacement therapy at a single center	2019	South Korea
30801484	Jauhari P, Saini AG, Suthar R, et al.	Thenar Hypertrophy and Electrical Myotonia in Pompe Disease	2019	India
31392188	Alandy-Dy J, Wencel M, Hall K, et al.	Variable clinical features and genotype-phenotype correlations in 18 patients with late-onset Pompe disease	2019	USA
31915562	Jay AM, Anne P, Stockton D.	Cardiac Murmur in a Boy with Normal Paternal Prenatal Carrier Screening for Pompe Disease	2019	USA
30897595	Lyu JW, Xu XB, Ji KQ, et al.	Activated mTOR signaling pathway in myofibers with inherited metabolic defect might be an evidence for mTOR inhibition therapies	2019	China
31606152	Gupta N, Kazi ZB, Nampoothiri S, et al.	Clinical and Molecular Disease Spectrum and Outcomes in Patients with Infantile-Onset Pompe Disease	2020	India

CHAPTER 5

Broad variation in phenotypes for common *GAA* genotypes in Pompe disease

Monica Y. Niño*, Stijn L.M. in 't Groen*, Douglas O.S. de Faria, Marianne Hoogeveen-Westerveld, Hannerieke J.M.P. van den Hout, Ans T. van der Ploeg, Atze J. Bergsma, and W.W.M. Pim Pijnappel

*These authors contributed equally to this study
Hum Mutat. 2021 Aug 18. Online ahead of print
doi: 10.1002/humu.24272

Received: 19 June 2021 | Revised: 30 July 2021 | Accepted: 15 August 2021

DOI: 10.1002/humu.24272

RESEARCH ARTICLE

Human Mutation HGVS HUMAN GENOME VARIATION SOCIETY WILEY

Broad variation in phenotypes for common *GAA* genotypes in Pompe disease

Monica Y. Niño[1,2,3] | Stijn L. M. in't Groen[1,2,3] | Douglas O. S. de Faria[1,2,3] | Marianne Hoogeveen-Westerveld[2] | Hannerieke J. M. P. van den Hout[1,3] | Ans T. van der Ploeg[1,3] | Atze J. Bergsma[1,2,3] | W. W. M. Pim Pijnappel[1,2,3]

[1]Department of Pediatrics, Erasmus University Medical Center, Rotterdam, The Netherlands

[2]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands

[3]Center for Lysosomal and Metabolic Diseases, Erasmus University Medical Center, Rotterdam, The Netherlands

Correspondence

W. W. M. Pim Pijnappel and Atze J. Bergsma, Department of Pediatrics, Erasmus University Medical Center, 3015 GE Rotterdam, The Netherlands.
Email: w.pijnappel@erasmusmc.nl and a.bergsma@erasmusmc.nl

Abstract

Patients with the common c.-32-13T > G/null *GAA* genotype have a broad variation in age at symptom onset, ranging from early childhood to late adulthood. Phenotypic variation for other common *GAA* genotypes remains largely unexplored. Here, we analyzed variation in age at symptom onset for the most common *GAA* genotypes using the updated and extended Pompe *GAA* variant database. Patients with the c.2647-7G > A/null genotype invariably presented symptoms at adulthood, while the c.-32-13T > G/null, c.546G > T/null, c.1076-22T > G/null, c.2238G > C/null, and c.2173C > T/null genotypes led to presentations from early childhood up to late adulthood. The c.1309C > T/null genotype was associated with onset at early to late childhood. Symptom onset shifted toward higher ages in homozygous patients. These findings indicate that a broad variation in symptom onset occurs for various common *GAA* genotypes, suggesting the presence of modifying factors. We identified three new compound heterozygous c.-32-13T > G/null patients who carried the genetic modifier c.510C > T and who showed symptom onset at childhood. While c.510C > T acted by lowering GAA enzyme activity, other putative genetic modifiers did not at the group level, suggesting that these act in trans on processes downstream of GAA enzyme activity.

KEYWORDS

GAA deficiency, genotype–phenotype, glycogenosis type II, mutation database, Pompe disease

1 | INTRODUCTION

Pompe disease is a monogenic lysosomal storage disorder caused by disease-associated variants in the acid α-glucosidase (*GAA*) gene (van der Ploeg & Reuser, 2008). A large number of disease-associated *GAA* variants have been identified and are listed in the recently updated Pompe disease *GAA* variant database at http://www.pompevariantdatabase.nl. The database provides *GAA* genotypes and phenotypes of all patients in which these were reported (de Faria et al., 2021; Niño et al., 2019). Two disease-associated *GAA* variants, one per allele, are required to cause Pompe disease.

The clinical spectrum of Pompe disease ranges from a classic infantile form to an adult-onset form (Gaeta et al., 2015; Gungor & Reuser, 2013; Herzog et al., 2012; M. A. Kroos et al., 2007; Laforêt

Monica Y. Niño and Stijn L. M. in't Groen contributed equally to this study.

et al., 2013; Reuser et al., 2018; van der Beek et al., 2009; van der Ploeg & Reuser, 2008; van der Ploeg et al., 2017). The most severe classic infantile form of Pompe disease is caused by two very severe disease-associated *GAA* variants that completely abrogate GAA enzymatic activity. Classic infantile patients present with generalized skeletal muscle weakness and hypertrophic cardiomyopathy shortly after birth, and die within the first year of life if left untreated (Kishnani et al., 2006; H. M. van den Hout et al., 2003). Patients with some residual GAA enzyme activity have a variable onset of symptoms due to the presence of at least one milder disease-associated *GAA* variant. These patients develop skeletal muscle weakness, leading to impaired mobility and respiration, resulting in a wheelchair and/or ventilator dependency at some point in their life (van der Beek et al., 2009, 2012). Cardiac symptoms are usually absent in this patient group.

Treatment of Pompe disease with enzyme replacement therapy (ERT) is available since 2006. ERT improves survival and is effective in normalizing cardiac hypertrophy and improving skeletal muscle function, resulting in increased mobility and stabilization of respiratory function (Amalfitano et al., 2001; Kishnani et al., 2007; Nicolino et al., 2009; Strothotte et al., 2010; van Capelle et al., 2010; H. van den Hout et al., 2000; J. M. van den Hout et al., 2001; van der Ploeg et al., 2010). The invasive nature of ERT treatment and its high costs are important factors that need to be taken into account in determining when treatment of patients with childhood or adulthood onset of symptoms should be started. Prediction of the phenotype that is associated with a particular genotype is an important aspect of decision-making.

While most disease-associated *GAA* variants are rare, a number of variants occur more frequently. Common variants are enriched in distinct populations, including three variants in the Caucasian population: c.-32-13T > G (p.[0, p.=]) (also termed IVS1), c.525del (p.(Glu176Argfs*45)), and c.2481 + 102_2646 + 31del (p.(Gly828_Asn882del)) (also termed delex18); two variants in the Asian population: c.1935C > A (p.(Asp645Glu)) and c.2238G > C (p.(Trp746Cys)); and one variant in patients from African-American descent: c.2560C > T (p.(Arg854*)) (Ausems et al., 2001; Becker et al., 1998; Dagnino et al., 2000; Huie et al., 1994; M. A. Kroos et al., 1995; Laforet et al., 2000; Lin & Shieh, 1996; Liu et al., 2014; Müller-Felber et al., 2007; Pittis & Filocamo, 2007; Reuser et al., 2018; Shieh & Lin, 1998; Van der Kraan et al., 1994; Wokke et al., 1995). The Caucasian c.-32-13T > G disease-associated variant is associated with a very broad range of age at symptom onset (Herzog et al., 2012; M. A. Kroos et al., 2007; Semplicini et al., 2018). This phenotypic variation is also observed within and between families as is illustrated by a study on 22 families with two or three siblings carrying the c.-32-13T > G variant (Wens et al., 2013). These findings have led to the hypothesis that modifying factors—genetic, epigenetic, or environmental—may alter the disease course in patients that carry the c.-32-13T > G variant (Herzog et al., 2012; Huie et al., 1994; Ko et al., 1999; M. Kroos et al., 2012; M. A. Kroos et al., 2007; Musumeci et al., 2015; Rairikar et al., 2017; Shieh & Lin, 1998; van der Ploeg & Reuser, 2008; Wens et al., 2013). It is, however, largely unknown whether phenotypic variation occurs in groups of patients with specific disease-associated variants other than c.-32-13T > G. We recently identified the *GAA* c.510C > T (p.(=)) variant as a genetic modifier that accelerates symptom onset in compound heterozygous and homozygous c.-32-13T > G patients (Bergsma et al., 2019), but additional putative variants that alter the course of Pompe disease remain enigmatic so far.

In the present study, we used the information included in the Pompe disease *GAA* variant database to compare variations in symptom onset for the most common *GAA* disease-associated variants in Pompe disease. We analyzed compound heterozygous and homozygous patients, investigated the effect of the second *GAA* allele, and made a prediction whether additional modifying factors for c.-32-13T > G patients may be present in cis or in trans. The findings on phenotypic variations associated with *GAA* genotypes will be important for diagnosis, genetic counseling, decision-making on the start of treatment with ERT, and for research into the identification of additional modifying factors.

2 | METHODS

2.1 | Patient information

To analyze genotypic variation, we used the most recent version of the Pompe disease *GAA* variant database (http://www.pompevariantdatabase.nl, update January 2020) (de Faria et al., 2021). References to the publications that contain the original patient information in this manuscript are listed in Tables S1 and S2. Geographical origins were indicated or were categorized in the following four groups: Caucasian (for patients from Europe, North America, and Australia), Latin American (for patients from Central and South America), African (for patients from the African continent), and Asian (for patients from the Asian continent). For the analysis of GAA enzyme activity in fibroblasts from patients with the c.-32-13T > G variant and childhood or adulthood onset in the absence or presence of the modifier c.510C > T, we analyzed the patient cohort that was described in Bergsma et al. (2019).

2.2 | Nomenclature

Variant annotations and classification conform to recommendations of the Human Genome Variation Society (HGVS) (den Dunnen et al., 2016). NM_000152.5 was used as a reference sequence for *GAA* mRNA, LRG_673t1.1 was used for intronic regions, and NP_000143.2 for GAA protein. Position c.1 represents the first nucleotide of the translation start codon located in *GAA* exon 2.

2.3 | GAA enzyme activity assay

For the analysis of GAA enzyme activity in fibroblasts from patients with the c.-32-13T > G variant and childhood or adulthood onset in the absence or presence of the modifier c.510C > T, GAA enzyme activities were determined by the diagnostic department of Clinical Genetics of the Erasmus MC in fibroblasts using

4-methylumbelliferone-α-D-glucopyranoside (4-MU) (Sigma-Aldrich) as a substrate as previously described (M. A. Kroos et al., 2007).

3 | RESULTS AND DISCUSSION

3.1 | Phenotypic variation in compound heterozygous patients

The most frequent compound heterozygous genotypes in the Pompe disease *GAA* variant database (http://www.pompevariantdatabase.nl) were listed from patients in whom a classified disease-associated *GAA* variant was combined with a null allele, which is defined as an allele that produces no detectable GAA enzyme activity and that occurs in classic infantile patients that have the most severe form of Pompe disease (Niño et al., 2019). The ages at symptom onset were plotted for five types of *GAA* variants, based on their association with the following groups of phenotypes: adult, childhood or adult, childhood, classic infantile or childhood, and classic infantile (Figure 1 and Table 1). It should be noted that the definition of symptoms can vary between publications, and as a result could contribute to the variability in age at symptom onset, which is a limitation of this analysis. A minimum number of four patients per *GAA* genotype was required in order for a variant to be included in the analysis. We note that other frequent compound heterozygous genotypes were also present in the database; in these cases, the second alleles were not null alleles. Since the clinical phenotype is the result of the combined action of two alleles, the absence of a null allele as the second variant precludes the classification of the variant in question.

In patients with the noncoding splice variant c.2647-7G > A (Caucasian), which affects splicing of *GAA* exon 18 and is associated with the adult phenotype in combination with a null allele, age at symptom onset ranged from 35 years to 53 years of age (median, 40 years) (Figure 1 and Table 1). The patients with this variant all

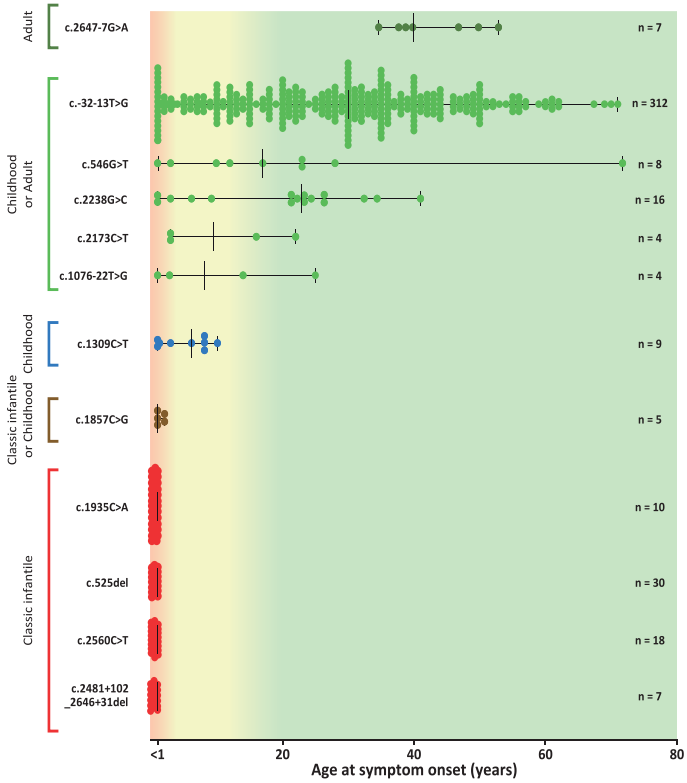


FIGURE 1 Variations in age at symptom onset in compound heterozygous patients. Age at symptom onset is indicated for patients that carry the most frequently occurring variants at compound heterozygous state in combination with a null allele (minimum of four patients). Clinical phenotype groups are shown. Median ages at symptom onset are indicated by long vertical lines; ranges by short vertical lines. One exceptional case was excluded from the figure, this involved an adult patient with two very severe *GAA* variants (genotype: c.1935C > A/c.2560C > T, age at symptom onset: 25 years, phenotype: adult), who was considered to be an enigmatic case (Hermans et al., 1993)

TABLE 1 Information on variants and patients described in Figure 1

DNA/protein nomenclature	Predicted effect	Phenotype with a null allele	Total number of patients	Phenotype of patients	Population[a]
c.2647-7G>A	Splicing	Adult	7	Adult (7)	Italian (7)
Combined with 1 null allele					
c.-32-13T>G	Splicing	Childhood or adult	463	Classic infantile (1) Childhood (101) Adult (306) Unknown (55)	French (79), United States (74), Italian (56), Dutch (48), Caucasian (30), German (28), Polish (17), Brazilian (15), Austrian (6), Hispanic (5), Iranian (4), Turkish (4), Greek (2), Colombian (2), Danish (2), British (2), Moroccan (2), Romanian (1), Chinese (1), Costa Rican (1), Serbian (1), Canadian (1), Unknown (82)
Combined with 67 null alleles					
c.2238G>C, p.(Trp746Cys)	Missense	Childhood or adult	20	Childhood (9) Adult (11)	Chinese (14), Taiwanese (4), South Korean (1), Caucasian (1)
Combined with 9 null alleles					
c.2173C>T, p.(Arg725Trp)	Missense	Childhood or adult	5	Childhood (4) Adult (1)	Hispanic (2), French (2), British (1)
Combined with 3 null alleles					
c.1076-22T>G	Splicing	Childhood	4	Childhood (3) Adult (1)	United States (1), Caucasian (1), German (1), Austrian (1)
Combined with 3 null alleles					
c.1309C>T, p.(Arg437Cys)	Missense	Childhood	10	Childhood (10)	Chinese (5), Japanese (3), Korean (2)
Combined with 4 null alleles					
c.546G>T	Splicing	Childhood	9	Childhood (5) Adult (4)	Japanese (7), Korean (2)
Combined with 6 null alleles					
c.1857C>G, p.(Ser619Arg)	Missense	Classic infantile or childhood	5	Classic infantile (2) Childhood (3)	Japanese (4), Korean (1)
Combined with 3 null alleles					
c.1935C>A, p.(Asp645Glu)	Missense	Classic infantile	47	Classic infantile (41) Childhood (1) Adult (1) Unknown (4)	Taiwanese (32), Chinese (12), United States (2), Thai (1)
Combined with 28 null alleles					
c.525del	Frameshift	Classic infantile	27	Classic infantile (26) Childhood (1)	Dutch (11), United States (7), Italian (4), Caucasian (2), Australian (2), British (1)
Combined with 19 null alleles					
c.2560C>T, p.(Arg854*)	Nonsense	Classic infantile	25	Classic infantile (23) Childhood (1) Adult (1)	French Guianese (9), United States (8), Brazilian (4), Caucasian (2), Colombian (1), Unknown (1)
Combined with 13 null alleles					
c.2481+102_2646+31del	Gross deletion	Classic infantile	27	Classic infantile (27)	Dutch (11), United States (7), Italian (4), Brazilian (2), Hispanic (1), German (1), Unknown (1)
Combined with 20 null alleles					

[a]Population is reported as indicated in the original publication. In certain cases, only Caucasian origin has been reported.

came from a single Italian family (Sampaolo et al., 2013). It is interesting that the age at symptom onset can vary greatly even within the same family. A similar broad range of age at symptom onset has previously been observed within families with another single *GAA* genotype c.-32-13T > G/c.525del (Wens et al., 2013). This indicates that phenotypic variation may occur between members of the same family with the same *GAA* genotype.

Five variants that were associated with onset during either childhood or adulthood, if combined with a null allele, were subjected to further study. The combination of the c.-32-13T > G variant with a null allele was encountered most frequently. In total, 312 compound heterozygous patients in combination with a null allele and from which age at symptom onset was described were included from the Pompe disease *GAA* variant database. Approximately 90% of Caucasian patients with childhood or adulthood onset Pompe disease have been reported to carry this variant (see Bergsma et al., 2019; M. A. Kroos et al., 2007; Montalvo et al., 2006; Semplicini et al., 2018, and references therein). The c.-32-13T > G variant causes aberrant splicing of *GAA* exon 2, resulting in at least eight distinct aberrant splice products (Boerkoel et al., 1995; Huie et al., 1994; van der Wal et al., 2017). It also allows a low level of normal splicing, resulting in the expression of 10%–15% functional GAA protein compared to average healthy control values. The median age at symptom onset of compound heterozygous c.-32-13T > G patients combined with a null allele was 30 years, and an exceptionally broad range between <1 and 71 years of age was found, similarly as reported previously (Herzog et al., 2012; M. A. Kroos et al., 2007; Semplicini et al., 2018). We further tested whether there were sex-related differences in age at symptom onset within this group of patients. No significant differences between males and females could be detected (Figure S1; Kaplan–Meier log-rank test: $p = .3$).

The c.546G > T (p.(=)) variant, which is common among Japanese patients (Fukuhara et al., 2018), also showed a particularly broad range of age at symptom onset ranging from <1 to 72 years of age (median 17 years, $n = 9$ patients). This variant also causes aberrant splicing of exon 2, in which some aberrant splicing products are identical to those caused by the c.-32-13T > G variant (leaky wild-type splicing, complete skipping of exon 2), while other splicing products are unique to c.546G > T (Bergsma et al., 2021; Maimaiti et al., 2009).

Patients with the missense variants c.2238G > C (p.(Trp746Cys)) (Asian) and c.2173C > T (p.(Arg725Trp)) (Caucasian) and the splice variant c.1076-22T > G (Caucasian) also showed a broad variation in age at symptom onset from <1 year to 41 years of age (ranges <1–41, 3–22, and <1–25 years, and medians 22.5, 8.5, and 8 years, respectively). These data indicate that patients that carry *GAA* variants that are associated with childhood or adult onset of the disease show a minimum range in age at symptom onset of 19 years and a maximum of 72 years, in both Asian and Caucasian populations. We note that relatively low patient numbers have been reported to date for these variants, except for c.-32-13T > G, and that it is possible that the inclusion of more patients will further broaden the range of age at symptom onset.

For only one variant, the missense variant c.1309C > T (p.(Arg437Cys)) (Asian), sufficient cases were described ($n = 9$) to predict that this variant in combination with a null allele is associated with childhood onset of symptoms. The age at symptom onset ranged from <1 to 10 years (median 6 years). It remains to be seen in a larger patient cohort whether this variant is always associated with childhood onset or whether it can also occur in adult-onset patients when combined with a null allele.

The c.1857C > G (p.(Ser619Arg)) missense variant was found in combination with a null allele in five Asian patients. The patients diagnosed with this combination presented either with the classic infantile form or a slightly milder form (Fukuhara et al., 2018). In compound heterozygous patients with the c.1857C > G/null genotype, there was very little phenotypic variation with symptom onset <1 year of age in all five patients. However, homozygous patients showed more variation in symptom onset (see below), indicating that the c.1857C > G (p.(Ser619Arg)) variant allows for the production of some residual GAA enzymatic activity, and can thus be classified as a classic infantile or childhood variant.

By definition, the classic infantile variants c.1935C > A (p.(Asp645Glu)) (missense), c.525del (frameshift causing deletion), c.2560C > T (p.(Arg854*)) (nonsense), and c.2481 + 102_2646 + 31del (in-frame deletion) were invariably associated with an age at symptom onset of <1 year of age.

3.2 | Phenotypic variation in homozygous patients

Patients with specific homozygous variants can help to provide more insight into phenotypes of these variants since these can be assessed for only a single disease-associated variant without the presence of a second different *GAA* disease-associated variant. A priori, one would expect that homozygous patients with childhood or adulthood onset would have a median age at symptom onset that is shifted toward a higher age relative to their compound heterozygous counterpart in which the variant is combined with a null allele. This shift is driven by a higher level of residual GAA enzymatic activity produced from the second allele compared to activity in compound heterozygous patients that carry the one copy of the variant combined with a null allele.

One such example is the c.-32-13T > G variant. Twenty-two symptomatic homozygous c.-32-13T > G patients are described in the database. The median age at symptom onset of these patients was 37 years, which was 7 years older compared to patients with the c.-32-13T > G/null allele *GAA* genotype (Figure 2 and Table 2). The range of age at symptom onset was very broad, from <1 year to 58 years of age, which was comparable to the range of age at symptom onset in compound heterozygous patients (range <1–71 years).

As asymptomatic individuals (e.g., from newborn screening, or siblings of patients, or individuals who are homozygous for c.-32-13T > G, see below) were not taken into account in the present analysis and often remain undiagnosed or are not reported, the median age at symptom onset in Figure 2 and Table 2 is likely biased and is in reality higher. The common c.-32-13T > G variant leads to a residual activity of 10%–15% per allele (Boerkoel et al., 1995; Huie et al., 1994; van der Wal et al., 2017). When present at homozygous state, patients with the c.-32-13T > G variant should have 20%–30%

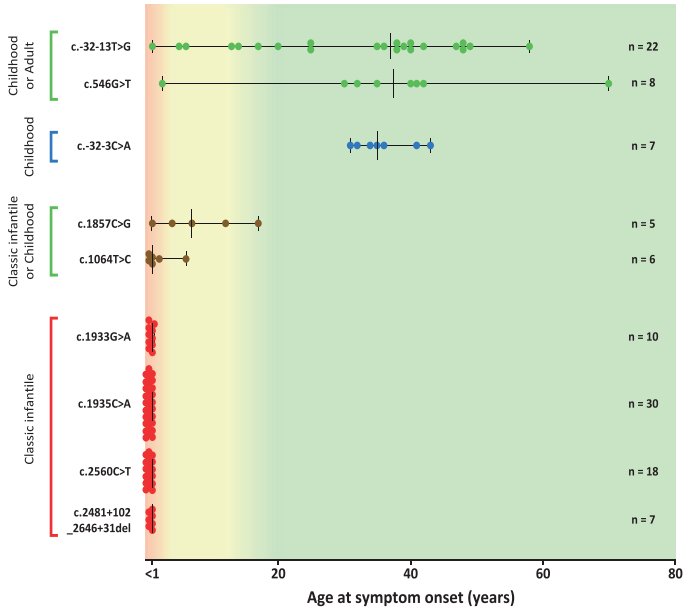


FIGURE 2 Variation in the age of symptom onset in homozygous patients. Age at symptom onset is indicated for patients who carry the most frequently occurring variants at homozygous state (minimum of four patients). Clinical phenotype groups are shown. Median ages at symptom onset are indicated by long vertical lines; ranges by short vertical lines

residual enzyme activity, which is just above the disease threshold of 20% (threshold value according to patient ranges used at the Erasmus MC assessed in fibroblasts with 4MU as substrate). This leads to the prediction that the majority of homozygous c.-32-13T > G patients remain asymptomatic. Indeed, it was previously noted that the frequency of the common c.-32-13T > G variant in the Dutch population is 1 in 154, which implies the existence of 170 individuals in the Netherlands that carry c.-32-13T > G in a homozygous state (Ausems et al., 1999). Currently, our center does not see any patients with this genotype, which suggests that it most often does not lead to a disease phenotype. Others have drawn similar conclusions (Musumeci et al., 2015; Rairikar et al., 2017; Semplicini et al., 2018). The broad range of age at symptom onset in the present analysis formally excluded the possibility that the second allele was responsible for the broad range of age at symptom onset in compound heterozygous c.-32-13T > G patients in this study and implies the involvement of modifying factors that delay or accelerate disease progression in c.-32-13T > G patients, as suggested previously (Herzog et al., 2012; Ko et al., 1999; M. Kroos et al., 2012; M. A. Kroos et al., 2007; Musumeci et al., 2015; Rairikar et al., 2017; Semplicini et al., 2018; Shieh & Lin, 1998; van der Ploeg & Reuser, 2008; Wens et al., 2013). A previous analysis of homozygous c.-32-13T > G patients showed that the c.510C > T variant acts as such a modifying factor and is overrepresented in patients with this genotype (Bergsma et al., 2019). The c.510C > T variant lowers enzymatic activity and can lead to a residual GAA enzymatic activity below the disease threshold. Given the spread of homozygous c.-32-13T > G patients that develop symptoms across several European countries, the United States, Colombia, Algeria, and Turkey (Table 2), it will be interesting to assess whether genetic modifiers, such as c.510C > T, have a particular geographical distribution. Because the minor allele frequency of c.510C > T is too low to be reported, this will require a more extensive genetic analysis of a larger cohort of patients.

A similar conclusion can be drawn for other variants that were not associated with the classic infantile phenotype. In the case of c.546G > T, the median age at symptom onset in homozygous patients was higher than in compound heterozygous patients in combination with a null allele (37.5 vs. 17 years, respectively) but the difference in age between the earliest and latest onset of disease was similar (67.5 vs. 72 years, respectively). The c.-32-3C > A splicing variant was associated with childhood onset when combined with a null allele, but with onset at adulthood when present at homozygous state (median 35, range 31–43 years). The range of age at symptom onset in patients with compound heterozygous c.-32-3C > A in combination with a null allele could not be assessed with certainty since there have been insufficient patients reported with this *GAA* genotype. Patients with the c.1857C > G (p.(Ser619Arg)) missense variant had a broad range of age at symptom onset in homozygous patients that ranged from <1 to 17 years (median 7 years), while it was only found in early childhood patients (median <1 year, range 2 years) when present at compound

TABLE 2 Information on variants and patients described in Figure 2

DNA/protein nomenclature	Predicted effect	Phenotype with a null allele	Total number of patients	Phenotype of the patients	Population[a]
c.-32-13T>G	Splicing	Childhood or adult	47	Childhood (7) Adult (16) Unknown (15)	German (10), French (3), Italian (3), Portuguese (2), United States (2), Colombian (1), Caucasian (1), Algerian (1), Turkish (1), Unknown (14)
c.546G>T	Splicing	Childhood	8	Childhood (1) Adult (7)	Japanese (8)
c.-32-3C>A	Splicing	Childhood	7	Adult (7)	Brazilian (7)
c.1857C>G, p.(Ser619Arg)	Missense	Classic infantile or childhood	5	Classic infantile (1) Childhood (4)	Japanese (5)
c.1064T>C, p.(Leu355Pro)	Missense	Classic infantile or childhood	10	Classic infantile (4) Childhood (6)	Italian (3), Arab (3), Syrian (1), Colombian (1), Portuguese (1), Unknown (1)
c.2015G>A, p.(Arg672Gln)	Missense	Classic infantile or childhood	5	Classic infantile (1) Childhood (4)	Japanese (3), Chinese (1), Unknown (1)
c.1933G>A, p.(Asp645Asn)	Missense	Classic infantile	11	Classic infantile (10) Childhood (1)	United States (4), Italian (3), Indian (2), Unknown (2)
c.1935C>A, p.(Asp645Glu)	Missense	Classic infantile	37	Classic infantile (34) Childhood (2) Unknown (1)	Taiwanese (29), Thai (3), Chinese (2), Japanese (1), Unknown (2)
c.2560C>T, p.(Arg854*)	Nonsense	Classic infantile	23	Classic infantile (17) Unknown (6)	United States (12), Brazilian (4), African (2), Pakistani (1), Arab (1), Nigerian (1), French Guianese (1), Unknown (1)
c.525del	Frameshift	Classic infantile	11	Classic infantile (11)	Dutch (6), Italian (2), Caucasian (2), Unknown (1)

[a]Population is reported as indicated in the original publication. In certain cases, only Caucasian origin has been reported.

heterozygous state in combination with a null allele (Fukuhara et al., 2018). The c.1064T > C (p.(Leu355Pro)) variant was found in six homozygous patients but the information in compound heterozygous patients who carried a null allele was lacking; homozygous patients showed a median age at symptom onset of <1 year (range <1–6 years).

Confirming their classification as null alleles, the variants c.1935C > A (p.(Asp645Glu)), c.2560C > T (p.(Arg854*)), and c.2481 + 102_2646 + 31del showed invariable symptom onset before the age of 1 year when present at the homozygous state or at compound heterozygous state. In addition, 10 patients were homozygous for the null allele c.1933G > A (p.(Asp645Asn)), and all patients had an age at symptom onset of <1 year.

We conclude from these analyses that the common *GAA* variants examined display phenotypic variation in age at symptom onset of up to 71 years between the youngest and the oldest patient when combined with a null allele and/or when present at a homozygous state. Broad phenotypic variation with respect to age at symptom onset and severity of disease was seen both in patients with splicing variants (c.-32-13T > G, c.546G > T, c.1076-22T > G, c.-32-3C > A) and missense variants (c.2238G > C (p.(Trp746Cys)), c.2173C > T (p.(Arg725Trp)), c.1309C > T (p.(Arg437Cys)), c.1857C > G (p.(Ser619Arg)), c.1064T > C (p.(Leu355-Pro))) and was independent of ethnicity. This suggests that disease-associated and potentially variant-specific modifying factors exist for Pompe disease that can delay or accelerate the progression of the disease course.

3.3 | Phenotypic variation in patients with the c.-32-13T > G variant: Effect of c.510C > T and of the second allele

We previously reported that the *GAA* c.510C > T variant is a silent variant that occurs in cis with the c.-32-13T > G variant in a subset of c.-32-13T > G patients. The c.510C > T variant was found in 8/136 compound

heterozygous c.-32-13T > G patients and in 2/4 homozygous c.-32-13T > G patients (Bergsma et al., 2019). Here, we report three additional c.-32-13T > G/null allele patients with Pompe disease who carried c.510C > T and who developed symptoms at a relatively young age during childhood at 3, 9, and 12 years of age. GAA enzyme activities in fibroblasts were determined for two of the three patients and averaged 8.9 nmol/h/mg, which was lower than in fibroblasts from patients with c.-32-13T > G/null *GAA* genotypes that lacked c.510C > T (12.4 nmol/h/mg) (Bergsma et al., 2019). Our findings support our earlier conclusion that the c.510C > T *GAA* variant is associated with a childhood disease presentation of symptoms in patients with the c.-32-13T > G/null allele genotype. The updated ages at symptom onset in compound heterozygous c.-32-13T > G + c.510C > T patients collected from the Pompe disease *GAA* variant database are plotted in Figure 3. Seven of these patients carried a null allele as the second allele. The median age at symptom onset of these patients was 3 years (range <1–13), similar to what was reported previously (Bergsma et al., 2019).

To assess a possible effect of the second allele in the total population of compound heterozygous c.-32-13T > G patients present in the database, we analyzed symptom onset in all patients with the same second allele (c.525del), any null allele, an unknown (disease-associated) allele, and a verified non-null allele (i.e., an allele that is confirmed to produce at least some residual enzymatic activity). Importantly and in agreement with previous reports (Wens et al., 2013), patients with the *GAA* c.-32-13T > G/c.525del genotype showed a similar median age at symptom onset and age range (median 30 years for both groups and an age range of <1–70 and <1–71 years, respectively). This confirmed that phenotypic variation in either patient group was not caused by differences in severity of the second allele (Bergsma et al., 2019; Wens et al., 2013). Patients with the genotype c.-32-13T > G/unknown (disease-associated) also showed a similar median age at symptom onset (28.5 years) and range (79 years). This suggested that the majority of unknown (disease-associated) variants in the database are null alleles. Identification of classic infantile patients that carry one of these alleles is required to confirm this. c.-32-13T > G patients in which the second allele was not a null allele showed a slightly higher median age at symptom onset of 37 years compared to c.-32-13T > G patients with a null allele (age range <1–70) (Figure 3).

We conclude that the very broad range of age at symptom onset in patients with the c.-32-13T > G variant is not caused by the second allele, and that most c.-32-13T > G patients carry a second null allele. Furthermore, when the c.510C > T variant occurs in cis with c.-32-13T > G, it acts as a negative modifying factor that is associated with an onset of symptoms during childhood.

3.4 | In search of additional modifying factors for c.-32-13T > G patients

The c.510C > T variant occurs in a subset of c.-32-13T > G patients and lowers GAA enzyme activity by worsening splicing outcomes. The presence of the c.510C > T variant predicted symptom onset at childhood, but not all patients with childhood onset carried c.510C > T (Bergsma et al., 2019). This suggested that additional modifying factors exist that may explain childhood presentation in c.-32-13T > G patients that do not carry c.510C > T. To test whether such putative modifying factors would act by lowering GAA enzyme activity, like c.510C > T does, we compared GAA enzyme activities in fibroblasts from compound heterozygous c.-32-13T > G patients with childhood onset but without c.510C > T with compound heterozygous patients with adulthood onset without c.510C > T (Figure 4). This showed no difference in GAA enzyme activity between the two groups. We conclude that putative modifying factors associated with early onset of symptoms other than c.510C > T do not lower GAA enzyme activity in fibroblasts at the group level, suggesting that the majority of such putative modifying factors likely act in trans, for the following reason. While cis-acting

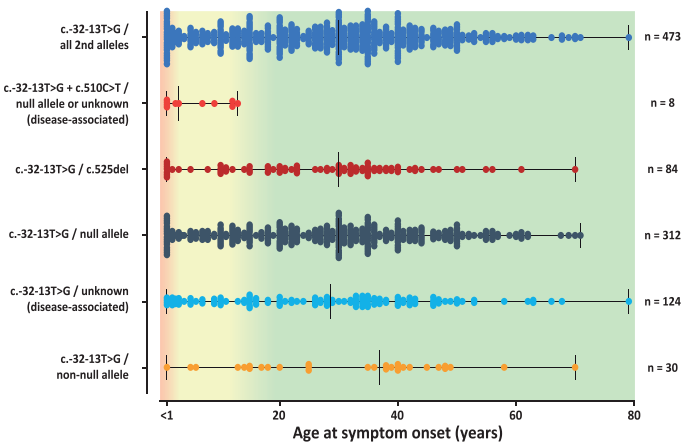


FIGURE 3 Effects of the second allele in patients carrying the c.-32-13T>G variant. Age range is indicated with a black horizontal line. Median ages at symptom onset are indicated by long vertical lines; ranges by short vertical lines

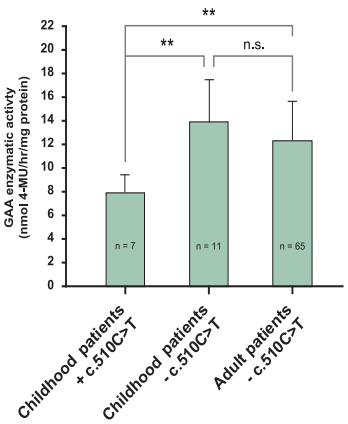


FIGURE 4 GAA enzyme activity in fibroblasts from patients with the c.-32-13T>G variant and childhood or adulthood onset in the absence or presence of the modifier c.510C>T. All patients carried the c.-32-13T>G variant on one allele, and a deleterious variant on the second allele. Values represent mean ± SD. n.s., not significant; $^{**}p < .01$ (*t* test)

genetic factors could still play a role in patients lacking c.510C > T, such putative factors, for example, distant enhancers, can be expected to alter *GAA* expression, which would be reflected at the level of GAA enzyme activity. The fact that GAA enzyme activity was not different between patients with childhood onset that lack c.510C > T and with adulthood onset that lack c.510C > T led us to conclude that at the group level, trans-acting factors that act downstream of regulating GAA activity, for example, factors that regulate lysosomal or skeletal muscle homeostasis, seem to be the most likely explanation for the differences in symptom onset among such patients. This conclusion does not exclude the possibility that for individual patients, cis-acting factors could still play a role. In addition, our conclusion relies on the assumption that enzyme activity in fibroblasts in vitro is a good proxy of enzyme activity in skeletal muscles in vivo, which remains to be determined. It will be interesting to verify these results in muscle cells derived from primary biopsies or patient-derived induced pluripotent stem cells (van der Wal et al., 2018) because muscle cells rather than fibroblasts better represent the primary affected cell type. Genome-wide studies using large patient cohorts will help to identify putative modifying factors, and advanced in vitro model systems, such as 2D or 3D models, derived from induced pluripotent stem cells along with gene-corrected controls (Iuliano et al., 2020; van der Wal et al., 2018) are required to unequivocally demonstrate their biological effect and to study their mechanism of action. We speculate that candidate modifying factors may act in trans and may modulate cellular pathways that are known to be involved in the progression of Pompe disease, such as glucose metabolism, lysosomal biogenesis, autophagy, and skeletal muscle strength and endurance.

4 | CONCLUSION

Previous work has reported on the large variability in age at symptom onset of patients with the c.-32-13T > G *GAA* variant. Here, we compared disease onset of patients carrying the c.-32-13T > G variant with patients carrying other variants and conclude that nine distinct *GAA* disease-associated variants (c.2647-7G > A; c.546G > T; c.2238G > C (p.(Trp746Cys)); c.2173C > T (p.(Arg725Trp)); c.1076-22T > G; c.1309C > T (p.(Arg437Cys)); c.-32-3C > A; c.1857C > G (p.(Ser619Arg)); c.1064T > C (p.(Leu355Pro))) are associated with a large variation in age at symptom onset when present at compound heterozygous state in combination with a null allele and/or at homozygous state. Altogether, these findings suggest a strong influence of modifying factors that affect the age at symptom onset in patients with Pompe disease. Furthermore, additional patients carrying c.510C > T in cis with the c.-32-13T > G variant were reported, confirming that the c.510C > T variant is a disease modifier that accelerates age at symptom onset when combined with the c.-32-13T > G variant. Importantly, the obtained evidence showed that the majority of putative additional genetic modifiers besides c.510C > T do not lower GAA enzyme activity in fibroblasts, suggesting that these putative modifiers act in trans.

ACKNOWLEDGMENTS

We thank Dr. Arnold J. J. Reuser for his critical review of the manuscript and Ton Verkerk for his invaluable assistance with generating the database. We thank Philip Lijnzaad for his help with the statistical analysis. This study was supported by grants from the Administrative Department of Science, Technology and Innovation (Colciencias, Colombia) and Sanofi Genzyme. The collaboration project is co-funded by the PPP Allowance made available by Health-Holland, Top Sector Life Sciences & Health, to the Prinses Beatrix Spierfonds to stimulate public–private partnerships (project numbers: LSHM17075 and LSHM19015). Research on Pompe disease at Erasmus MC is financially supported by "Prinses Beatrix Spierfonds" (project numbers: W.OR13-21, W.OR15-10, W.OR16-07); Tex Net; Sophia Foundation for Medical Research (SSWO) (project number: s17-32); Metakids (project number: 2016-063); Conselho Nacional de Desenvolvimento Científico e Tecnológico–"National Counsil of Technological and Scientific Development," Brasil (grant number: 234407/2014-0); Colciencias and Sanofi Genzyme. Ans T. van der Ploeg 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.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

DATA AVAILABILITY STATEMENT

The data described in this study are freely accessible in the Pompe disease *GAA* variant database at http://www.pompevariantdatabase.nl

WEB RESOURCE

Pompe disease *GAA* variant database: http://www.pompevariantdatabase.nl

ORCID

Atze J. Bergsma http://orcid.org/0000-0002-8733-5078

W. W. M. Pim Pijnappel http://orcid.org/0000-0002-7042-2482

REFERENCES

Amalfitano, A., Bengur, A. R., Morse, R. P., Majure, J. M., Case, L. E., Veerling, D. L., Mackey, J., Kishnani, P., Smith, W., McVie-Wylie, A., Sullivan, J. A., Hoganson, G. E., Phillips, J. A., Schaefer, G. B., Charrow, J., Ware, R. E., Bossen, E. H., & Chen, Y. T. (2001). Recombinant human acid alpha-glucosidase enzyme therapy for infantile glycogen storage disease type II: Results of a phase I/II clinical trial. *Genetics in Medicine*, *3*(2), 132–138.

Ausems, M. G., ten Berg, K., Sandkuijl, L. A., Kroos, M. A., Bardoel, A. F., Roumelioti, K. N., Reuser, A. J., Sinke, R., & Wijmenga, C. (2001). Dutch patients with glycogen storage disease type II show common ancestry for the 525delT and del exon 18 mutations. *Journal of Medical Genetics*, *38*(8), 527–529.

Ausems, M. G., Verbiest, J., Hermans, M. P., Kroos, M. A., Beemer, F. A., Wokke, J. H., Sandkuijl, L. A., Reuser, A. J., & van der Ploeg, A. T. (1999). Frequency of glycogen storage disease type II in The Netherlands: Implications for diagnosis and genetic counselling. *European Journal of Human Genetics*, *7*(6), 713–716.

Becker, J. A., Vlach, J., Raben, N., Nagaraju, K., Adams, E. M., Hermans, M. M., Reuser, A. J., Brooks, S. S., Tifft, C. J., Hirschhorn, R., Huie, M. L., Nicolino, M., & Plotz, P. H. (1998). The African origin of the common mutation in African American patients with glycogen-storage disease type II. *American Journal of Human Genetics*, *62*(4), 991–994. https://doi.org/10.1086/301788

Bergsma, A. J., in 't Groen, S., van den Dorpel, J., van den Hout, H., van der Beek, N., Schoser, B., Toscano, A., Musumeci, O., Bembi, B., Dardis, A., Morrone, A., Tummolo, A., Pasquini, E., van der Ploeg, A. T., & Pijnappel, W. W. M. P. (2019). A genetic modifier of symptom onset in Pompe disease. *EBioMedicine*, *43*, 553–561. https://doi.org/10.1016/j.ebiom.2019.03.048

Bergsma, A. J., in 't Groen, S. L. M., Catalano, F., Yamanaka, M., Takahashi, S., Okumiya, T., & Pijnappel, W. W. M. P. (2021). A generic assay for the identification of splicing variants that induce nonsense-mediated decay in Pompe disease. *European Journal of Human Genetics*, *29*, 422–433.

Boerkoel, C. F., Exelbert, R., Nicastri, C., Nichols, R. C., Miller, F. W., Plotz, P. H., & Raben, N. (1995). Leaky splicing mutation in the acid maltase gene is associated with delayed onset of glycogenosis type II. *American Journal of Human Genetics*, *56*(4), 887–897.

Dagnino, F., Stroppiano, M., Regis, S., Bonuccelli, G., & Filocamo, M. (2000). Evidence for a founder effect in Sicilian patients with glycogen storage disease type II. *Human Heredity*, *50*(6), 331–333.

de Faria, D. O. S., t Groen, S., Hoogeveen-Westerveld, M., Nino, M. Y., van der Ploeg, A. T., Bergsma, A. J., & Pijnappel, W. W. M. P. (2021). Update of the Pompe variant database for the prediction of clinical phenotypes: Novel disease-associated variants, common sequence variants, and results from newborn screening. *Human Mutation*, *42*(2), 119–134. https://doi.org/10.1002/humu.24148

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. (2016). HGVS recommendations for the description of sequence variants: 2016 update. *Human Mutation*, *37*(6), 564–569. https://doi.org/10.1002/humu.22981

Fukuhara, Y., Fuji, N., Yamazaki, N., Hirakiyama, A., Kamioka, T., Seo, J. H., Mashima, R., Kosuga, M., & Okuyama, T. (2018). A molecular analysis of the GAA gene and clinical spectrum in 38 patients with Pompe disease in Japan. *Molecular Genetics and Metabolism Reports*, *14*, 3–9. https://doi.org/10.1016/j.ymgmr.2017.10.009

Gaeta, M., Musumeci, O., Mondello, S., Ruggeri, P., Montagnese, F., Cucinotta, M., Vinci, S., Milardi, D., & Toscano, A. (2015). Clinical and pathophysiological clues of respiratory dysfunction in late-onset Pompe disease: New insights from a comparative study by MRI and respiratory function assessment. *Neuromuscular Disorders*, *25*(11), 852–858. https://doi.org/10.1016/j.nmd.2015.09.003

Gungor, D., & Reuser, A. J. (2013). How to describe the clinical spectrum in Pompe disease? *American Journal of Medical Genetics. Part A*, *161A*(2), 399–400. https://doi.org/10.1002/ajmg.a.35662

Hermans, M. M., de Graaff, E., Kroos, M. A., Wisselaar, H. A., Willemsen, R., Oostra, B. A., & Reuser, A. J. (1993). The conservative substitution Asp-645→Glu in lysosomal alpha-glucosidase affects transport and phosphorylation of the enzyme in an adult patient with glycogen-storage disease type II. *Biochemical Journal*, *289*(Pt 3), 687–693. http://www.ncbi.nlm.nih.gov/pubmed/8094613

Herzog, A., Hartung, R., Reuser, A. J., Hermanns, P., Runz, H., Karabul, N., Gökce, S., Pohlenz, J., Kampmann, C., Lampe, C., Beck, M., & Mengel, E. (2012). A cross-sectional single-centre study on the spectrum of Pompe disease, German patients: Molecular analysis of the GAA gene, manifestation and genotype-phenotype correlations. *Orphanet Journal of Rare Diseases*, *7*(1), 35. https://doi.org/10.1186/1750-1172-7-35

Huie, M. L., Chen, A. S., Tsujino, S., Shanske, S., DiMauro, S., Engel, A. G., & 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. *Human Molecular Genetics*, *3*(12), 2231–2236.

Iuliano, A., Van der, Wal, Ruijmbeek, E., C. W. B., in't Groen, S. L. M., Pijnappel, W. W. M. P., de Greef, J. C., & Saggiomo, V. (2020). Coupling 3D printing and novel replica molding for in house fabrication of skeletal muscle tissue engineering devices. *Advanced Materials Technologies*, *121*, 800. https://doi.org/10.1002/admt.202000344

Kishnani, P. S., Corzo, D., Nicolino, M., Byrne, B., Mandel, H., Hwu, W. L., Leslie, N., Levine, J., Spencer, C., McDonald, M., Li, J., Dumontier, J., Halberthal, M., Chien, Y. H., Hopkin, R., Vijayaraghavan, S., Gruskin, D., Bartholomew, D., van der Ploeg, A., … Wraith, J. E. (2007). Recombinant human acid α-glucosidase: Major clinical benefits in infantile-onset Pompe disease. *Neurology*, *68*(2), 99–109. https://doi.org/10.1212/01.wnl.0000251268.41188.04

Kishnani, P. S., Hwu, W. L., Mandel, H., Nicolino, M., Yong, F., Corzo, D., & Infantile-Onset Pompe Disease Natural History Study Group. (2006). A retrospective, multinational, multicenter study on the natural history of infantile-onset Pompe disease. *Journal of Pediatrics*, *148*(5), 671–676. https://doi.org/10.1016/j.jpeds.2005.11.033

Ko, T. M., Hwu, W. L., Lin, Y. W., Tseng, L. H., Hwa, H. L., Wang, T. R., & Chuang, S. M. (1999). Molecular genetic study of Pompe disease in Chinese patients in Taiwan. *Human Mutation*, *13*(5), 380–384.

Kroos, M., Hoogeveen-Westerveld, M., van der Ploeg, A., & Reuser, A. J. (2012). The genotype-phenotype correlation in Pompe disease. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics*, *160C*(1), 59–68. https://doi.org/10.1002/ajmg.c.31318

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., Halley, D. J., Van der Ploeg, A. T., & Reuser, A. J. (2007). Broad spectrum of Pompe disease in patients with the same c.-32-13T→G haplotype. *Neurology*, *68*(2), 110–115. https://doi.org/10.1212/01.wnl.0000252798.25690.76

Kroos, M. A., Van der Kraan, M., Van Diggelen, O. P., Kleijer, W. J., Reuser, A. J., Van den Boogaard, M. J., Ausems, M. G., Ploos van Amstel, H. K., Poenaru, L., & Nicolino, M. (1995). Glycogen

storage disease type II: Frequency of three common mutant alleles and their associated clinical phenotypes studied in 121 patients. *Journal of Medical Genetics*, *32*(10), 836–837.

Laforêt, P., Laloui, K., Granger, B., Hamroun, D., Taouagh, N., Hogrel, J. Y., Bouhour, F., Lacour, A., Salort-Campana, E., Penisson-Besnier, I., Sacconi, S., Zagnoli, F., Chapon, F., Eymard, B., Desnuelle, C., Pouget, J., & French Pompe Registry Study Group. (2013). The French Pompe registry. Baseline characteristics of a cohort of 126 patients with adult Pompe disease. *Revue Neurologique (Paris)*, *169*(8-9), 595–602. https://doi.org/10.1016/j.neurol.2013.07.002

Laforet, P., Nicolino, M., Eymard, P. B., Puech, J. P., Caillaud, C., Poenaru, L., & Fardeau, M. (2000). Juvenile and adult-onset acid maltase deficiency in France: Genotype-phenotype correlation. *Neurology*, *55*(8), 1122–1128.

Lin, C. Y., & Shieh, J. J. (1996). Molecular study on the infantile form of Pompe disease in Chinese in Taiwan. *Chung-Hua Min Kuo Hsiao Erh Ko i Hsueh Hui Tsa Chih*, *37*(2), 115–121.

Liu, X., Wang, Z., Jin, W., Lv, H., Zhang, W., Que, C., Huang, Y., & Yuan, Y. (2014). Clinical and GAA gene mutation analysis in mainland Chinese patients with late-onset Pompe disease: Identifying c.2238G > C as the most common mutation. *BMC Medical Genetics*, *15*(1), 141. https://doi.org/10.1186/s12881-014-0141-2

Maimaiti, M., Takahashi, S., Okajima, K., Suzuki, N., Ohinata, J., Araki, A., Tanaka, H., Mukai, T., & Fujieda, K. (2009). Silent exonic mutation in the acid-α-glycosidase gene that causes glycogen storage disease type II by affecting mRNA splicing. *Journal of Human Genetics*, *54*(8), 493–496. https://doi.org/10.1038/jhg.2009.66

Montalvo, A. L., Bembi, B., Donnarumma, M., Filocamo, M., Parenti, G., Rossi, M., Merlini, L., Buratti, E., De Filippi, P., Dardis, A., Stroppiano, M., Ciana, G., & Pittis, M. G. (2006). Mutation profile of the GAA gene in 40 Italian patients with late onset glycogen storage disease type II. *Human Mutation*, *27*(10), 999–1006. https://doi.org/10.1002/humu.20374

Müller-Felber, W., Horvath, R., Gempel, K., Podskarbi, T., Shin, Y., Pongratz, D., Walter, M. C., Baethmann, M., Schlotter-Weigel, B., Lochmüller, H., & Schoser, B. (2007). Late onset Pompe disease: Clinical and neurophysiological spectrum of 38 patients including long-term follow-up in 18 patients. *Neuromuscular Disorders*, *17*(9–10), 698–706.

Musumeci, O., Thieme, A., Claeys, K. G., Wenninger, S., Kley, R. A., Kuhn, M., Lukacs, Z., Deschauer, M., Gaeta, M., Toscano, A., Gläser, D., & Schoser, B. (2015). Homozygosity for the common GAA gene splice site mutation c.-32-13T > G in Pompe disease is associated with the classical adult phenotypical spectrum. *Neuromuscular Disorders*, *25*(9), 719–724.

Nicolino, M., Byrne, B., Wraith, J. E., Leslie, N., Mandel, H., Freyer, D. R., Arnold, G. L., Pivnick, E. K., Ottinger, C. J., Robinson, P. H., Loo, J. C., Smitka, M., Jardine, P., Tatò, L., Chabrol, B., McCandless, S., Kimura, S., Mehta, L., Bali, D., … Kishnani, P. S. (2009). Clinical outcomes after long-term treatment with alglucosidase alfa in infants and children with advanced Pompe disease. *Genetics in Medicine*, *11*(3), 210–219. https://doi.org/10.1097/GIM.0b013e31819d0996

Niño, M. Y., In 't Groen, S., Bergsma, A. J., van der Beek, N., Kroos, M., Hoogeveen-Westerveld, M., van der Ploeg, A. T., & Pijnappel, W. W. M. P. (2019). Extension of the Pompe mutation database by linking disease-associated variants to clinical severity. *Human Mutation*, *40*(11), 1954–1967. https://doi.org/10.1002/humu.23854

Pittis, M. G., & Filocamo, M. (2007). Molecular genetics of late onset glycogen storage disease II in Italy. *Acta Myologica: Myopathies and Cardiomyopathies*, *26*(1), 67–71.

Rairikar, M. V., Case, L. E., Bailey, L. A., Kazi, Z. B., Desai, A. K., Berrier, K. L., Coats, J., Gandy, R., Quinones, R., & Kishnani, P. S. (2017). Insight into the phenotype of infants with Pompe disease identified by newborn screening with the common c.-32-13T > G "late-onset" GAA variant. *Molecular Genetics and Metabolism*, *122*(3), 99–107.

Reuser, A. J. J., Hirschhorn, R., & Kroos, M. A. (2018). Pompe disease: Glycogen storage disease type II, acid α-glucosidase (acid maltase) deficiency, *The online metabolic and molecular bases of inherited disease* (8th ed.). McGraw-Hill Medical.

Sampaolo, S., Esposito, T., Farina, O., Formicola, D., Diodato, D., Gianfrancesco, F., Cipullo, F., Cremone, G., Cirillo, M., Del Viscovo, L., Toscano, A., Angelini, C., & Di Iorio, G. (2013). Distinct disease phenotypes linked to different combinations of GAA mutations in a large late-onset GSDII sibship. *Orphanet Journal of Rare Diseases*, *8*, 159. https://doi.org/10.1186/1750-1172-8-159

Semplicini, C., Letard, P., De Antonio, M., Taouagh, N., Perniconi, B., Bouhour, F., Orlikowski, D., Sacconi, S., Salort-Campana, E., Solé, G., Zagnoli, F., Hamroun, D., Froissart, R., Caillaud, C., Laforêt, P., & French Pompe Study Group. (2018). Late-onset Pompe disease in France: Molecular features and epidemiology from a nationwide study. *Journal of Inherited Metabolic Disease*, *41*(6), 937–946. https://doi.org/10.1007/s10545-018-0243-7

Shieh, J. J., & Lin, C. Y. (1998). Frequent mutation in Chinese patients with infantile type of GSD II in Taiwan: Evidence for a founder effect. *Human Mutation*, *11*(4), 306–312.

Strothotte, S., Strigl-Pill, N., Grunert, B., Kornblum, C., Eger, K., Wessig, C., Deschauer, M., Breunig, F., Glocker, F. X., Vielhaber, S., Brejova, A., Hilz, M., Reiners, K., Müller-Felber, W., Mengel, E., Spranger, M., & Schoser, B. (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. *Journal of Neurology*, *257*(1), 91–97. https://doi.org/10.1007/s00415-009-5275-3

van Capelle, C. I., van der Beek, N. A., Hagemans, M. L., Arts, W. F., Hop, W. C., Lee, P., Jaeken, J., Frohn-Mulder, I. M., Merkus, P. J., Corzo, D., Puga, A. C., Reuser, A. J., & van der Ploeg, A. T. (2010). Effect of enzyme therapy in juvenile patients with Pompe disease: A three-year open-label study. *Neuromuscular Disorders*, *20*(12), 775–782. https://doi.org/10.1016/j.nmd.2010.07.277

van den Hout, H., Reuser, A. J., Vulto, A. G., Loonen, M. C., Cromme-Dijkhuis, A., & Van der Ploeg, A. T. (2000). Recombinant human α-glucosidase from rabbit milk in Pompe patients. *Lancet*, *356*(9227), 397–398. https://doi.org/10.1016/s0140-6736(00)02533-2

van den Hout, H. M., Hop, W., van Diggelen, O. P., Smeitink, J. A., Smit, G. P., Poll-The, B. T., Bakker, H. D., Loonen, M. C., de Klerk, J. B., Reuser, A. J., & van der Ploeg, A. T. (2003). The natural course of infantile Pompe's disease: 20 original cases compared with 133 cases from the literature. *Pediatrics*, *112*(2), 332–340. https://doi.org/10.1542/peds.112.2.332

van den Hout, J. M., Reuser, A. J., de Klerk, J. B., Arts, W. F., Smeitink, J. A., & Van der Ploeg, A. T. (2001). Enzyme therapy for Pompe disease with recombinant human α-glucosidase from rabbit milk. *Journal of Inherited Metabolic Disease*, *24*(2), 266–274. https://doi.org/10.1023/a:1010383421286

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., Notermans, N. C., Faber, K. G., Verschuuren, J. J., Reuser, A. J., van der Ploeg, A. T., & van Doorn, P. A. (2012). Clinical features and predictors for disease natural progression in adults with Pompe disease: A nationwide prospective observational study. *Orphanet Journal of Rare Diseases*, *7*, 88. https://doi.org/10.1186/1750-1172-7-88

van der Beek, N. A., Hagemans, M. L., Reuser, A. J., Hop, W. C., Van der Ploeg, Van Doorn, A. T., P. A., & Wokke, J. H. (2009). Rate of disease progression during long-term follow-up of patients with late-onset Pompe disease. *Neuromuscular Disorders*, *19*(2), 113–117. https://doi.org/10.1016/j.nmd.2008.11.007

Van der Kraan, M., Kroos, M. A., Joosse, M., Bijvoet, A. G., Verbeet, M. P., Kleijer, W. J., & Reuser, A. J. (1994). Deletion of exon 18 is a frequent mutation in glycogen storage disease type II. *Biochemical & Biophysical Research Communications*, *203*(3), 1535–1541.

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., Lange, D. J., Leshner, R. T., Mayhew, J. E., Morgan, C., Nozaki, K., Park, D. J., Pestronk, A., Rosenbloom, B., Skrinar, A., … Zivkovic, S. A. (2010). A randomized study of alglucosidase alfa in late-onset Pompe's disease. *New England Journal of Medicine*, *362*(15), 1396–1406. https://doi.org/10.1056/NEJMoa0909859

van der Ploeg, A. T., Kruijshaar, M. E., Toscano, A., Laforêt, P., Angelini, C., Lachmann, R. H., Pascual Pascual, S. I., Roberts, M., Rösler, K., Stulnig, T., van Doorn, P. A., Van den Bergh, P., Vissing, J., Schoser, B., & European Pompe, C. (2017). European consensus for starting and stopping enzyme replacement therapy in adult patients with Pompe disease: A 10-year experience. *European Journal of Neurology*, *24*(6), 768–e31. https://doi.org/10.1111/ene.13285

van der Ploeg, A. T., & Reuser, A. J. (2008). Pompe's disease. *Lancet*, *372*(9646), 1342–1353. https://doi.org/10.1016/S0140-6736(08)61555-X

van der Wal, E., Bergsma, A. J., van Gestel, T., In 't Groen, S., Zaehres, H., Araúzo-Bravo, M. J., Schöler, H. R., van der Ploeg, A. T., & Pijnappel, W. W. M. P. (2017). GAA deficiency in Pompe disease is alleviated by exon inclusion in iPSC-derived skeletal muscle cells. *Molecular Therapy. Nucleic Acids*, *7*, 101–115. https://doi.org/10.1016/j.omtn.2017.03.002

van der Wal, E., Herrero-Hernandez, P., Wan, R., Broeders, M., In 't Groen, S., van Gestel, T., van IJcken, W., Cheung, T. H., van der Ploeg, A. T., Schaaf, G. J., & Pijnappel, W. W. M. P. (2018). Large-scale expansion of human iPSC-derived skeletal muscle cells for disease modeling and cell-based therapeutic strategies. *Stem Cell Reports*, *10*(6), 1975–1990. https://doi.org/10.1016/j.stemcr.2018.04.002

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., & Brusse, E. (2013). Phenotypical variation within 22 families with Pompe disease. *Orphanet Journal of Rare Diseases*, *8*(1), 182. https://doi.org/10.1186/1750-1172-8-182

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., & Ploos van Amstel, H. K. (1995). Genotype-phenotype correlation in adult-onset acid maltase deficiency. *Annals of Neurology*, *38*(3), 450–454.

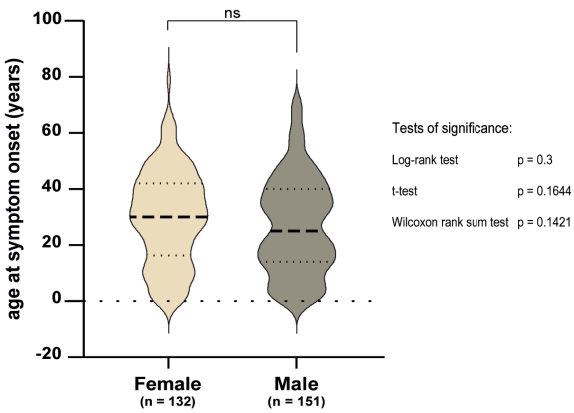
SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Niño, M. Y., in't Groen, S. L. M., de Faria, D. O. S., Hoogeveen-Westerveld, M., van den Hout, H. J. M. P., van der Ploeg, A. T., Bergsma, A. J., & Pijnappel, W. W. M. P. (2021). Broad variation in phenotypes for common *GAA* genotypes in Pompe disease. *Human Mutation*, 1–12. https://doi.org/10.1002/humu.24272

SUPPLEMENTARY DATA

Supplementary Figure S1



Supplementary Figure S1: Analysis of sex-related differences in age at symptom onset. The dip test indicated that the data was distributed unimodally. The Log-rank test, t-test and Wilcoxon rank sum test were performed to determine significance.

Supplementary Table S1: References to patient information for compound heterozygous variants in combination with a null allele

Variant	Pubmed IDs
c.2647-7G>A	24107549 (Sampaolo, et al., 2013)
c.-32-13T>G	16531044 (Kostera-Pruszczyk, et al., 2006), 16838077 (Kroos, et al., 2006), 16917947 (Montalvo, et al., 2006), 17027861 (Shin, 2006), 17056254 (Palmer, et al., 2007), 17616415 (Gort, et al., 2007), 17643989 (Muller-Felber, et al., 2007), 17723315 (McCready, et al., 2007), 18607768 (Joshi, et al., 2008), 18757064 (van der Beek, et al., 2008), 19588081 (Oba-Shinjo, et al., 2009), 20559845 (Sacconi, et al., 2010), 21550241 (Orlikowski, et al., 2011), 22676651 (Herzog, et al., 2012), 22980766 (Alejaldre, et al., 2012), 23000108 (Wens, et al., 2012), 23160972 (Schneider, et al., 2013), 23430493 (Nino, et al., 2013), 23566438 (Spada, et al., 2013), 24245577 (Wens, et al., 2013), 24685124 (Andreassen, et al., 2014), 25155446 (Horvath, et al., 2015), 25455803 (Vill, et al., 2015), 25526786 (Liu, et al., 2014), 25673129 (Montagnese, et al., 2015), 25681614 (Turaca, et al., 2015), 25783438 (Musumeci, et al., 2016), 25786784 (Echaniz-Laguna, et al., 2015), 25835646 (Hobson-Webb, et al., 2015), 25846667 (Bandyopadhyay, et al., 2015), 26622091 (Tecellioglu and Kamisli, 2015), 26800218 (Angelini, et al., 2016), 26809617 (Levesque, et al., 2016), 26873529 (Stepien, et al., 2016), 27649523 (Nazari, et al., 2017), 27862019 (Taisne, et al., 2017), 27927596 (Schanzer, et al., 2017), 28032299 (Bekircan-Kurt, et al., 2017), 28380188 (Sixel, et al., 2017), 28694071 (Torrealba-Acosta, et al., 2017), 28763149 (McIntosh, et al., 2017), 28838325 (Byrne, et al., 2017), 28856460 (Pichiecchio, et al., 2017), 28951071 (Rairikar, et al., 2017), 29122469 (Mori, et al., 2017), 29143201 (Burlina, et al., 2018), 29149851 (Johnson, et al., 2017), 29181627 (Loscher, et al., 2018), 29742245 (Lorenzoni, et al., 2018), 29880332 (Savarese, et al., 2018), 30155607 (Semplicini, et al., 2018), 30367637 (Szklanny and Tylki-Szymanska, 2018), 30564623 (Nallamilli, et al., 2018), 30697769 (Ebert, et al., 2019), 30922962 (Bergsma, et al., 2019), 31125121 (Jastrzebska, et al., 2019), 31392188 (Alandy-Dy, et al., 2019), 9660056 (Kroos, et al., 1998)
c.2238G>C	18458862 (Wan, et al., 2008), 21232767 (Chien, et al., 2011), 25526786 (Liu, et al., 2014), 27692865 (Chu, et al., 2016), 28433475 (Lee, et al., 2017), 29451150 (Liu, et al., 2018), 30897595 (Lyu, et al., 2019), 31392188 (Alandy-Dy, et al., 2019)
c.2173C>T	17616415 (Gort, et al., 2007), 8401535 (Hermans, et al., 1993b), 3865697 (Trend, et al., 1985), 30155607 (Semplicini, et al., 2018)
c.1076-22T>G	9259196 (Adams, et al., 1997), 17643989 (Muller-Felber, et al., 2007), 25455803 (Vill, et al., 2015), 29181627 (Loscher, et al., 2018)
c.1309C>T	25526786 (Liu, et al., 2014), 12601120 (Lam, et al., 2003), 23884227 (Park, et al., 2013), 21940687 (Cho, et al., 2012), 24169249 (Liu, et al., 2013), 29124014 (Fukuhara, et al., 2018), 27692865 (Chu, et al., 2016), 30360039 (Kim, et al., 2019)
c.546G>T	19609281 (Maimaiti, et al., 2009), 20202878 (Kobayashi, et al., 2010), 25388776 (Park, et al., 2015), 29124014 (Fukuhara, et al., 2018), 28433475 (Lee, et al., 2017)
c.1857C>G	23884227 (Park, et al., 2013), 21676566 (Ishigaki, et al., 2012), 29124014 (Fukuhara, et al., 2018)
c.1935C>A	21039225 (Amarinthnukrowh, et al., 2010), 21232767 (Chien, et al., 2011), 18458862 (Wan, et al., 2008), 16782080 (Dou, et al., 2006), 25466677 (Chien, et al., 2015), 24243590 (Yang, et al., 2014), 8604985 (Shieh and Lin, 1996), 24706590 (Chien, et al., 2014), 24269976 (Fu, et al., 2014), 10338092 (Ko, et al., 1999), 8094613 (Hermans, et al., 1993a), 9554747 (Shieh and Lin, 1998), 11053688 (Tsujino, et al., 2000), 26685070 (Yang, et al., 2016), 29122469 (Mori, et al., 2017), 28394184 (Chen, et al., 2017), 31510962 (Ngiwsara, et al., 2019), 31342611 (Reuser, et al., 2019), 19948615 (Chien, et al., 2009)
c.525del	8558570 (Kroos, et al., 1995), 9660056 (Kroos, et al., 1998), 10377006 (Huie, et al., 1999), 14695532 (Hermans, et al., 2004), 15121988 (Van den Hout, et al., 2004), 18429042 (Pittis, et al., 2008), 21687968 (Manwaring, et al., 2012), 22237443 (Messinger, et al., 2012), 22555271 (Kindel, et al., 2012), 23430912 (Swarr, et al., 2012), 23825616 (Banugaria, et al., 2013), 29122469 (Mori, et al., 2017), 29422078 (Parini, et al., 2018), 30902109 (Poelman, et al., 2019), 31342611 (Reuser, et al., 2019)
c.2560C>T	8094613 (Hermans, et al., 1993a), 9529346 (Becker, et al., 1998), 10528311 (Castro-Gago, et al., 1999), 16860134 (Kishnani, et al., 2006), 17723315 (McCready, et al., 2007), 18535739 (Pereira, et al., 2008), 19588081 (Oba-Shinjo, et al., 2009), 22237443 (Messinger, et al., 2012), 22658377 (Palermo, et al., 2012), 23430493 (Nino, et al., 2013), 23601496 (Elder, et al., 2013), 23825616 (Banugaria, et al., 2013), 24273659 (Alansari, et al., 2013), 29122469 (Mori, et al., 2017), 29390460 (Martinez, et al., 2017), 29637184 (Elenga, et al., 2018), 31342611 (Reuser, et al., 2019)
c.2481+102_ 2646+31del	7603531 (Raben, et al., 1995), 7717400 (Boerkoel, et al., 1995), 7981676 (Huie, et al., 1994), 8558570 (Kroos, et al., 1995), 9660056 (Kroos, et al., 1998), 14695532 (Hermans, et al., 2004), 14972326 (Montalvo, et al., 2004), 15121988 (Van den Hout, et al., 2004), 17616415 (Gort, et al., 2007), 18285536 (Nascimbeni, et al., 2008), 18429042 (Pittis, et al., 2008), 19588081 (Oba-Shinjo, et al., 2009), 27927596 (Schanzer, et al., 2017), 28657663 (Tsai, et al., 2017), 29122469 (Mori, et al., 2017), 29422078 (Parini, et al., 2018), 30902109 (Poelman, et al., 2019), 31342611 (Reuser, et al., 2019)

Supplementary Table S2: References to patient information for homozygous variants

Homozygous variant	Pubmed IDs
c.-32-13T>G	11071489 (Laforet, et al., 2000), 15986226 (Sharma, et al., 2005), 17643989 (Muller-Felber, et al., 2007), 18285536 (Nascimbeni, et al., 2008), 22676651 (Herzog, et al., 2012), 23430493 (Nino, et al., 2013), 23430949 (Wittmann, et al., 2012), 23465405 (Scott, et al., 2013), 25455803 (Vill, et al., 2015), 25677830 (Golden-Grant, et al., 2015), 25786784 (Echaniz-Laguna, et al., 2015), 26231297 (Musumeci, et al., 2015), 28032299 (Bekircan-Kurt, et al., 2017), 28265479 (Sifi, et al., 2017), 28554557 (Almeida, et al., 2017), 28838325 (Byrne, et al., 2017), 28951071 (Rairikar, et al., 2017), 29326002 (Golsari, et al., 2018), 30093709 (Wasserstein, et al., 2019), 30155607 (Semplicini, et al., 2018), 31342611 (Reuser, et al., 2019)
c.546G>T	23465405 (Scott, et al., 2013), 29124014 (Fukuhara, et al., 2018), 30093193 (Hossain, et al., 2018)
c.-32-3C>A	19588081 (Oba-Shinjo, et al., 2009), 20464284 (Grzesiuk, et al., 2010), 25681614 (Turaca, et al., 2015)
c.1857C>G	14643388 (Pipo, et al., 2003), 18495398 (Nabatame, et al., 2009), 29124014 (Fukuhara, et al., 2018)
c.1064T>C	14972326 (Montalvo, et al., 2004), 16917947 (Montalvo, et al., 2006), 17723315 (McCready, et al., 2007), 18429042 (Pittis, et al., 2008), 23430493 (Nino, et al., 2013), 24016645 (Loureiro Neves, et al., 2013), (Al-Hassnan, et al., 2018), 31342611 (Reuser, et al., 2019)
c.2015G>A	9535769 (Huie, et al., 1998), 11053688 (Tsujino, et al., 2000), 28937052 (Quan, et al., 2017), 31342611 (Reuser, et al., 2019)
c.1933G>A	16860134 (Kishnani, et al., 2006), 17723315 (McCready, et al., 2007), (Del Rizzo, et al., 2010), 18285536 (Nascimbeni, et al., 2008), 29122469 (Mori, et al., 2017), 29422078 (Parini, et al., 2018), 31342611 (Reuser, et al., 2019)
c.1935C>A	21039225 (Amarinthnukrowh, et al., 2010), 9554747 (Shieh and Lin, 1998), 11053688 (Tsujino, et al., 2000), 18458862 (Wan, et al., 2008), 21232767 (Chien, et al., 2011), 10338092 (Ko, et al., 1999), 28394184 (Chen, et al., 2017), 31342611 (Reuser, et al., 2019)
c.2560C>T	9529346 (Becker, et al., 1998), 17723315 (McCready, et al., 2007), 18535739 (Pereira, et al., 2008), 19588081 (Oba-Shinjo, et al., 2009), 22237443 (Messinger, et al., 2012), 23601496 (Elder, et al., 2013), 24273659 (Alansari, et al., 2013), 29122469 (Mori, et al., 2017), 29390460 (Martinez, et al., 2017), 29637184 (Elenga, et al., 2018), 31342611 (Reuser, et al., 2019)
c.525del	8558570 (Kroos, et al., 1995), 18429042 (Pittis, et al., 2008), 15121988 (Van den Hout, et al., 2004), 31342611 (Reuser, et al., 2019), 30902109 (Poelman, et al., 2019)

REFERENCES

Adams EM, Becker JA, Griffith L, Segal A, Plotz PH, Raben N. 1997. Glycogenosis type II: a juvenile-specific mutation with an unusual splicing pattern and a shared mutation in African Americans. Hum Mutat 10(2):128-34.

Al-Hassnan ZN, Khalifa OA, Bubshait DK, Tulbah S, Alkorashy M, Alzaidan H, Alowain M, Rahbeeni Z, Al-Sayed M. 2018. The phenotype, genotype, and outcome of infantile-onset Pompe disease in 18 Saudi patients. Mol Genet Metab Rep 15:50-54.

Alandy-Dy J, Wencel M, Hall K, Simon J, Chen Y, Valenti E, Yang J, Bali D, Lakatos A, Goyal N and others. 2019. Variable clinical features and genotype-phenotype correlations in 18 patients with late-onset Pompe disease. Ann Transl Med 7(13):276.

Alansari A, Al-Rawahi S, Ba-Omar T, Al-Nabhani M, Date A. 2013. The Identification of Pompe Disease Mutations in Archival Tissues and Development of a Rapid Molecular-based Test. Sultan Qaboos Univ Med J 13(4):502-9.

Alejaldre A, Diaz-Manera J, Ravaglia S, Tibaldi EC, D'Amore F, Moris G, Muelas N, Vilchez JJ, Garcia-Medina A, Uson M and others. 2012. Trunk muscle involvement in late-onset Pompe disease: study of thirty patients. Neuromuscul Disord 22 Suppl 2:S148-54.

Almeida V, Conceicao I, Fineza I, Coelho T, Silveira F, Santos M, Valverde A, Geraldo A, Mare R, Aguiar TC and others. 2017. Screening for Pompe disease in a Portuguese high risk population. Neuromuscul Disord 27(8):777-781.

Amarinthnukrowh P, Tongkobpetch S, Kongpatanayothin A, Suphapeetiporn K, Shotelersuk V. 2010. p.D645E of acid alpha-glucosidase is the most common mutation in thai patients with infantile-onset pompe disease. Genet Test Mol Biomarkers 14(6):835-7.

Andreassen CS, Schlutter JM, Vissing J, Andersen H. 2014. Effect of enzyme replacement therapy on isokinetic strength for all major muscle groups in four patients with Pompe disease-a long-term follow-up. Mol Genet Metab 112(1):40-3.

Angelini C, Savarese M, Fanin M, Nigro V. 2016. Next generation sequencing detection of late onset pompe disease. Muscle Nerve 53(6):981-3.

Bandyopadhyay S, Wicklund M, Specht CS. 2015. Novel presentation of Pompe disease: Inclusion-body myositis-like clinical phenotype. Muscle Nerve 52(3):466-7.

Banugaria SG, Prater SN, Patel TT, Dearmey SM, Milleson C, Sheets KB, Bali DS, Rehder CW, Raiman JA, Wang RA and others. 2013. Algorithm for the early diagnosis and treatment of patients with cross reactive immunologic material-negative classic infantile pompe disease: a step towards improving the efficacy of ERT. PLoS One 8(6):e67052.

Becker JA, Vlach J, Raben N, Nagaraju K, Adams EM, Hermans MM, Reuser AJ, Brooks SS, Tifft CJ, Hirschhorn R and others. 1998. The African origin of the common mutation in African American patients with glycogen-storage disease type II. Am J Hum Genet 62(4):991-4.

Bekircan-Kurt CE, Gunes HN, Yildiz FG, Saka E, Tan E, Erdem-Ozdamar S. 2017. New mutations and genotype-phenotype correlation in late-onset Pompe patients. Acta Neurol Belg 117(1):269-275.

Bergsma AJ, In 't Groen SLM, van den Dorpel JJA, van den Hout H, van der Beek N, Schoser B, Toscano A, Musumeci O, Bembi B, Dardis A and others. 2019. A genetic modifier of symptom onset in Pompe disease. EBioMedicine 43:553-561.

Boerkoel CF, Exelbert R, Nicastri C, Nichols RC, Miller FW, Plotz PH, Raben N. 1995. Leaky splicing mutation in the acid maltase gene is associated with delayed onset of glycogenosis type II. Am J Hum Genet 56(4):887-97.

Burlina AB, Polo G, Salviati L, Duro G, Zizzo C, Dardis A, Bembi B, Cazzorla C, Rubert L, Zordan R and others. 2018. Newborn screening for lysosomal storage disorders by tandem mass spectrometry in North East Italy. J Inherit Metab Dis 41(2):209-219.

Byrne BJ, Geberhiwot T, Barshop BA, Barohn R, Hughes D, Bratkovic D, Desnuelle C, Laforet P, Mengel E, Roberts M and others. 2017. A study on the safety and efficacy of reveglucosidase alfa in patients with late-onset Pompe disease. Orphanet J Rare Dis 12(1):144.

Castro-Gago M, Eiris-Punal J, Rodriguez-Nunez A, Pintos-Martinez E, Benlloch-Marin T, Barros-Angueira F. 1999. [Severe form of juvenile type II glycogenosis in a compound-heterozygous boy (Tyr-292--> Cys/Arg-854-->Stop)]

Forma grave de glucogenosis tipo II juvenil en un nino heterocigoto compuesto (Tyr-292-->Cys/Arg-854-->Stop). Rev Neurol 29(1):46-9.

Chen X, Liu T, Huang M, Wu J, Zhu J, Guo Y, Xu X, Li F, Wang J, Fu L. 2017. Clinical and Molecular Characterization of Infantile-Onset Pompe Disease in Mainland Chinese Patients: Identification of Two Common Mutations. Genet Test Mol Biomarkers 21(6):391-396.

Chien YH, Lee NC, Chen CA, Tsai FJ, Tsai WH, Shieh JY, Huang HJ, Hsu WC, Tsai TH, Hwu WL. 2015. Long-term prognosis of patients with infantile-onset Pompe disease diagnosed by newborn screening and treated since birth. J Pediatr 166(4):985-91 e1-2.

Chien YH, Lee NC, Huang HJ, Thurberg BL, Tsai FJ, Hwu WL. 2011. Later-onset Pompe disease: early detection and early treatment initiation enabled by newborn screening. J Pediatr 158(6):1023-1027 e1.

Chien YH, Lee NC, Thurberg BL, Chiang SC, Zhang XK, Keutzer J, Huang AC, Wu MH, Huang PH, Tsai FJ and others. 2009. Pompe disease in infants: improving the prognosis by newborn screening and early treatment. Pediatrics 124(6):e1116-25.

Chien YH, Lee NC, Tsai YJ, Thurberg BL, Tsai FJ, Hwu WL. 2014. Prominent vacuolation of the eyelid levator muscle in an early-treated child with infantile-onset Pompe disease. Muscle Nerve 50(2):301-2.

Cho A, Kim SJ, Lim BC, Hwang H, Park JD, Kim GB, Jin DK, Lee J, Ki CS, Kim KJ and others. 2012. Infantile Pompe disease: clinical and genetic characteristics with an experience of enzyme replacement therapy. J Child Neurol 27(3):319-24.

Chu YP, Sheng B, Lau KK, Chan HF, Kam GY, Lee HH, Mak CM. 2016. Clinical manifestation of late onset Pompe disease patients in Hong Kong. Neuromuscul Disord 26(12):873-879.

Del Rizzo M, Fanin M, Cerutti A, Cazzorla C, Milanesi O, Nascimbeni AC, Angelini C, Giordano L, Bordugo A, Burlina AB. 2010. Long-term follow-up results in enzyme replacement therapy for Pompe disease: a case report. J Inherit Metab Dis 33 Suppl 3:S389-93.

Dou W, Peng C, Zheng J, Gu X, Fu L, Martiniuk F, Sheng HZ. 2006. A novel missense mutation in the acid alpha-glucosidase gene causing the classic infantile form of Pompe disease. Clin Chim Acta 374(1-2):145-6.

Ebert SE, Brenzy K, Cartwright MS. 2019. Neuromuscular ultrasound as an initial evaluation for suspected myopathy: A case report. Muscle Nerve 59(5):E31-E32.

Echaniz-Laguna A, Carlier RY, Laloui K, Carlier P, Salort-Campana E, Pouget J, Laforet P. 2015. Should patients with asymptomatic pompe disease be treated? A nationwide study in France. Muscle Nerve 51(6):884-9.

Elder ME, Nayak S, Collins SW, Lawson LA, Kelley JS, Herzog RW, Modica RF, Lew J, Lawrence RM, Byrne BJ. 2013. B-Cell depletion and immunomodulation before initiation of enzyme replacement therapy blocks the immune response to acid alpha-glucosidase in infantile-onset Pompe disease. J Pediatr 163(3):847-54 e1.

Elenga N, Verloes A, Mrsic Y, Basurko C, Schaub R, Cuadro-Alvarez E, Kom-Tchameni R, Carles G, Lambert V, Boukhari R and others. 2018. Incidence of infantile Pompe disease in the Maroon population of French Guiana. BMJ Paediatr Open 2(1):e000182.

Fu L, Qiu W, Yu Y, Guo Y, Zhao P, Zhang X, Liu C, Li F, Huang H, Huang M and others. 2014. Clinical and molecular genetic study of infantile-onset Pompe disease in Chinese patients: identification of 6 novel mutations. Gene 535(1):53-9.

Fukuhara Y, Fuji N, Yamazaki N, Hirakiyama A, Kamioka T, Seo JH, Mashima R, Kosuga M, Okuyama T. 2018. A molecular analysis of the GAA gene and clinical spectrum in 38 patients with Pompe disease in Japan. Mol Genet Metab Rep 14:3-9.

Golden-Grant K, Merritt JL, 2nd, Scott CR. 2015. Ethical considerations of population screening for late-onset genetic disease. Clin Genet 88(6):589-92.

Golsari A, Nasimzadah A, Thomalla G, Keller S, Gerloff C, Magnus T. 2018. Prevalence of adult Pompe disease in patients with proximal myopathic syndrome and undiagnosed muscle biopsy. Neuromuscul Disord 28(3):257-261.

Gort L, Coll MJ, Chabas A. 2007. Glycogen storage disease type II in Spanish patients: high frequency of c.1076-1G>C mutation. Mol Genet Metab 92(1-2):183-7.

Grzesiuk AK, Shinjo SM, da Silva R, Machado M, Galera MF, Marie SK. 2010. Homozygotic intronic GAA mutation in three siblings with late-onset Pompe's disease. Arq Neuropsiquiatr 68(2):194-7.

Hermans MM, de Graaff E, Kroos MA, Wisselaar HA, Willemsen R, Oostra BA, Reuser AJ. 1993a. The conservative substitution Asp-645-->Glu in lysosomal alpha-glucosidase affects transport and

phosphorylation of the enzyme in an adult patient with glycogen-storage disease type II. Biochem J 289 (Pt 3):687-93.

Hermans MM, Kroos MA, de Graaff E, Oostra BA, Reuser AJ. 1993b. Two mutations affecting the transport and maturation of lysosomal alpha-glucosidase in an adult case of glycogen storage disease type II. Hum Mutat 2(4):268-73.

Hermans MM, van Leenen D, Kroos MA, Beesley CE, Van Der Ploeg AT, Sakuraba H, Wevers R, Kleijer W, Michelakakis H, Kirk EP and others. 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(1):47-56.

Herzog A, Hartung R, Reuser AJ, Hermanns P, Runz H, Karabul N, Gokce S, Pohlenz J, Kampmann C, Lampe C and others. 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.

Hobson-Webb LD, Austin SL, Jain S, Case LE, Greene K, Kishnani PS. 2015. Small-fiber neuropathy in pompe disease: first reported cases and prospective screening of a clinic cohort. Am J Case Rep 16:196-201.

Horvath JJ, Austin SL, Case LE, Greene KB, Jones HN, Soher BJ, Kishnani PS, Bashir MR. 2015. Correlation between quantitative whole-body muscle magnetic resonance imaging and clinical muscle weakness in Pompe disease. Muscle Nerve 51(5):722-30.

Hossain MA, Miyajima T, Akiyama K, Eto Y. 2018. A Case of Adult-onset Pompe Disease with Cerebral Stroke and Left Ventricular Hypertrophy. J Stroke Cerebrovasc Dis 27(11):3046-3052.

Huie ML, Chen AS, Brooks SS, Grix A, 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(7):1081-7.

Huie ML, Kasper JS, Arn PH, Greenberg CR, Hirschhorn R. 1999. Increased occurrence of cleft lip in glycogen storage disease type II (GSDII): exclusion of a contiguous gene syndrome in two patients by presence of intragenic mutations including a novel nonsense mutation Gln58Stop. Am J Med Genet 85(1):5-8.

Huie ML, Tsujino S, Sklower Brooks S, Engel A, Elias E, Bonthron DT, Bessley C, Shanske S, DiMauro S, Goto YI and others. 1998. Glycogen storage disease type II: identification of four novel missense mutations (D645N, G648S, R672W, R672Q) and two insertions/deletions in the acid alpha-glucosidase locus of patients of differing phenotype. Biochem Biophys Res Commun 244(3):921-7.

Ishigaki K, Murakami T, Nakanishi T, Oda E, Sato T, Osawa M. 2012. Close monitoring of initial enzyme replacement therapy in a patient with childhood-onset Pompe disease. Brain Dev 34(2):98-102.

Jastrzebska A, Potulska-Chromik A, Lusakowska A, Jastrzebski M, Lipowska M, Kierdaszuk B, Kaminska A, Kostera-Pruszczyk A. 2019. Screening for late-onset Pompe disease in Poland. Acta Neurol Scand 140(4):239-243.

Johnson K, Topf A, Bertoli M, Phillips L, Claeys KG, Stojanovic VR, Peric S, Hahn A, Maddison P, Akay E and others. 2017. Identification of GAA variants through whole exome sequencing targeted to a cohort of 606 patients with unexplained limb-girdle muscle weakness. Orphanet J Rare Dis 12(1):173.

Joshi PR, Glaser D, Schmidt S, Vorgerd M, Winterholler M, Eger K, Zierz S, Deschauer M. 2008. Molecular diagnosis of German patients with late-onset glycogen storage disease type II. J Inherit Metab Dis 31 Suppl 2:S261-5.

Kim MS, Song A, Im M, Huh J, Kang IS, Song J, Yang A, Kim J, Kwon EK, Choi EJ and others. 2019. Clinical and molecular characterization of Korean children with infantile and late-onset Pompe disease: 10 years of experience with enzyme replacement therapy at a single center. Korean J Pediatr 62(6):224-234.

Kindel SJ, Miller EM, Gupta R, Cripe LH, Hinton RB, Spicer RL, Towbin JA, Ware SM. 2012. Pediatric cardiomyopathy: importance of genetic and metabolic evaluation. J Card Fail 18(5):396-403.

Kishnani PS, Nicolino M, Voit T, Rogers RC, Tsai AC, Waterson J, Herman GE, Amalfitano A, Thurberg BL, Richards S and others. 2006. Chinese hamster ovary cell-derived recombinant human acid alpha-glucosidase in infantile-onset Pompe disease. J Pediatr 149(1):89-97.

Ko TM, Hwu WL, Lin YW, Tseng LH, Hwa HL, Wang TR, Chuang SM. 1999. Molecular genetic study of Pompe disease in Chinese patients in Taiwan. Hum Mutat 13(5):380-4.

Kobayashi H, Shimada Y, Ikegami M, Kawai T, Sakurai K, Urashima T, Ijima M, Fujiwara M, Kaneshiro E, Ohashi T and others. 2010. Prognostic factors for the late onset Pompe disease with enzyme replacement therapy: from our experience of 4 cases including an autopsy case. Mol Genet Metab 100(1):14-9.

Kostera-Pruszczyk A, Opuchlik A, Lugowska A, Nadaj A, Bojakowski J, Tylki-Szymanska A, Kaminska A. 2006. Juvenile onset acid maltase deficiency presenting as a rigid spine syndrome. Neuromuscul Disord 16(4):282-5.

Kroos M, Manta P, Mavridou I, Muntoni F, Halley D, Van der Helm R, Zaifeiriou D, Van der Ploeg A, Reuser A, Michelakakis H. 2006. Seven cases of Pompe disease from Greece. J Inherit Metab Dis 29(4):556-63.

Kroos MA, Van der Kraan M, Van Diggelen OP, Kleijer WJ, Reuser AJ, Van den Boogaard MJ, Ausems MG, Ploos van Amstel HK, Poenaru L, Nicolino M and others. 1995. Glycogen storage disease type II: frequency of three common mutant alleles and their associated clinical phenotypes studied in 121 patients. J Med Genet 32(10):836-7.

Kroos MA, van Leenen D, Verbiest J, Reuser AJ, Hermans MM. 1998. Glycogen storage disease type II: identification of a dinucleotide deletion and a common missense mutation in the lysosomal alpha-glucosidase gene. Clin Genet 53(5):379-82.

Laforet P, Nicolino M, Eymard PB, Puech JP, Caillaud C, Poenaru L, Fardeau M. 2000. Juvenile and adult-onset acid maltase deficiency in France: genotype-phenotype correlation. Neurology 55(8):1122-8.

Lam CW, Yuen YP, Chan KY, Tong SF, Lai CK, Chow TC, Lee KC, Chan YW, Martiniuk F. 2003. Juvenile-onset glycogen storage disease type II with novel mutations in acid alpha-glucosidase gene. Neurology 60(4):715-7.

Lee JH, Shin JH, Park HJ, Kim SZ, Jeon YM, Kim HK, Kim DS, Choi YC. 2017. Targeted population screening of late onset Pompe disease in unspecified myopathy patients for Korean population. Neuromuscul Disord 27(6):550-556.

Levesque S, Auray-Blais C, Gravel E, Boutin M, Dempsey-Nunez L, Jacques PE, Chenier S, Larue S, Rioux MF, Al-Hertani W and others. 2016. Diagnosis of late-onset Pompe disease and other muscle disorders by next-generation sequencing. Orphanet J Rare Dis 11:8.

Liu HX, Pu CQ, Shi Q, Zhang YT, Ban R. 2018. Identification of Seven Novel Mutations in the Acid Alpha-glucosidase Gene in Five Chinese Patients with Late-onset Pompe Disease. Chin Med J (Engl) 131(4):448-453.

Liu Q, Zhao J, Wang ZX, Zhang W, Yuan Y. 2013. [Clinical features and acid alpha-glucosidase gene mutation in 7 Chinese patients with glycogen storage disease type II]. Zhonghua Yi Xue Za Zhi 93(25):1981-5.

Liu X, Wang Z, Jin W, Lv H, Zhang W, Que C, Huang Y, Yuan Y. 2014. Clinical and GAA gene mutation analysis in mainland Chinese patients with late-onset Pompe disease: identifying c.2238G > C as the most common mutation. BMC Med Genet 15:141.

Lorenzoni PJ, Kay CSK, Higashi NS, D'Almeida V, Werneck LC, Scola RH. 2018. Late-onset Pompe disease: what is the prevalence of limb-girdle muscular weakness presentation? Arq Neuropsiquiatr 76(4):247-251.

Loscher WN, Huemer M, Stulnig TM, Simschitz P, Iglseder S, Eggers C, Moser H, Moslinger D, Freilinger M, Lagler F and others. 2018. Pompe disease in Austria: clinical, genetic and epidemiological aspects. J Neurol 265(1):159-164.

Loureiro Neves F, Garcia PC, Madureira N, Araujo H, Rodrigues F, Estevao MH, Lacerda L, Diogo Matos LM. 2013. [Juvenile Pompe disease: retrospective clinical study]

Doenca de Pompe juvenil: estudo retrospetivo de casuistica clinica. Acta Med Port 26(4):361-70.

Lyu JW, Xu XB, Ji KQ, Zhang N, Sun Y, Zhao DD, Zhao YY, Yan CZ. 2019. Activated mTOR signaling pathway in myofibers with inherited metabolic defect might be an evidence for mTOR inhibition therapies. Chin Med J (Engl) 132(7):805-810.

Maimaiti M, Takahashi S, Okajima K, Suzuki N, Ohinata J, Araki A, Tanaka H, Mukai T, 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(8):493-6.

Manwaring V, Prunty H, Bainbridge K, Burke D, Finnegan N, Franses R, Lam A, Vellodi A, Heales S. 2012. Urine analysis of glucose tetrasaccharide by HPLC; a useful marker for the investigation of patients with Pompe and other glycogen storage diseases. J Inherit Metab Dis 35(2):311-6.

Martinez M, Romero MG, Guereta LG, Cabrera M, Regojo RM, Albajara L, Couce ML, Pipaon MS. 2017. Infantile-onset Pompe disease with neonatal debut: A case report and literature review. Medicine (Baltimore) 96(51):e9186.

McCready ME, Carson NL, Chakraborty P, Clarke JT, Callahan JW, Skomorowski MA, Chan AK, Bamforth F, Casey R, Rupar CA and others. 2007. Development of a clinical assay for detection of GAA mutations and characterization of the GAA mutation spectrum in a Canadian cohort of individuals with glycogen storage disease, type II. Mol Genet Metab 92(4):325-35.

McIntosh P, Austin S, Sullivan J, Bailey L, Bailey C, Viskochil D, Kishnani PS. 2017. Three cases of multi-generational Pompe disease: Are current practices missing diagnostic and treatment opportunities? Am J Med Genet A 173(10):2628-2634.

Messinger YH, Mendelsohn NJ, Rhead W, Dimmock D, Hershkovitz E, Champion M, Jones SA, Olson R, White A, Wells C and others. 2012. Successful immune tolerance induction to enzyme replacement therapy in CRIM-negative infantile Pompe disease. Genet Med 14(1):135-42.

Montagnese F, Barca E, Musumeci O, Mondello S, Migliorato A, Ciranni A, Rodolico C, De Filippi P, Danesino C, Toscano A. 2015. Clinical and molecular aspects of 30 patients with late-onset Pompe disease (LOPD): unusual features and response to treatment. J Neurol 262(4):968-78.

Montalvo AL, Bembi B, Donnarumma M, Filocamo M, Parenti G, Rossi M, Merlini L, Buratti E, De Filippi P, Dardis A and others. 2006. Mutation profile of the GAA gene in 40 Italian patients with late onset glycogen storage disease type II. Hum Mutat 27(10):999-1006.

Montalvo AL, Cariati R, Deganuto M, Guerci V, Garcia R, Ciana G, Bembi B, Pittis MG. 2004. Glycogenosis type II: identification and expression of three novel mutations in the acid alpha-glucosidase gene causing the infantile form of the disease. Mol Genet Metab 81(3):203-8.

Mori M, Haskell G, Kazi Z, Zhu X, DeArmey SM, Goldstein JL, Bali D, Rehder C, Cirulli ET, Kishnani PS. 2017. Sensitivity of whole exome sequencing in detecting infantile- and late-onset Pompe disease. Mol Genet Metab 122(4):189-197.

Muller-Felber W, Horvath R, Gempel K, Podskarbi T, Shin Y, Pongratz D, Walter MC, Baethmann M, Schlotter-Weigel B, Lochmuller H and others. 2007. Late onset Pompe disease: clinical and neurophysiological spectrum of 38 patients including long-term follow-up in 18 patients. Neuromuscul Disord 17(9-10):698-706.

Musumeci O, la Marca G, Spada M, Mondello S, Danesino C, Comi GP, Pegoraro E, Antonini G, Marrosu G, Liguori R and others. 2016. LOPED study: looking for an early diagnosis in a late-onset Pompe disease high-risk population. J Neurol Neurosurg Psychiatry 87(1):5-11.

Musumeci O, Thieme A, Claeys KG, Wenninger S, Kley RA, Kuhn M, Lukacs Z, Deschauer M, Gaeta M, Toscano A and others. 2015. Homozygosity for the common GAA gene splice site mutation c.-32-13T>G in Pompe disease is associated with the classical adult phenotypical spectrum. Neuromuscul Disord 25(9):719-24.

Nabatame S, Taniike M, Sakai N, Kato-Nishimura K, Mohri I, Kagitani-Shimono K, Okinaga T, Tachibana N, Ozono K. 2009. Sleep disordered breathing in childhood-onset acid maltase deficiency. Brain Dev 31(3):234-9.

Nallamilli BRR, Chakravorty S, Kesari A, Tanner A, Ankala A, Schneider T, da Silva C, Beadling R, Alexander JJ, Askree SH and others. 2018. Genetic landscape and novel disease mechanisms from a large LGMD cohort of 4656 patients. Ann Clin Transl Neurol 5(12):1574-1587.

Nascimbeni AC, Fanin M, Tasca E, Angelini C. 2008. Molecular pathology and enzyme processing in various phenotypes of acid maltase deficiency. Neurology 70(8):617-26.

Nazari F, Sinaei F, Nilipour Y, Fatehi F, Streubel B, Ashrafi MR, Aryani O, Nafissi S. 2017. Late-onset pompe disease in Iran: A clinical and genetic report. Muscle Nerve 55(6):835-840.

Ngiwsara L, Wattanasirichaigoon D, Tim-Aroon T, Rojnueangnit K, Noojaroen S, Khongkraparn A, Sawangareetrakul P, Ketudat-Cairns JR, Charoenwattanasatien R, Champattanachai V and others. 2019. Clinical course, mutations and its functional characteristics of infantile-onset Pompe disease in Thailand. BMC Med Genet 20(1):156.

Nino MY, Mateus HE, Fonseca DJ, Kroos MA, Ospina SY, Mejia JF, Uribe JA, Reuser AJ, Laissue P. 2013. Identification and Functional Characterization of GAA Mutations in Colombian Patients Affected by Pompe Disease. JIMD Rep 7:39-48.

Oba-Shinjo SM, da Silva R, Andrade FG, Palmer RE, Pomponio RJ, Ciociola KM, M SC, Gutierrez PS, Porta G, Marrone CD and others. 2009. Pompe disease in a Brazilian series: clinical and molecular analyses with identification of nine new mutations. J Neurol 256(11):1881-90.

Orlikowski D, Pellegrini N, Prigent H, Laforet P, Carlier R, Carlier P, Eymard B, Lofaso F, Annane D. 2011. Recombinant human acid alpha-glucosidase (rhGAA) in adult patients with severe respiratory failure due to Pompe disease. Neuromuscul Disord 21(7):477-82.

Palermo AT, Palmer RE, So KS, Oba-Shinjo SM, Zhang M, Richards B, Madhiwalla ST, Finn PF, Hasegawa A, Ciociola KM and others. 2012. Transcriptional response to GAA deficiency (Pompe disease) in infantile-onset patients. Mol Genet Metab 106(3):287-300.

Palmer RE, Amartino HM, Niizawa G, Blanco M, Pomponio RJ, Chamoles NA. 2007. Pompe disease (glycogen storage disease type II) in Argentineans: clinical manifestations and identification of 9 novel mutations. Neuromuscul Disord 17(1):16-22.

Parini R, De Lorenzo P, Dardis A, Burlina A, Cassio A, Cavarzere P, Concolino D, Della Casa R, Deodato F, Donati MA and others. 2018. Long term clinical history of an Italian cohort of infantile onset Pompe disease treated with enzyme replacement therapy. Orphanet J Rare Dis 13(1):32.

Park HD, Lee DH, Choi TY, Lee YK, Lee SY, Kim JW, Ki CS, Lee YW. 2013. Three patients with glycogen storage disease type II and the mutational spectrum of GAA in Korean patients. Ann Clin Lab Sci 43(3):311-6.

Park JS, Kim HG, Shin JH, Choi YC, Kim DS. 2015. Effect of enzyme replacement therapy in late onset Pompe disease: open pilot study of 48 weeks follow-up. Neurol Sci 36(4):599-605.

Pereira SJ, Berditchevisky CR, Marie SK. 2008. Report of the first Brazilian infantile Pompe disease patient to be treated with recombinant human acid alpha-glucosidase. J Pediatr (Rio J) 84(3):272-5.

Pichiecchio A, Sacco S, De Filippi P, Caverzasi E, Ravaglia S, Bastianello S, Danesino C. 2017. Late-onset Pompe disease: a genetic-radiological correlation on cerebral vascular anomalies. J Neurol 264(10):2110-2118.

Pipo JR, Feng JH, Yamamoto T, Ohsaki Y, Nanba E, Tsujino S, Sakuragawa N, Martiniuk F, Ninomiya H, Oka A and others. 2003. New GAA mutations in Japanese patients with GSDII (Pompe disease). Pediatr Neurol 29(4):284-7.

Pittis MG, Donnarumma M, Montalvo AL, Dominissini S, Kroos M, Rosano C, Stroppiano M, Bianco MG, Donati MA, Parenti G and others. 2008. Molecular and functional characterization of eight novel GAA mutations in Italian infants with Pompe disease. Hum Mutat 29(6):E27-36.

Poelman E, Hoogeveen-Westerveld M, van den Hout JMP, Bredius RGM, Lankester AC, Driessen GJA, Kamphuis SSM, Pijnappel WWM, van der Ploeg AT. 2019. Effects of immunomodulation in classic infantile Pompe patients with high antibody titers. Orphanet J Rare Dis 14(1):71.

Quan JJ, Liu LJ, Lyu TW, Huang XP, Tian J. 2017. Atypical Infantile-onset Pompe Disease with Hypertrophic Cardiomyopathy. Chin Med J (Engl) 130(19):2393-2394.

Raben N, Nichols RC, Boerkoel C, Plotz P. 1995. Genetic defects in patients with glycogenosis type II (acid maltase deficiency). Muscle Nerve Suppl 3:S70-4.

Rairikar MV, Case LE, Bailey LA, Kazi ZB, Desai AK, Berrier KL, Coats J, Gandy R, Quinones R, Kishnani PS. 2017. Insight into the phenotype of infants with Pompe disease identified by newborn screening with the common c.-32-13T>G "late-onset" GAA variant. Mol Genet Metab 122(3):99-107.

Reuser AJJ, van der Ploeg AT, Chien YH, Llerena J, Jr., Abbott MA, Clemens PR, Kimonis VE, Leslie N, Maruti SS, Sanson BJ and others. 2019. GAA variants and phenotypes among 1,079 patients with Pompe disease: Data from the Pompe Registry. Hum Mutat 40(11):2146-2164.

Sacconi S, Bocquet JD, Chanalet S, Tanant V, Salviati L, Desnuelle C. 2010. Abnormalities of cerebral arteries are frequent in patients with late-onset Pompe disease. J Neurol 257(10):1730-3.

Sampaolo S, Esposito T, Farina O, Formicola D, Diodato D, Gianfrancesco F, Cipullo F, Cremone G, Cirillo M, Del Viscovo L and others. 2013. Distinct disease phenotypes linked to different combinations of GAA mutations in a large late-onset GSDII sibship. Orphanet J Rare Dis 8:159.

Savarese M, Torella A, Musumeci O, Angelini C, Astrea G, Bello L, Bruno C, Comi GP, Di Fruscio G, Piluso G and others. 2018. Targeted gene panel screening is an effective tool to identify undiagnosed late onset Pompe disease. Neuromuscul Disord 28(7):586-591.

Schanzer A, Kaiser AK, Muhlfeld C, Kulessa M, Paulus W, von Pein H, Rohrbach M, Viergutz L, Mengel E, Marquardt T and others. 2017. Quantification of muscle pathology in infantile Pompe disease. Neuromuscul Disord 27(2):141-152.

Schneider I, Hanisch F, Muller T, Schmidt B, Zierz S. 2013. Respiratory function in late-onset Pompe disease patients receiving long-term enzyme replacement therapy for more than 48 months. Wien Med Wochenschr 163(1-2):40-4.

Scott CR, Elliott S, Buroker N, Thomas LI, Keutzer J, Glass M, Gelb MH, Turecek F. 2013. Identification of infants at risk for developing Fabry, Pompe, or mucopolysaccharidosis-I from newborn blood spots by tandem mass spectrometry. J Pediatr 163(2):498-503.

Semplicini C, Letard P, De Antonio M, Taouagh N, Perniconi B, Bouhour F, Echaniz-Laguna A, Orlikowski D, Sacconi S, Salort-Campana E and others. 2018. Late-onset Pompe disease in France: molecular features and epidemiology from a nationwide study. J Inherit Metab Dis 41(6):937-946.

Sharma MC, Schultze C, von Moers A, Stoltenburg-Didinger G, Shin YS, Podskarbi T, Isenhardt K, Tews DS, Goebel HH. 2005. Delayed or late-onset type II glycogenosis with globular inclusions. Acta Neuropathol 110(2):151-7.

Shieh JJ, Lin CY. 1996. Identification of a small deletion in one allele of patients with infantile form of glycogen storage disease type II. Biochem Biophys Res Commun 219(2):322-6.

Shieh JJ, Lin CY. 1998. Frequent mutation in Chinese patients with infantile type of GSD II in Taiwan: evidence for a founder effect. Hum Mutat 11(4):306-12.

Shin YS. 2006. Glycogen storage disease: clinical, biochemical, and molecular heterogeneity. Semin Pediatr Neurol 13(2):115-20.

Sifi Y, Medjroubi M, Froissart R, Taghane N, Sifi K, Benhabiles A, Lemai S, Semra S, Benmekhebi H, Bouderda Z and others. 2017. Clinical Analysis of Algerian Patients with Pompe Disease. J Neurodegener Dis 2017:9427269.

Sixel BS, Silva LD, Cavalcanti NC, Penque GM, Lisboa S, Horovitz DD, Llerena JC, Jr. 2017. Respiratory manifestations in late-onset Pompe disease: a case series conducted in Brazil. J Bras Pneumol 43(1):54-59.

Spada M, Porta F, Vercelli L, Pagliardini V, Chiado-Piat L, Boffi P, Pagliardini S, Remiche G, Ronchi D, Comi G and others. 2013. Screening for later-onset Pompe's disease in patients with paucisymptomatic hyperCKemia. Mol Genet Metab 109(2):171-3.

Stepien KM, Hendriksz CJ, Roberts M, Sharma R. 2016. Observational clinical study of 22 adult-onset Pompe disease patients undergoing enzyme replacement therapy over 5years. Mol Genet Metab 117(4):413-8.

Swarr DT, Kaufman B, Fogel MA, Finkel R, Ganesh J. 2012. Unusual cardiac "masses" in a newborn with infantile pompe disease. JIMD Rep 5:17-20.

Szklanny K, Tylki-Szymanska A. 2018. Follow-up analysis of voice quality in patients with late-onset Pompe disease. Orphanet J Rare Dis 13(1):189.

Taisne N, Desnuelle C, Juntas Morales R, Ferrer Monasterio X, Sacconi S, Duval F, Sole G, Flipo RM, Lacour A, Vermersch P and others. 2017. Bent spine syndrome as the initial symptom of late-onset Pompe disease. Muscle Nerve 56(1):167-170.

Tecellioglu M, Kamisli O. 2015. Familial Pompe Disease. Med Arch 69(5):342-4.

Torrealba-Acosta G, Rodriguez-Roblero MC, Bogantes-Ledezma S, Carazo-Cespedes K, Desnuelle C. 2017. First clinical and genetic description of a family diagnosed with late-onset Pompe disease from Costa Rica. Neuromuscul Disord 27(10):951-955.

Trend PS, Wiles CM, Spencer GT, Morgan-Hughes JA, Lake BD, Patrick AD. 1985. Acid maltase deficiency in adults. Diagnosis and management in five cases. Brain 108 (Pt 4):845-60.

Tsai AC, Hung YW, Harding C, Koeller DM, Wang J, Wong LC. 2017. Next generation deep sequencing corrects diagnostic pitfalls of traditional molecular approach in a patient with prenatal onset of Pompe disease. Am J Med Genet A 173(9):2500-2504.

Tsujino S, Huie M, Kanazawa N, Sugie H, Goto Y, Kawai M, Nonaka I, Hirschhorn R, Sakuragawa N. 2000. Frequent mutations in Japanese patients with acid maltase deficiency. Neuromuscul Disord 10(8):599-603.

Turaca LT, de Faria DO, Kyosen SO, Teixeira VD, Motta FL, Pessoa JG, Rodrigues ESM, de Almeida SS, D'Almeida V, Munoz Rojas MV and others. 2015. Novel GAA mutations in patients with Pompe disease. Gene 561(1):124-31.

Van den Hout JM, Kamphoven JH, Winkel LP, Arts WF, De Klerk JB, Loonen MC, Vulto AG, Cromme-Dijkhuis A, Weisglas-Kuperus N, Hop W and others. 2004. Long-term intravenous treatment of Pompe disease with recombinant human alpha-glucosidase from milk. Pediatrics 113(5):e448-57.

van der Beek NA, Soliman OI, van Capelle CI, Geleijnse ML, Vletter WB, Kroos MA, Reuser AJ, Frohn-Mulder IM, van Doorn PA, van der Ploeg AT. 2008. Cardiac evaluation in children and adults with Pompe disease sharing the common c.-32-13T>G genotype rarely reveals abnormalities. J Neurol Sci 275(1-2):46-50.

Vill K, Schessl J, Teusch V, Schroeder S, Blaschek A, Schoser B, Muller-Felber W. 2015. Muscle ultrasound in classic infantile and adult Pompe disease: a useful screening tool in adults but not in infants. Neuromuscul Disord 25(2):120-6.

Wan L, Lee CC, Hsu CM, Hwu WL, Yang CC, Tsai CH, Tsai FJ. 2008. Identification of eight novel mutations of the acid alpha-glucosidase gene causing the infantile or juvenile form of glycogen storage disease type II. J Neurol 255(6):831-8.

Wasserstein MP, Caggana M, Bailey SM, Desnick RJ, Edelmann L, Estrella L, Holzman I, Kelly NR, Kornreich R, Kupchik SG and others. 2019. The New York pilot newborn screening program for lysosomal storage diseases: Report of the First 65,000 Infants. Genet Med 21(3):631-640.

Wens SC, Kroos MA, de Vries JM, Hoogeveen-Westerveld M, Wijgerde MG, van Doorn PA, van der Ploeg AT, Reuser AJ. 2012. Remarkably low fibroblast acid alpha-glucosidase activity in three adults with Pompe disease. Mol Genet Metab 107(3):485-9.

Wens SC, van Gelder CM, Kruijshaar ME, de Vries JM, van der Beek NA, Reuser AJ, van Doorn PA, van der Ploeg AT, Brusse E. 2013. Phenotypical variation within 22 families with Pompe disease. Orphanet J Rare Dis 8:182.

Wittmann J, Karg E, Turi S, Legnini E, Wittmann G, Giese AK, Lukas J, Golnitz U, Klingenhager M, Bodamer O and others. 2012. Newborn screening for lysosomal storage disorders in hungary. JIMD Rep 6:117-25.

Yang CF, Liu HC, Hsu TR, Tsai FC, Chiang SF, Chiang CC, Ho HC, Lai CJ, Yang TF, Chuang SY and others. 2014. A large-scale nationwide newborn screening program for Pompe disease in Taiwan: towards effective diagnosis and treatment. Am J Med Genet A 164A(1):54-61.

Yang CF, Yang CC, Liao HC, Huang LY, Chiang CC, Ho HC, Lai CJ, Chu TH, Yang TF, Hsu TR and others. 2016. Very Early Treatment for Infantile-Onset Pompe Disease Contributes to Better Outcomes. J Pediatr 169:174-80 e1.

CHAPTER 6

The haplotype paradox in Pompe disease: differential effects of haplotypes on the penetrance of European ancestry disease-associated variants

Douglas O.S. de Faria[1,2,3], Stijn L.M. in't Groen[1,2,3], Atze J. Bergsma[1,2,3], Monica Y. Nino[1,2,3], Ans T. van der Ploeg[2,3], Oscar Lao[4,5,6*], W.W.M. Pijnappel[1,2,3*]

1 Department of Pediatrics, Erasmus University Medical Center, Rotterdam, the Netherlands

2 Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, the Netherlands

3 Center for Lysosomal and Metabolic Diseases, Erasmus University Medical Center, Rotterdam, the Netherlands

4 CNAG-CRG, Centre for Genomic Regulation, C/ Baldiri i Reixach 4, 08028, Barcelona, Spain

5 Barcelona Institute of Science and Technology (BIST), Barcelona, Spain

6 Universitat Pompeu Fabra (UPF), Barcelona, Spain

Correspondence: WWMP Pijnappel, Department of Clinical Genetics, Erasmus University Medical Center, P.O.Box 2040, 3000 CA Rotterdam. The Netherlands E-mail: w.pijnappel@erasmusmc.nl

Oscar Lao, CNAG-CRG, Centre for Genomic Regulation, C/ Baldiri i Reixach 4, 08028, Barcelona, Spain Barcelona Institute of Science and Technology (BIST), Barcelona, Spain, Universitat Pompeu Fabra (UPF), Barcelona, Spain E-mail: oscar.lao@cnag.crg.eu

Manuscript in preparation

The haplotype paradox in Pompe disease: differential effects of haplotypes on the penetrance of European ancestry disease-associated variants

Douglas O. S. de Faria, Stijn L.M. in 't Groen, Atze J. Bergsma, Monica Y. Nino, Ans T. van der Ploeg, Oscar Lao, W.W.M. Pijnappel

Abstract

Ninety percent of European ancestry patients with Pompe disease with childhood and adult disease onset of symptoms carry the c.-32-13T>G *GAA* (IVS1) variant on one allele combined with another *GAA* disease-associated variant on the second allele. These patients have a broad clinical spectrum, suggesting that modifying factors could play a role in disease progression. It has been previously observed in other mendelian diseases that the haplotype context can act in shaping the penetrance of the disease. Here, we mapped haplotypes that are related to disease-associated variants and determined their possible impact on GAA expression. Using a minor allele frequency (MAF) ≥ 5%, we defined 26 haplotypes among Northern and Western European (CEU) individuals based on 83 single nucleotide variants. The haplotype proportions of Pompe disease-associated variants deviated from the expected observed frequencies in the general CEU population. Interestingly, the IVS1 variant was present in the most common observed haplotype in healthy individuals. In contrast, most other disease-associated variants concentrated in evolutionarily related less frequent haplotypes. By examining the presence of expression quantitative trait loci (eQTLs) in these haplotypes, we found that haplotypes of most disease-associated variants negatively affected GAA expression. In contrast, the haplotype containing the IVS1 variant does not modify the GAA expression. These results demonstrate that haplotypes can modulate the severity of disease-associated variants in Pompe disease. Because the common IVS1 variant has followed an atypical evolutionary behaviour, we speculate that it might be subject to positive selection as a result of a linked property that confers a selective evolutionary advantage.

INTRODUCTION

Pompe disease (glycogen storage disease type II; OMIM 232300) is an autosomal recessive disorder caused by disease-associated variants in the acid α-glucosidase (*GAA*) gene. These variants lead to a partial or complete lack of GAA enzymatic activity, resulting in the accumulation of lysosomal glycogen. This accumulation progressively affects many tissues, most prominently skeletal muscle. Depending on the disease-associated variants' severity, there are different forms, including classic infantile, childhood onset and adult onset forms of the disease [1]. The classic infantile form occurs when there are two severe disease-associated variants that completely abrogate GAA enzyme activity. For the childhood and adulthood onset forms the patients still have some residual enzyme activity but, insufficient to prevent disease onset [2].

Pompe disease is a rare disease with an estimated frequency of 1 per 14,000 to 1 per 250,000 newborns [3]–[7]. The Pompe variant database currently reports 911 *GAA* variants, of which 648 are disease-associated [8]. The most common variant worldwide is the c.-32-13T>G variant, commonly referred to as IVS1, located in intron 1. It affects around 45% of the worldwide population of patients with Pompe disease. It is present in 90% of European ancestry patients with childhood or adult phenotypes [9]. Nevertheless, the presence of some private *GAA* variants in certain populations points to single mutational events in those populations.

The IVS1 variant leads to aberrant pre-mRNA splicing in which *GAA* exon 2 is partially or completely skipped. This variant does allow for some (10-15%) residual canonical splicing, which explains the later onset of disease in patients carrying this variant [10]–[12]. The clinical heterogeneity of patients with late onset of Pompe disease that carries IVS1 combined with a null-variant on the second allele has previously been reported [13]–[15]. The onset of symptoms can vary from the age <1 to 60 years old, even for patients with identical *GAA* disease-associated variants. This suggests the presence of modifying factors. A recently identified modifying factor is the c.510C>T; p.(=) variant, which has been identified in patients carrying the IVS1 [15]. c.510C>T decreases GAA enzyme activity by reducing the extent of leaky wild-type splicing in the context of the IVS1 variant. The *ACE* and *ACTN3* genes have also been suggested as potential modifying factors for Pompe disease [16]. However, a different study did not find evidence of the modifier effect of *ACE* [17].

The *GAA* gene contains various single nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) higher than 1% [18], [19]. Although these common SNPs do not cause Pompe disease, they might modify the disease progression when present *in cis* with a disease-associated variant. The c.271G>A; p.(Asp91Asn) variant (also termed GAA2), which has a MAF of 3% in the European population, results in retention of the N-terminal fragment that is normally cleaved during processing of the protein. c.271G>A

was shown to be present *in cis* with c.307T>G; p.(Cys103Gly), a disease-associated variant that fully abrogated GAA enzyme activity [8], [20]. The c.1726G>A; p.(Gly576Ser) and c.2065G>A; p.(Glu689Lys) variants are common in the East Asian population, with MAFs of 15% and 24%, respectively. Homozygotes for these variants have an estimated 60-80% reduction of acid alpha-glucosidase compared to average normal, and they are known as Asian pseudodeficiency alleles [21].

A previous study investigating haplotypes has been performed in patients with the IVS1 variant and included a mixture of 18 disease-associated and common SNPs within the coding region of *GAA* [12]. Here, we analysed the haplotype background of childhood and adult Pompe disease-associated variants found in the Dutch population combining results from the SNP array and allele-specific PCR and compared the haplotype composition with the one of the general population. Additionally, we further studied the effect of the haplotypes in the expression of GAA by analysing expression quantitative trait loci (eQTLs).

METHODS

Identification of GAA haplotypes in CEU population from the 1000genomes

Haplotype information of the *GAA* (GRCh38 chr17:80,101,581-80,119,881) from 99 CEU (Utah Residents with Northern and Western European Ancestry) of the 1000 Genome project [22] were downloaded from hgdownload.soe.ucsc.edu/gbdb/hg38/1000Genomes/. A filter was applied to extract variants with MAF ≥ 5%. After applying the filter, 83 variants were considered (**Table S1**).

Pompe patients

The 130 Pompe patients enrolled in this study were part of a cohort of individuals followed up at the Erasmus MC University Medical Center. They were initially described in Bergsma *et al* 2019 [15]. All patients were self-declared Dutch ancestry and were heterozygous for the c.-32-13T>G (IVS1) variant, with a confirmed fully deleterious variant on the second allele. North European genetic ancestry of each patient was confirmed by means of a classical Multidimensional Scaling (MDS) analysis on an identical by state (IBS) matrix computed with Pompe patients and 1000G individuals using Plink software (see Supplementary **Figure S1**). Informed consent was obtained from all patients. The medical ethical committee of the Erasmus MC approved the study. A list of patients with genotypes can be found in Supplementary **Table S2**. More information on the severity of disease-associated variants can be found at www.pompevariantdatabase.nl .

SNP array analysis of the Dutch Pompe patients and haplotype assignment

DNA was isolated from leukocytes and/or fibroblasts. The Illumina GSA Beadchip (Illumina GSA arrays "Infinium iSelect 24x1 HTS Custom Beadchip kit"), corresponding to 654k variants, was applied to each patient. The Beadstudio GenCall algorithm implemented at GenomeStudio Software was used for single nucleotide variants (SNVs) calling and quality control (QC) procedures. SNVs with a call rate < 97,5% and/or Hardy-Weinberg equilibrium p-value < 1x10-4 were excluded from further analyses. Each individual sample was checked for discordance in relation to the sex, excess of heterozygosity and/or homozygosity, genetic ancestry and familial relationships using Plink software [23]. In order to implement a comprehensive haplotype analysis of the *GAA* region, covering all common variants present within the *GAA* gene, we have extracted all variants with MAF ≥ 5% in the CEU population. 5% was used as a cutoff due to their good quality in the SNP array analysis, more reliable linkage disequilibrium analysis, and the possibility to detect tag SNPs (i.e. an SNP that is representative for a haplotype) [24]-[26].

Data imputation and phasing

SNV imputation and phasing was done by two steps procedure. Phasing was done with Minimac [27] and imputation used the Markov Chain package [28], using sequencing data from the 1000G V.3 CEU. The inferred phase of the IVS1 allele by Minimac was validated by allele-specific PCR from exon 1 to 3 as described in Bergsma *et al* 2019 [15]. Since all patients were compound heterozygous for the IVS1 (validated by trio analysis), the phase of other disease variants was inferred.

Haplotype network analysis

The relationships of 198 CEU haplotypes in the *GAA* region were represented using a minimum spanning network as implemented in the pegas package [29] using the default parameters. In order to place the ancestral haplotype, the ancestral state of each SNP was obtained from Ensembl.

Pairwise Linkage Disequilibrium (LD) and clustering analysis to determine sub-haplotype blocks within GAA gene

A SNV pairwise linkage disequilibrium (LD) analysis was performed between the common SNVs in the *GAA* gene with a MAF equal to or higher than 5% in the CEU population. The LDlink 3.8 web-based tool LDmatrix (https://ldlink.nci.nih.gov/?tab=home) was used to estimate the LD as *r* squared (r^2). From the LD matrix between pairs of SNVs, we defined LD blocks and tag SNPs using the algorithm implemented in Haploreg v4.1 [30]. We could infer sub-haplotype blocks, regions with limited haplotype diversity, in which within a sub-haplotype block every variant could be determined as a tagging variant for the particular sub-haplotype block. We used 0.8 as a cut off for r^2.

Mapping eQTLs for GAA

The 83 variants selected within *GAA* with MAF ≥ 5% from CEU 1000G were analysed to identify the eQTLs for skeletal muscle tissue using GTEX (Genotype-Tissue Expression project – www.gtexportal.org). The GTEX release of October 2020 was used.

Statistical analyses

For each haplotype present in the CEU we counted the number of different *GAA* disease-associated variants observed in our dataset. In order to compute whether the observed number of disease-associated variants at each haplotype is the expected given frequency of the haplotype in the general population, we conducted a two-tailed Fisher exact test. $P < .05$ was considered as statistically significant.

RESULTS

Defining GAA haplotypes in the CEU population and in Pompe patients

Many common SNPs with a MAF ≥ 1% are present in the *GAA* gene, and those are not considered to cause Pompe disease [8], [18], [19]. However, these SNPs might influence the severity of disease-associated variants. To analyze *GAA* haplotypes in a systematic manner, we investigated haplotypes spanning the *GAA* (approximately 18Kb) to determine the frequency of common haplotypes and the distribution of Pompe disease-associated variants within these haplotypes. We set a threshold for the variants that have common MAF (≥ 5%). The population used in this analysis was derived from 1000G European individuals (CEU) as this group has a similar genetic ancestrality to the Dutch population [31]. We screened 83 common variants (≥ 5%) in the *GAA* that cover both coding and non-coding regions from position 80,101,556 to 80,119,881 on chromosome 17 (genome build GRCh38). Analysing the 83 variants on the 198 CEU phased alleles, we detected 26 distinct haplotypes, in which 16 haplotypes were present in more than one allele, and 10 were rare haplotypes being found in only one allele. Three haplotypes stand out in the CEU population: H1 (52/198; 26%), H2 (33/198; 17%) and H3 (29/198; 15%) (**Figure 1a**; **Table S3**).

To define the haplotypes of the 130 Pompe patients, we used the SNPs defined in the 654k SNP array. The allele-specific PCR could determine the phase of the alleles from exon 1 to 3, covering approximately 4kb of the *GAA* gene. Using the estimated haplotypes by phasing data from the SNP array, we could reflect the rest of the gene and confirming it by linkage analysis. Variants with MAF ≥ 5% were selected within *GAA* boundaries. There are 24 different *GAA* genotypes (i.e. one disease-associated in one chromosome and the other on second) in this patient cohort. All patients contain the common c.-32-13T>G (IVS1) variant on one allele and another Pompe disease-associated variant on the second allele

(**Table S2**). Each disease allele was manually annotated to its haplotype. Comparison between the observed number of different Pompe disease-associated variants by haplotype and the observed frequency of such haplotype in the general population showed statistically significant differences (**Figure 1b**, **Table 1**). Analysing all haplotypes in the CEU population, we could detect a significant difference (Fisher exact test – p-value 0.014) when compared to Pompe haplotypes.

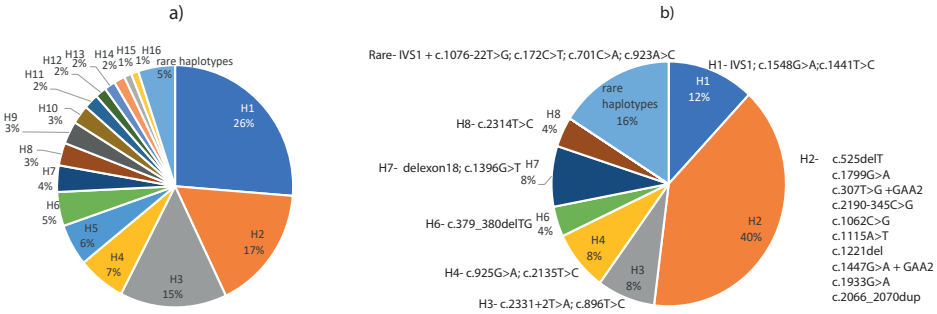


Figure 1. Comparison of *GAA* haplotypes in the CEU population and in Dutch Pompe patients. a) Analysis of 198 CEU *GAA* phased alleles based on 83 common variants distributed over 26 different haplotypes. Haplotypes are numbered and their frequency is indicated. b) As A, but now using 25 *GAA* alleles derived from 130 Dutch patients with childhood/adult onset Pompe disease.

Haplotype network analysis

The minimum spanning haplotype network of the haplotypes present in the CEU population and their enrichment/depletion in Pompe disease-associated variants was built to describe the topological relationship between both. The 26 distinct haplotypes identified in the 198 CEU alleles arose on two distal branches from the ancestral haplotype (**Figure 2**). The most common haplotype in CEU (H1) was present in a branch in which three disease-associated variants (3/21; 14,28%) were found: the variants IVS1 (124 alleles), c.1548G>A (8) and c.1441T>C (1). The other branch of haplotypes appears to be divided into two clusters, in which some recombination events are observed. All other Pompe disease alleles (18/21; 85,7%) were present on these clusters. Therefore, most Pompe disease alleles were present in clusters that did not contain the most common haplotypes in the normal CEU population. On the other hand, the IVS1 allele, which is the most frequent variant in European late-onset patients, was present in the most common haplotype, particularly in the CEU population.

Table 1. Distribution of the disease-associated alleles on the CEU haplotypes. The number of alleles within the total population of 130 patients is indicated.

Number of the disease-associated alleles on the haplotypes							
Haplotype 1	**Haplotype 2**	**Haplotype 3**	**Haplotype 4**	**Haplotype 6**	**Haplotype 7**	**Haplotype 8**	**Alleles in a rare haplotype**
IVS1 (130)	c.525delT (57)	c.2331+2T>A (2)	c.925G>A (7)	c.2314T>C (1)	c.2481+102_2646+31del538 (19)	c.379_380delTG (9)	IVS1 + c.1076-22T>G (2)
c.1548G>A (8)	c.1799G>A (6)	c.896T>C (1)	c.2135T>C (2)		c.1396G>T (1)		c.172C>T (2)
c.1441T>C (1)	c.307T>G + GAA2 (3)						c.701C>A (1)
	c.2190-345C>G (2)						c.923A>C (1)
	c.1062C>G (1)						
	c.1115A>T (1)						
	c.1221del (1)						
	c.1447G>A + GAA2 (1)						
	c.1933G>A (1)						
	c.2066_2070dup (1)						

() number of disease-associated alleles present in 130 Pompe patients.
IVS1+c.1076-22T>G, c.172C>T, c.701C>A and c.923A>C are present in a rare haplotypes not found in the 26 CEU haplotypes.

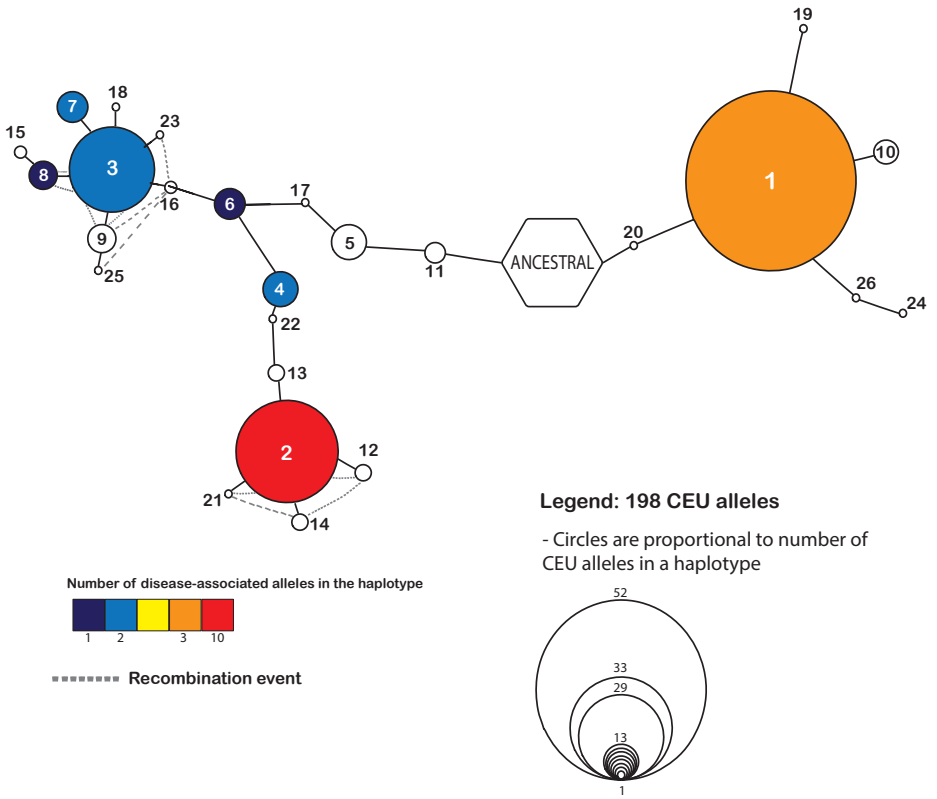


Figure 2. Network analysis of 26 different CEU haplotypes in the *GAA* illustrating the relationship between haplotypes that contain Pompe disease-associated variants in the Dutch population. Circles are proportional to the number of CEU alleles within a certain haplotype. The hexagon is the reference point for the ancestral

continues on next page

haplotype which was based on 100 ancestral haplotypes. Coloured circles represent the haplotypes harbouring disease-associated variants, in which red contains the largest number and dark blue the smallest number of disease-associated variants. Related haplotypes are connected by lines, in which the distance between the haplotypes reflects their relationship. Dashed lines indicate recombination events.

Linkage analysis and sub-haplotype blocks definition in the GAA gene

To understand how the variants relate to each other, we performed linkage disequilibrium (LD) analysis in a pairwise comparison for the 83 common sequence variants studied. We found that some parts of the *GAA* gene showed linkage of common variants, notably between exon 2 to exon 15, covering the N-terminal β-sheet domain, and parts encoding the catalytic GH31 domain (**Figure 3**; **Table S4**- CEU r^2 and D' values). The pairwise analysis showed that most of the common sequence variants were strongly linked. To understand better the linkage pattern, we performed a more in-depth analysis in order to identify possible sub-haplotype blocks within the *GAA*.

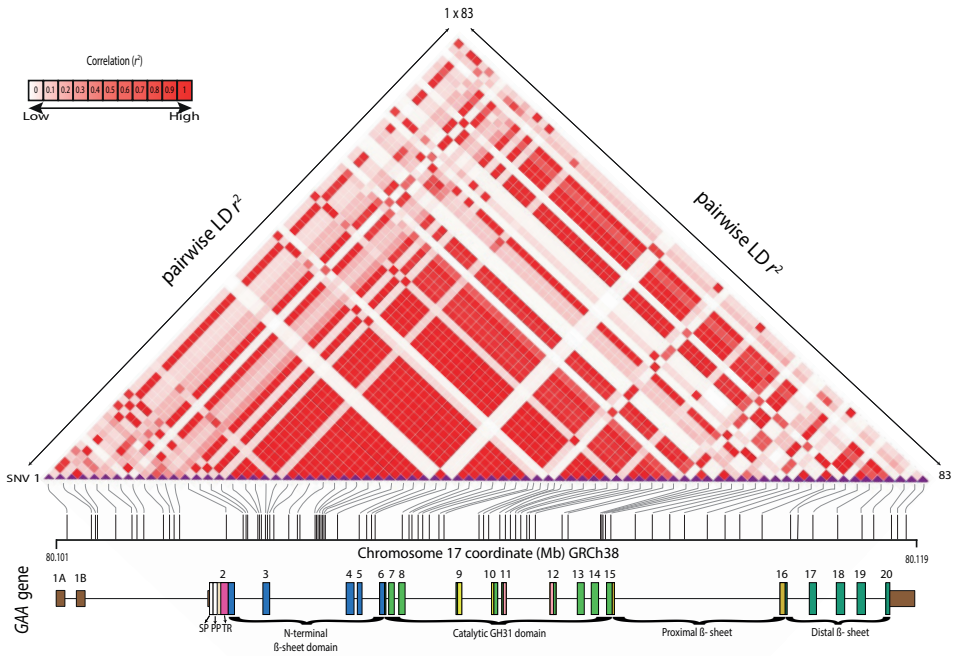


Figure 3. Heatmap matrix of pairwise Linkage Disequilibrium (LD) analysis based on the coefficient (r^2) among common SNVs in the GAA. 83 SNVs are shown in the heatmap and their position in the *GAA* is indicated. Exons are numbered and indicated, and protein domains are annotated. Coefficient (r^2) values range from 0 to 1, in which higher values indicate a higher degree of correlation and 0 indicates no correlation between variants.

To define sub-haplotypes, and to analyse the possibility that sub-haplotypes are present in blocks in the *GAA* gene, we built population-specific haplotype blocks. The values of these sub-haplotypes measures of linkage disequilibrium (D' and r2 CEU 1000G phase 3) are given in Supplemental **Figure S2**. We found that 69 out of 83 variants were present in sub-haplotype blocks. Those variants have similar allele frequencies within the global population (GnomAD) as well as in the Dutch population (GoNL) (**Table 2**). The analysis revealed the presence of seven sub-haplotype blocks. One block, block A, is present throughout the *GAA*, while the others cover distinct portions of the gene (**Figure 4**). The other variants for this population are considered to be independent or linked to a variant outside the *GAA* region. Next, sub-haplotypes were categorized per haplotype along with disease-associated variants. Haplotype 1 do not contain sub-haplotypes; haplotype 2 is composed of sub-blocks A, B and E; haplotype 3 and 8 are composed of sub-blocks A, D, E and F; haplotype 4 and 6 constitute of sub-haplotypes A and E; haplotype 7 is composed of sub-haplotypes A, E and F (**Table 3**).

Sub-haplotype blocks enriched by eQTLs

It is well known that the expression of a gene can be potentially influenced by common sequence variants (MAF > 1%) acting *in cis* [32]-[34]. DNA variants that affect gene expression are known as expression quantitative trait loci (eQTLs). eQTLs *in cis* can be detected as far as 1Mb from the gene transcription start site (TSS), and can act in a tissue-specific pattern. To test whether common variants present in *GAA* haplotypes are eQTLs, we analysed these variants using results from the Genotype-Tissue Expression (GTEx) project. The GTEx project collected 54 different tissues from nearly 1000 individuals. We chose to analyse the eQTLs for skeletal muscle which is the tissue that is the most affected in Pompe disease [2]. The p-value quantifies how the variant significantly changes gene expression in a particular tissue, and the normalised effect size computes the alternative allele's effect on expression relative to the reference allele. There were 337 eQTLs for skeletal muscle in the *GAA* within a 110kb window from the TSS. We found 64 eQTLs for skeletal muscle within the *GAA* coding region, of which 57 were in LD that constituted the sub-haplotype blocks. All eQTLs present in haplotype 2, sub-blocks A, B and E negatively affected GAA expression with 20-50% based on p-value and the normalised effect size. Similarly, sub-haplotype blocks A and E of haplotypes 3, 4, 6, 7, and 8 were associated with 20% lower GAA expression relative to the reference allele (**Figure 5**). Interestingly, this implies that disease-associated variants in haplotype 1, including IVS1, c.1548G>A, and c.1441T>C, are devoid of eQTLs that lower GAA expression, i.e. alleles containing these variants have 20 to 50% higher expression level relative to all other disease-associated variants detected in this cohort.

Table 2. Variants within the sub-haplotype blocks.

Sub-Haplotype Block A					Allele Frequency	
					GnomAD	GoNL
n	c. Position	chr 17 Grc 38	rs	where	Global	Dutch
1	c.-33+219G>C	80102109	rs4889961	Intron 1B	0,75	0,77
2	c.-33+1104A>G	80102994	rs11150841	Intron 1B	0,75	0,77
3	c.-33+1309T>C	80103199	rs1442314	Intron 1B	0,75	0,77
4	c.324T>C ; p.Cys108=	80104910	rs1800300	exon 2	0,72	0,77
5	c.547-243C>G	80105506	rs8065426	intron 2	0,66	0,75
6	c.547-67C>G	80105682	rs8069491	intron 2	0,66	0,76
7	c.547-39T>G	80105710	rs12452721	intron 2	0,66	0,76
8	c.547-4C>G	80105745	rs3816256	intron 2	0,66	0,76
9	c.596A>G ; p.His199Arg	80105798	rs1042393	exon 3	0,66	0,76
10	c.668G>A ; p.Arg223His	80105870	rs1042395	exon 3	0,66	0,76
11	c.692+144A>G	80106038	rs2304847	intron 3	0,66	0,75
12	c.692+509T>G	80106403	rs8082405	intron 3	0,66	0,76
13	c.692+674G>C	80106568	rs8078350	intron 3	0,67	0,76
14	c.692+751T>C	80106645	rs8068051	intron 3	0,65	0,75
15	c.693-585T>C	80106972	rs8068555	intron 3	0,65	0,75
16	c.693-559C>T	80106998	rs12602422	intron 3	0,65	0,75
17	c.693-491G>A	80107066	rs12948631	intron 3	0,67	0,75
18	c.693-441C>T	80107116	rs12602440	intron 3	0,67	0,75
19	c.693-434C>A	80107123	rs12941269	intron 3	0,65	0,75
20	c.693-414C>G	80107143	rs12941289	intron 3	0,65	0,75
21	c.693-413A>G	80107144	rs12937590	intron 3	0,66	0,71
22	c.693-216T>C	80107341	rs11150844	intron 3	0,67	0,75
23	c.955+12G>A	80107908	rs2252455	intron 5	0,69	0,75
24	c.956-107G>A	80108183	rs2241888	intron 5	0,73	0,77
25	c.956-84C>T	80108206	rs2241887	intron 5	0,65	0,75
26	c.1203G>A ; p.Gln401=	80108705	rs1800304	exon 8	0,65	0,75
27	c.1326+132G>A	80108960	rs894306	intron 8	0,65	0,75
28	c.1327-356G>T	80109589	rs6565640	intron 8	0,72	0,77
29	c.1327-269A>G	80109676	rs6565641	intron 8	0,65	0,75
30	c.1327-18A>G	80109927	rs2278619	intron 8	0,72	0,77
31	c.1438-220A>G	80110507	rs2278618	intron 9	0,65	0,75
32	c.1438-108G>A	80110619	rs12944802	intron 9	0,65	0,75
33	c.1438-19G>C	80110708	rs2304844	intron 9	0,65	0,75
34	c.1551+49C>A	80110889	rs2304843	intron 10	0,67	0,75
35	c.1636+404A>G	80111429	rs4889818	intron 11	0,73	0,77
36	c.2040+20A>G	80113047	rs2304836	intron 14	0,72	0,77
37	c.2189+263G>A	80113629	rs7221604	intron 15	0,65	0,76
38	c.2190-444A>G	80116524	rs4889967	intron 15	0,72	0,77
39	c.2331+20G>A	80117129	rs2304832	intron 16	0,80	0,77
40	c.2332-198A>T	80117402	rs2304830	intron 16	0,73	0,77
41	c.2338G>A ; p.Val780Ile	80117606	rs1126690	exon 17	0,72	0,77
42	c.2800-227C>T	80119045	rs9890469	intron 19	0,66	0,75

Sub-Haplotype Block B					Allele Frequency	
					GnomAD	GoNL
n	c. Position	chr 17 Grc 38	rs	where	Global	Dutch
43	c.-260G>C	80101663	rs2304849	Exon 1A	0,16	0,28
44	c.-33+316C>A	80102206	rs8077055	Intron 1B	0,20	0,16
45	c.-33+317C>T	80102207	rs8077056	Intron 1B	0,20	0,19
46	c.546+293G>A	80105425	rs34746710	Intron 2	0,20	0,29
47	c.547-238T>C	80105511	rs12452263	Intron 2	0,20	0,29
48	c.1327-179G>A	80109766	rs2278620	Intron 8	0,20	0,29
49	c.1581G>A ; p.Arg527=	80110970	rs1042396	Exon 11	0,20	0,29
50	c.2190-686G>A	80116282	rs12452616	Intron 15	0,20	0,29
51	c.*419G>T	80119750	rs7567	Exon 20 UTR	0,20	0,29

Sub-Haplotype Block C					Allele Frequency	
					GnomAD	GoNL
n	Position	chr 17 Grc 38	rs	where	Global	Dutch
52	c.-33+1190G>T	80103080	rs12602593	Intron 1B	0,08	0,10
53	c.2190-647G>A	80116321	rs59362713	Intron 15	0,08	0,10
54	c.2190-536G>A	80116432	rs60429724	Intron 15	0,08	0,10
55	c.2331+24T>C	80117133	rs2304831	Intron 16	0,14	0,12

Sub-Haplotype Block D					Allele Frequency	
					GnomAD	GoNL
n	Position	chr 17 Grc 38	rs	where	Global	Dutch
56	c.-32-1124C>T	80103431	rs58959690	Intron 1B	0,20	0,25
57	c.642C>T ; p.Ser214=	80105844	rs1800301	Exon 3	0,18	0,22

Sub-Haplotype Block E					Allele Frequency	
					GnomAD	GoNL
n	Position	chr 17 Grc 38	rs	where	Global	Dutch
58	c.1636+389C>T	80111414	rs7221675	Intron 11	0,63	0,71
59	c.1636+390A>G	80111415	rs7209921	Intron 11	0,63	0,71
60	c.1637-185A>G	80111798	rs12951255	Intron 11	0,55	0,65
61	c.1754+144C>T	80112244	rs2304838	Intron 12	0,61	0,71
62	c.2553G>A ; p.Gly851=	80118264	rs1042397	Exon 18	0,57	0,65

Sub-Haplotype Block F					Allele Frequency	
					GnomAD	GoNL
n	Position	chr 17 Grc 38	rs	where	Global	Dutch
63	c.-33+671A>C	80102561	rs55751636	intron 1B	0,29	0,35
64	c.-32-1298G>C	80103257	rs12602610	intron 1B	0,31	0,35

Sub-Haplotype Block G					Allele Frequency	
					GnomAD	GoNL
n	Position	chr 17 Grc 38	rs	where	Global	Dutch
65	c.1327-118A>G	80109827	rs74003628	intron 8	0,06	0,06
66	c.1754+12G>A	80112112	rs2304840	intron 12	0,06	0,06
67	c.2040+66C>T	80113093	rs2304835	Intron 14	0,06	0,06
68	c.2040+69A>G	80113096	rs2304834	intron 14	0,06	0,06
69	c.2189+607G>A	80113973	rs112710614	intron 15	0,06	0,06

DISCUSSION

There are several ways how the genomic context surrounding a particular disease-associated variant can affect its penetrance. The type of disease-associated variant is the first relevant parameter. For example, non-coding variants can affect gene expression outcomes such as transcription, mRNA stability, splicing, and epigenetic regulation. Coding variants can affect the protein function and stability such as folding, half-life, translation,

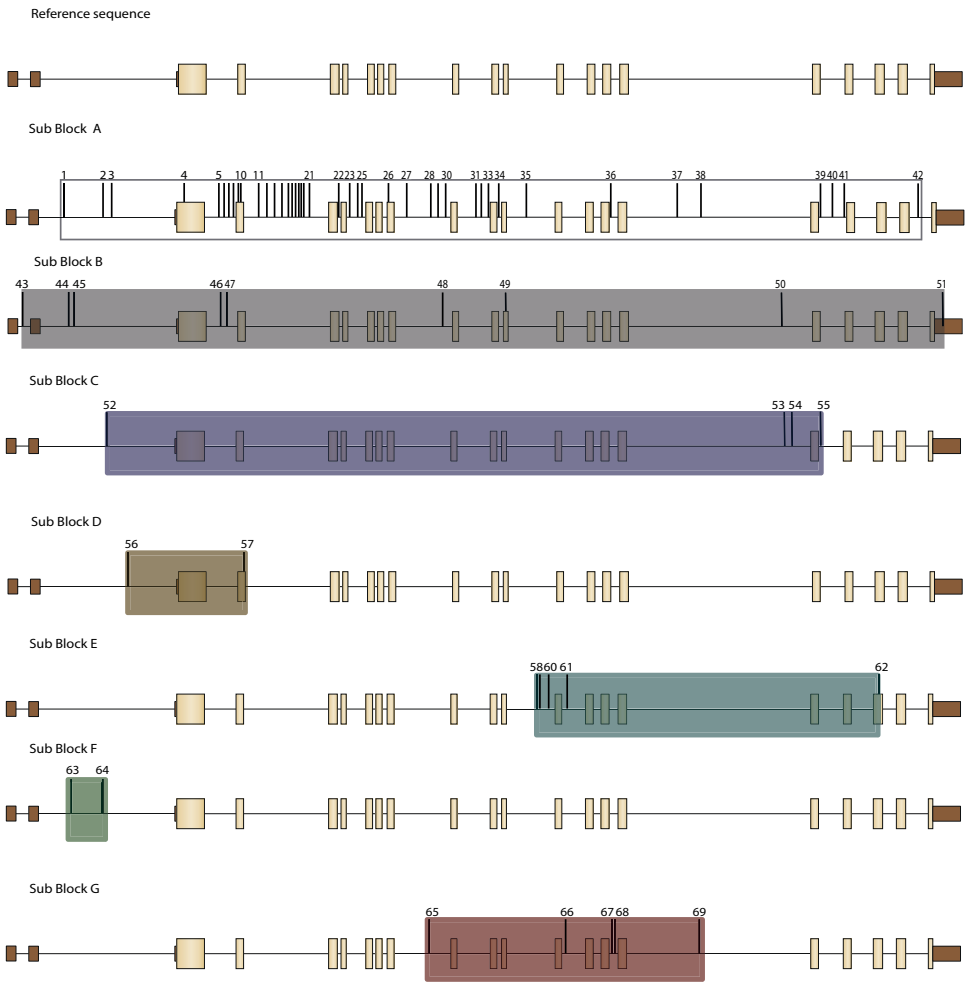


Figure 4. Analysis of sub-haplotype blocks in the *GAA*. The boundaries of the sub-haplotype blocks were defined by the first and last SNP for each haplotype block based on the LD matrix between pairs of SNVs, using tag SNPs to determine the sub-blocks using the algorithm implemented in Haploreg v4.1.

Table 3. Haplotype distribution of disease-associated variants

Distribution of the disease-associated variants on the haplotypes							
Haplotype 1	**sub block**	**Haplotype 2**	**sub block**	**Haplotype 3**	**sub block**	**Haplotype 4**	**sub block**
IVS1		c.525delT	A+B+E	c.2331+2T>A	A+D+E+F	c.925G>A	A+E
c.1548G>A		c.1799G>A	A+B+E	c.896T>C	A+D+E+F	c.2135T>C	A+E
c.1441T>C		c.307T>G + GAA2	A+B+E				
		c.2190-345C>G	A+B+E				
		c.1062C>G	A+B+E				
		c.1115A>T	A+B+E				
		c.1221del	A+B+E				
		c.1447G>A + GAA2	A+B+E				
		c.1933G>A	A+B+E				
		c.2066_2070dup	A+B+E				
Haplotype 6	**sub block**	**Haplotype 7**	**sub block**	**Haplotype 8**	**sub block**		
c.2314T>C	A+E	c.2481+102_2646+31del538	A+E+F	c.379_380delTG	A+D+E+F		
		c.1396G>T	A+E+F				

premature translation termination, post-translational modifications. Moreover, studies have shown that coding variants can also affect the penetrance of *cis*-acting regulatory variants [35], [36]. Irrespective of the underlying mechanism, it is known that disease-associated variants, in general, tend to be present in haplotypes that negatively affect their expression [36].

Gene-haplotype analysis could provide insight into allelic diversity among individuals. In the case of *GAA* and Pompe disease, haplotype analyses can help to identify potential modifiers for disease progression. Some modifying alleles have been already described: the Asian pseudodeficiency allele (c.[1726G> A (p.Gly576-Ser); 2065G>A (p.Glu689Lys)]) and the European GAA2 allele (c.271G>A, p.Asp91Asn)[19], [20]. The study performed by Labrousse *et al* 2010 showed that the Asian pseudodeficiency is present in a haplotype in which some disease-associated variants appeared to segregate together [19]. This led to the identification of a frequent haplotype that includes the Asian pseudodeficiency allele in Taiwanese newborns with low enzyme activity.

Two studies of haplotypes in the *GAA* present in Pompe patients of European descent have been reported previously [12], [37]. These studies analysed patients with the IVS1 disease-associated variant in one allele combined with a disease-associated variant on the second allele. The haplotypes were based on 18 variants. To build the *GAA* haplotypes based on all common SNVs in the general population, we have selected CEU individuals as they are genetically similar to Dutch individuals [31]. We performed an SNP array in our patient cohort to analyse the haplotypes that are linked to disease-associated variants. SNP array is a useful tool to screen common variants (detecting up to 99% positions); however, variants with low frequency is more prone to be missed due to the low sensitivity in detecting rare genetic variants [24]. In order to implement a comprehensive haplotype analysis, covering all common variants present within the *GAA*, we have extracted all variants with MAF ≥ 5% in the CEU population. However, this strategy

could also be a limitation because variants with a lower MAF will be lost. For example, we recently identified a genetic modifier of Pompe disease (*GAA* c.510C>T), which has a MAF of 0.005% in the CEU population. However, c.510C>T was present *in cis* with 27,3% of the IVS1 alleles in patients with symptom onset at childhood, while it was absent in patients with onset at adulthood, indicating that rare variants can modify the severity of disease-associated variants [15]. Despite the cutoff used in the present study, we could extract 83 common variants that defined 26 different haplotypes in the CEU population. All common SNVs identified were also common in the global population and they likely contribute to haplotypes in other non-CEU populations as well, which should be investigated in future studies.

Interestingly, the haplotype distribution of our patient cohort differed significantly from the haplotype distribution of the CEU population. The clustering of disease-associated alleles in a certain haplotype enriched by common derived variants, such as shown in **Figure 2**, might indicate that the haplotype combination could increase the penetrance of the disease-associated variants. In agreement, examples from the literature indicate that epistatic modifier effects can cause variable penetrance of an allele that can influence the severity of the phenotype [38]-[40]. This could be due to haplotype combinations whereby cis-regulatory variants increase the penetrance of disease-associated coding variants due to purifying selection [36], [41]. The presence of disease-associated variants in the *GAA* could be disadvantageous when present in specific haplotypes. Here, we show evidence of this phenomenon by analysing the eQTLS in our haplotypes studied in Pompe patients in the *GAA*, mainly for the haplotype 2, where this haplotype has a negative effect (**Figure 5**).

Furthermore, the IVS1 variant is present in the H1 haplotype. H1 can be related to the previous major *GAA* core haplotype published [12], in which our study compiles the entire haplotype that is present within the *GAA* boundaries. The H1 haplotype lacks SNPs, as was shown by allele-specific PCR covering approximately 4kb in IVS1 alleles [15]. Using SNP array phasing, we could relate the absence of derived variants to the rest of the *GAA* in our cohort. It is interesting to note that the IVS1 haplotype is closely related to the ancestral haplotype, yet it has a very high frequency for a disease-associated variant of 0.5% in the European population. This would be inconsistent with the deleterious effect of IVS1, which predicts removal from the population via purifying selection. Whether this may be explained by a positive selection due to a property that provides an evolutionary competitive advantage, such as has been reported for genes involved in energy metabolism (e.g. *LEPR* and *TCF7L2*), remains to be determined [42]-[46].

ACKNOWLEDGEMENTS

We thank Joana Carvalho Moreira de Mello for her considerable help in the bioinformatics part using the R programming language and in the discussions.

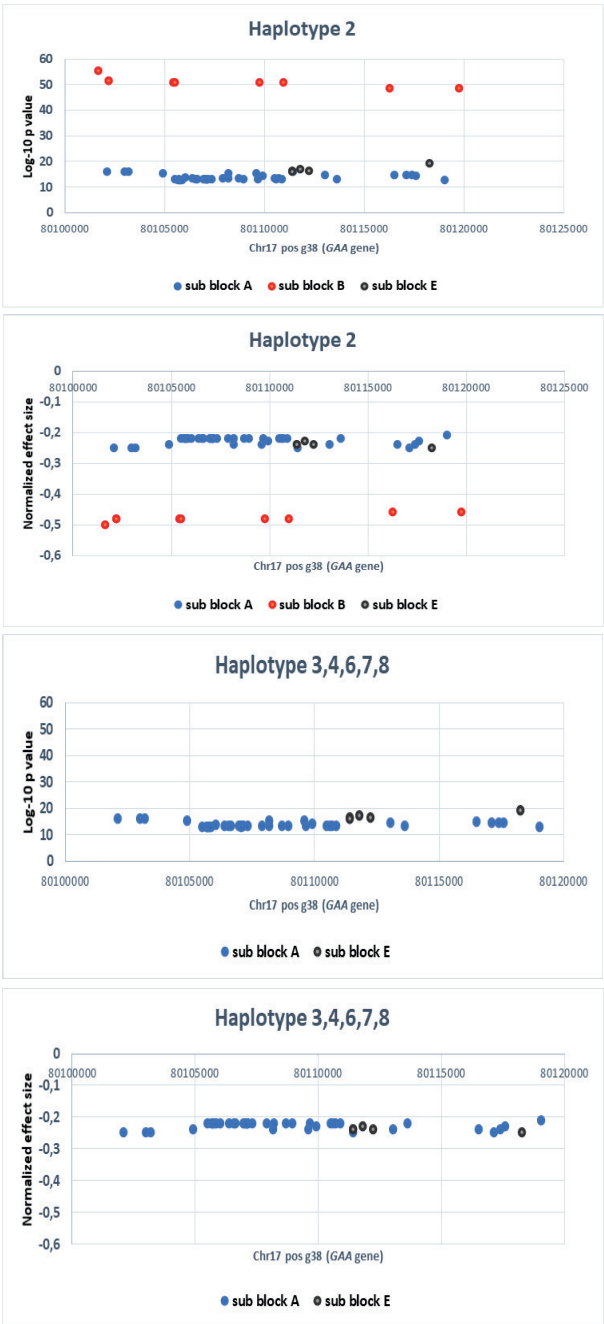


Figure 5. Distribution of *GAA* eQTLs for skeletal muscle tissue stratified for sub-haplotypes. On the first and third graphs, p-values are shown for the significance of the eQTL. On the second and fourth graphs, effect sizes on GAA expression are shown. eQTLs were extracted from the GTEx database.

REFERENCES

[1] D. Güngör and A. J. J. Reuser, How to describe the clinical spectrum in Pompe disease?, Am. J. Med. Genet. Part A, vol. 161, no. 2, pp. 399–400, Feb. 2013, doi: 10.1002/ajmg.a.35662.
[2] A. T. van der Ploeg and A. J. Reuser, "Pompe's disease," Lancet, vol. 372, no. 9646, pp. 1342–1353, 2008, doi: 10.1016/S0140-6736(08)61555-X.
[3] K. Momosaki et al., "Newborn screening for Pompe disease in Japan: report and literature review of mutations in the GAA gene in Japanese and Asian patients," J. Hum. Genet., vol. 64, no. 8, pp. 741–755, Aug. 2019, doi: 10.1038/s10038-019-0603-7.
[4] M. G. E. M. Ausems et al., "Frequency of glycogen storage disease type II in The Netherlands: Implications for diagnosis and genetic counselling," Eur. J. Hum. Genet., vol. 7, no. 6, pp. 713–716, 1999, doi: 10.1038/sj.ejhg.5200367.
[5] R. Pinto et al., "Prevalence of lysosomal storage diseases in Portugal," Eur. J. Hum. Genet., vol. 12, no. 2, pp. 87–92, Feb. 2004, doi: 10.1038/sj.ejhg.5201044.
[6] Y. H. Chien et al., "Long-term prognosis of patients with infantile-onset pompe disease diagnosed by newborn screening and treated since birth," J. Pediatr., vol. 166, no. 4, pp. 985-991.e2, 2015, doi: 10.1016/j.jpeds.2014.10.068.
[7] Z. Lukacs et al., "Prevalence of Pompe disease in 3,076 patients with hyperCKemia and limb-girdle muscular weakness," Neurology, vol. 87, no. 3, pp. 295–298, Jul. 2016, doi: 10.1212/WNL.0000000000002758.
[8] D. O. S. Faria et al., "Update of the Pompe variant database for the prediction of clinical phenotypes: Novel disease-associated variants, common sequence variants, and results from newborn screening," Hum. Mutat., vol. 42, no. 2, pp. 119–134, Feb. 2021, doi: 10.1002/humu.24148.
[9] M. Y. Niño et al., "Extension of the Pompe mutation database by linking disease-associated variants to clinical severity," Hum. Mutat., vol. 40, no. 11, pp. 1954–1967, Nov. 2019, doi: 10.1002/humu.23854.
[10] E. van der Wal, A. J. Bergsma, J. M. Pijnenburg, A. T. van der Ploeg, and W. W. M. P. Pijnappel, "Antisense Oligonucleotides Promote Exon Inclusion and Correct the Common c.-32-13T>G GAA Splicing Variant in Pompe Disease," Mol. Ther. - Nucleic Acids, vol. 7, pp. 90–100, Jun. 2017, doi: 10.1016/j.omtn.2017.03.001.
[11] C. F. Boerkoel et al., "Leaky splicing mutation in the acid maltase gene is associated with delayed onset of glycogenosis type II," Am. J. Hum. Genet., vol. 56, no. 4, pp. 887–897, 1995, Accessed: Apr. 19, 2021. [Online]. Available: /pmc/articles/PMC1801206/?report=abstract.
[12] M. A. Kroos et al., "Broad spectrum of Pompe disease in patients with the same c.-32-13T→G haplotype," Neurology, vol. 68, no. 2, pp. 110–115, Jan. 2007, doi: 10.1212/01.wnl.0000252798.25690.76.
[13] M. Kroos, M. Hoogeveen-Westerveld, A. van der Ploeg, and A. J. J. Reuser, "The genotype-phenotype correlation in Pompe disease," Am. J. Med. Genet. Part C Semin. Med. Genet., vol. 160 C, no. 1, pp. 59–68, Feb. 2012, doi: 10.1002/ajmg.c.31318.
[14] S. C. A. Wens et al., "Phenotypical variation within 22 families with Pompe disease," Orphanet J. Rare Dis., vol. 8, no. 1, 2013, doi: 10.1186/1750-1172-8-182.
[15] A. J. Bergsma et al., "A genetic modifier of symptom onset in Pompe disease," EBioMedicine, vol. 43, pp. 553–561, May 2019, doi: 10.1016/j.ebiom.2019.03.048.
[16] P. De Filippi et al., "Genotype-phenotype correlation in Pompe disease, a step forward," Orphanet J. Rare Dis., vol. 9, no. 1, p. 102, Aug. 2014, doi: 10.1186/s13023-014-0102-z.
[17] E. Kuperus et al., "The ACE I/D polymorphism does not explain heterogeneity of natural course and response to enzyme replacement therapy in Pompe disease," PLoS One, vol. 13, no. 12, Dec. 2018, doi: 10.1371/journal.pone.0208854.
[18] L. T. Turaça et al., "Novel GAA mutations in patients with Pompe disease," Gene, vol. 561, no. 1, 2015, doi: 10.1016/j.gene.2015.02.023.
[19] P. Labrousse et al., "Genetic heterozygosity and pseudodeficiency in the Pompe disease newborn screening pilot program," Mol. Genet. Metab., vol. 99, no. 4, pp. 379–383, Apr. 2010, doi: 10.1016/j.ymgme.2009.12.014.
[20] M. Kroos et al., "Seven cases of Pompe disease from Greece," J. Inherit. Metab. Dis., vol. 29, no. 4, pp. 556–563, Aug. 2006, doi: 10.1007/s10545-006-0280-5.

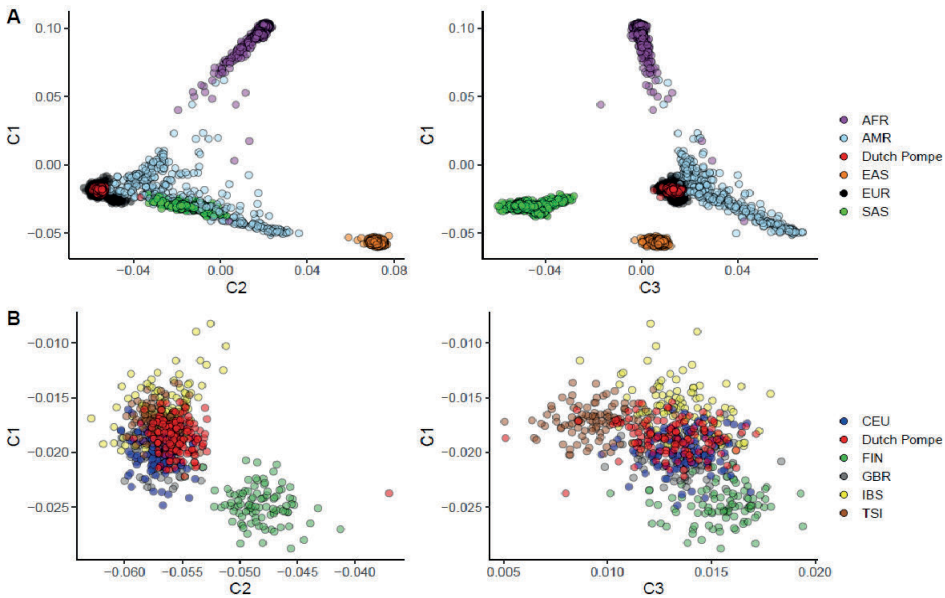
[21] M. A. Kroos et al., "p.[G576S; E689K]: Pathogenic combination or polymorphism in Pompe disease?," Eur. J. Hum. Genet., vol. 16, no. 8, pp. 875–879, Aug. 2008, doi: 10.1038/ejhg.2008.34.
[22] A. Auton et al., "A global reference for human genetic variation," Nature, vol. 526, no. 7571. Nature Publishing Group, pp. 68–74, Sep. 30, 2015, doi: 10.1038/nature15393.
[23] S. Purcell et al., "PLINK: A tool set for whole-genome association and population-based linkage analyses," Am. J. Hum. Genet., vol. 81, no. 3, pp. 559–575, 2007, doi: 10.1086/519795.
[24] M. W et al., "Use of SNP chips to detect rare pathogenic variants: retrospective, population based diagnostic evaluation," BMJ, vol. 372, 2021, doi: 10.1136/BMJ.N214.
[25] X. Sun, J. Namkung, X. Zhu, and R. C. Elston, "Capability of common SNPs to tag rare variants," BMC Proc., vol. 5, no. S9, Dec. 2011, doi: 10.1186/1753-6561-5-s9-s88.
[26] M. N. Saad, M. S. Mabrouk, A. M. Eldeib, and O. G. Shaker, "Identification of rheumatoid arthritis biomarkers based on single nucleotide polymorphisms and haplotype blocks: A systematic review and meta-analysis," Journal of Advanced Research, vol. 7, no. 1. Elsevier, pp. 1–16, Jan. 01, 2016, doi: 10.1016/j.jare.2015.01.008.
[27] B. Howie, C. Fuchsberger, M. Stephens, J. Marchini, and G. R. Abecasis, "Fast and accurate genotype imputation in genome-wide association studies through pre-phasing," Nat. Genet., vol. 44, no. 8, pp. 955–959, Aug. 2012, doi: 10.1038/ng.2354.
[28] Y. Li, C. J. Willer, J. Ding, P. Scheet, and G. R. Abecasis, "MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes," Genet. Epidemiol., vol. 34, no. 8, pp. 816–834, Dec. 2010, doi: 10.1002/gepi.20533.
[29] E. Paradis, "pegas: an R package for population genetics with an integrated-modular approach," Bioinformatics, vol. 26, no. 3, pp. 419–420, Feb. 2010, doi: 10.1093/bioinformatics/btp696.
[30] L. D. Ward and M. Kellis, "HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants," Nucleic Acids Res., vol. 40, no. D1, pp. D930–D934, Jan. 2012, doi: 10.1093/nar/gkr917.
[31] L. C. Francioli et al., "Whole-genome sequence variation, population structure and demographic history of the Dutch population," Nat. Genet., vol. 46, no. 8, pp. 818–825, Jun. 2014, doi: 10.1038/ng.3021.
[32] H. J. Westra and L. Franke, "From genome to function by studying eQTLs," Biochimica et Biophysica Acta - Molecular Basis of Disease, vol. 1842, no. 10. Elsevier B.V., pp. 1896–1902, Oct. 01, 2014, doi: 10.1016/j.bbadis.2014.04.024.
[33] F. Aguet et al., "The GTEx Consortium atlas of genetic regulatory effects across human tissues," Science (80-.)., vol. 369, no. 6509, pp. 1318–1330, Sep. 2020, doi: 10.1126/SCIENCE.AAZ1776.
[34] S. Kim-Hellmuth et al., "Cell type-specific genetic regulation of gene expression across human tissues," Science (80-.)., vol. 369, no. 6509, Sep. 2020, doi: 10.1126/SCIENCE.AAZ8528.
[35] C. Tcheandjieu et al., "A phenome-wide association study of 26 mendelian genes reveals phenotypic expressivity of common and rare variants within the general population," PLOS Genet., vol. 16, no. 11, p. e1008802, Nov. 2020, doi: 10.1371/journal.pgen.1008802.
[36] S. E. Castel et al., "Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk," Nat. Genet., vol. 50, no. 9, pp. 1327–1334, Sep. 2018, doi: 10.1038/s41588-018-0192-y.
[37] L. Gort, M. J. Coll, and A. Chabás, Glycogen storage disease type II in Spanish patients: High frequency of c.1076-1G>C mutation, Mol. Genet. Metab., vol. 92, no. 1–2, pp. 183–187, Sep. 2007, doi: 10.1016/j.ymgme.2007.05.011.
[38] A. E. Fish, J. A. Capra, and W. S. Bush, "Are Interactions between cis-Regulatory Variants Evidence for Biological Epistasis or Statistical Artifacts?," Am. J. Hum. Genet., vol. 99, no. 4, pp. 817–830, Oct. 2016, doi: 10.1016/j.ajhg.2016.07.022.
[39] D. N. Cooper, M. Krawczak, C. Polychronakos, C. Tyler-Smith, and H. Kehrer-Sawatzki, "Where genotype is not predictive of phenotype: Towards an understanding of the molecular basis of reduced penetrance in human inherited disease," Human Genetics, vol. 132, no. 10. Springer, pp. 1077–1130, Oct. 03, 2013, doi: 10.1007/s00439-013-1331-2.
[40] J. Domingo, P. Baeza-Centurion, and B. Lehner, "The Causes and Consequences of Genetic Interactions (Epistasis)," Annual Review of Genomics and Human Genetics, vol. 20. Annual Reviews Inc., pp. 433–460, Aug. 31, 2019, doi: 10.1146/annurev-genom-083118-014857.
[41] T. Lappalainen, S. B. Montgomery, A. C. Nica, and E. T. Dermitzakis, "Epistatic selection between coding and regulatory variation in human evolution and disease," Am. J. Hum. Genet., vol. 89,

no. 3, pp. 459–463, Sep. 2011, doi: 10.1016/j.ajhg.2011.08.004.
[42] L. Ségurel et al., "Positive selection of protective variants for type 2 diabetes from the Neolithic onward: A case study in Central Asia," Eur. J. Hum. Genet., vol. 21, no. 10, pp. 1146–1151, Oct. 2013, doi: 10.1038/ejhg.2012.295.
[43] A. M. Hancock et al., "Adaptations to Climate in Candidate Genes for Common Metabolic Disorders," PLoS Genet., vol. 4, no. 2, p. e32, Feb. 2008, doi: 10.1371/journal.pgen.0040032.
[44] A. Helgason et al., "Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution," Nat. Genet., vol. 39, no. 2, pp. 218–225, Feb. 2007, doi: 10.1038/ng1960.
[45] Y. C. Klimentidis, M. Abrams, J. Wang, J. R. Fernandez, and D. B. Allison, "Natural selection at genomic regions associated with obesity and type-2 diabetes: East Asians and sub-Saharan Africans exhibit high levels of differentiation at type-2 diabetes regions," Hum. Genet., vol. 129, no. 4, pp. 407–418, Apr. 2011, doi: 10.1007/s00439-010-0935-z.
[46] B. F. Voight, S. Kudaravalli, X. Wen, and J. K. Pritchard, "A Map of Recent Positive Selection in the Human Genome," PLoS Biol., vol. 4, no. 3, p. e72, Mar. 2006, doi: 10.1371/journal.pbio.0040072.

SUPPLEMENTARY INFORMATION

Table S1. 83 SNVs with MAF≥5% extracted from CEU in 1000genomes
Table S2. List of patients with genotype
Table S3. 83 SNVs and their distribution in the 26 CEU haplotypes
Table S4. Pairwise LD (CEU r^2 and D' values) from 83 SNVs

The supplementary tables S1, S2, S3 and S4 can be accessed at:
shorturl.at/lsMTW

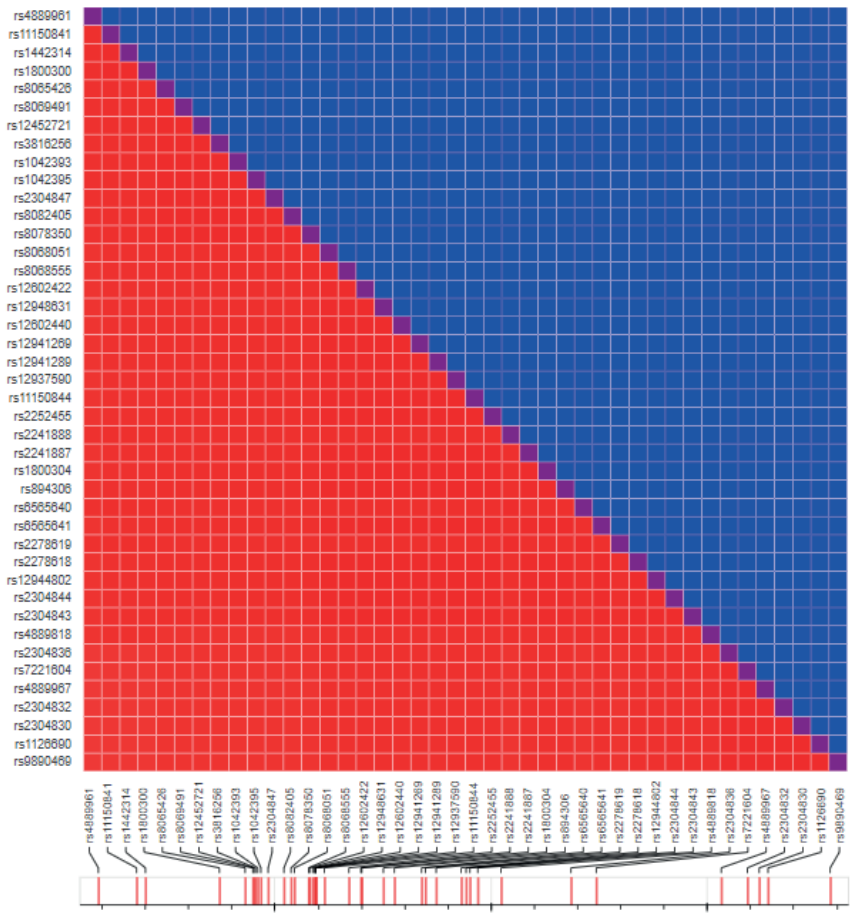


A) Major groups and Dutch Pompe patients
B) European sub-groups and Dutch Pompe patients

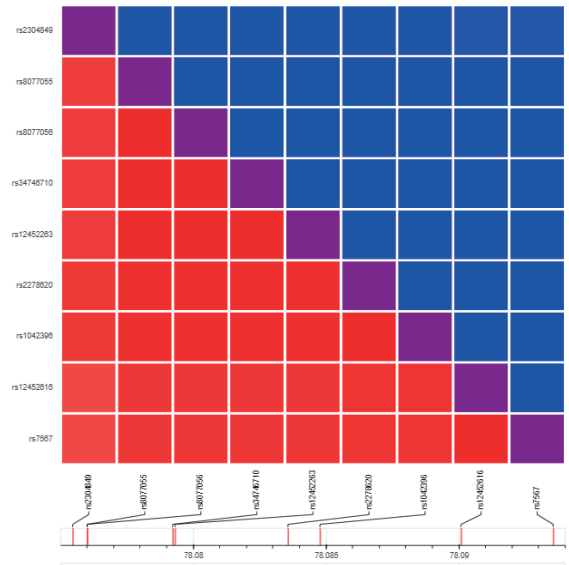
Figure S1. Multidimensional Scaling (MDS) analysis on an identical by state (IBS) matrix computed with Pompe patients and 1000G individuals.

Figure S2. The values of the sub-blocks measures of linkage disequilibrium (D'and r^2 CEU 1000G phase 3). Sub-blocks A to G

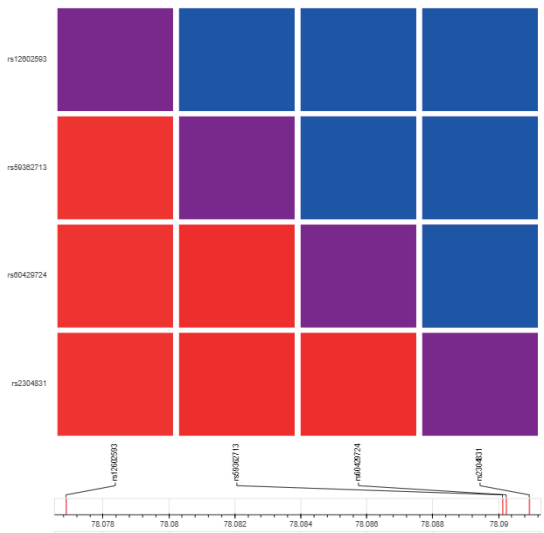
Haplotype sub-block A



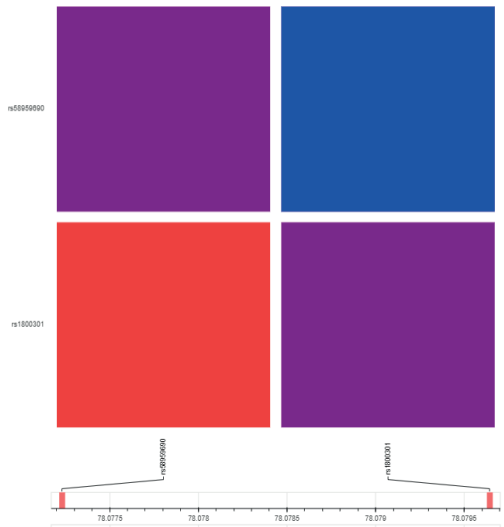
Haplotype sub-block B



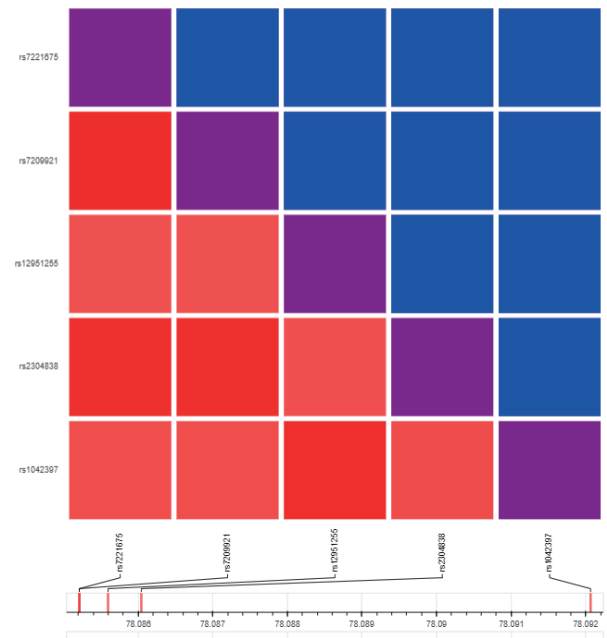
Haplotype sub-block C



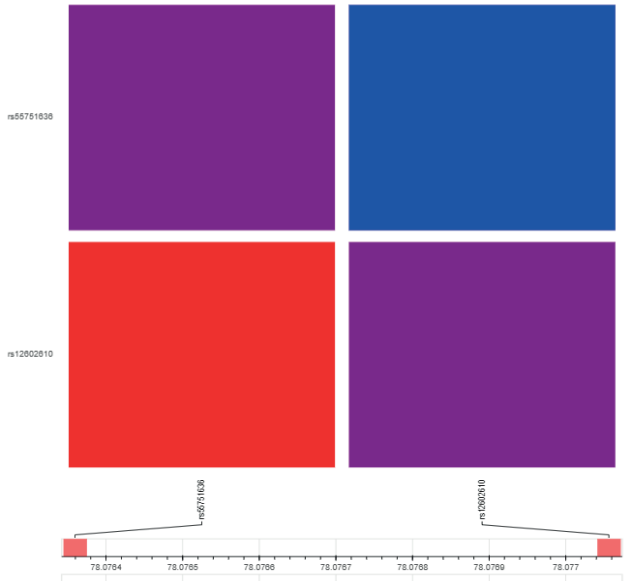
Haplotype sub-block D



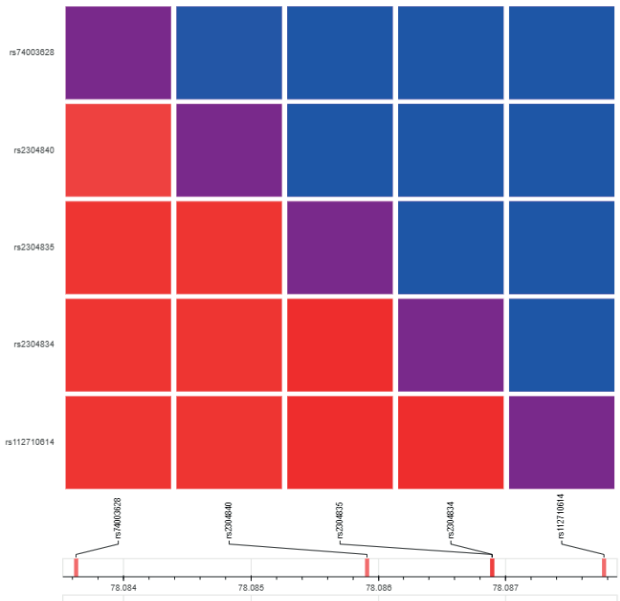
Haplotype sub-block E



Haplotype sub-block F



Haplotype sub-block G



CHAPTER 7

Clinical diversity of Pompe patients with c.-32-13T>G genotypes investigated by gene expression profiling

Monica Y. Niño[1, 2, 3], **Douglas O. S. de Faria[1,2, 3],** Marian Kroos[1], Marianne Hoogeveen-Westerveld [1], Stijn L.M. in 't Groen[1,2,3], J.M.P. van Den Hout[4], Wilfred Van IJcken[4], Ans T. van der Ploeg[2,3], W.W.M. Pijnappel[1,2,3]

1Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands

2Department of Pediatrics, Erasmus University Medical Center, Rotterdam, The Netherlands

3Center for Lysosomal and Metabolic Diseases, Erasmus University Medical Center, Rotterdam, The Netherlands

4Erasmus Center for Biomics, Erasmus MC University Medical Center, Rotterdam, The Netherlands.

Correspondence: WWMP Pijnappel, Department of Clinical Genetics, Erasmus University Medical Center, P.O.Box 2040, 3000 CA Rotterdam. The Netherlands E-mail: w.pijnappel@erasmusmc.nl

Manuscript in preparation

Clinical diversity of Pompe patients with c.-32-13T>G genotypes investigated by gene expression profiling

Monica Y. Niño, **Douglas O. S. de Faria**, Marian Kroos, Marianne Hoogeveen-Westerveld, Stijn L.M. in 't Groen, J.M.P. van Den Hout, Wilfred Van IJcken, Ans T. van der Ploeg, W.W.M. Pijnappel

ABSTRACT

Background: Pompe disease is an autosomal recessive neuromuscular disorder caused by acid a-glucosidase (GAA) deficiency. The disease presents as a clinical spectrum with regard to age of onset, symptom severity and rate of disease progression. Notably, clinical diversity occurs also among patients with *GAA* genotypes that exert the same functional effect on GAA activity. This observation points to the interaction of factors co-determining the clinical course of Pompe disease. We hypothesized that differences in gene expression profiles result in modulation of secondary responses and outcomes of disease progression while the level of GAA deficiency is the same. The differentially expressed genes (DEGs) provide opportunities for a better understanding of the molecular pathogenesis of Pompe disease and may at the same time serve as markers of disease progression.

Objective: The aim of our present study was to identify transcriptional differences in cultured fibroblasts from patients with childhood and adult onset Pompe disease.

Materials and methods: RNA sequencing (RNA−seq) was applied to cultured fibroblasts from patients with *GAA* genotypes (c.−32−13T>G / 'null', in 35/39 cases). The patients were selected on the basis of ages of onset. Seventeen had childhood onset (<16 years) and 22 adult onset (>35 years) Pompe disease. Patients and controls were analyzed in 2 separate analyses. Study 1 was the initial study, while study 2 was an independent study using individuals to validate the results of study 1. For comparison, we investigated gene expression also in fibroblasts from 19 childhood and 18 adult, age-matched, unaffected individuals. Gene ontology analysis was performed using Panther and Ingenuity software.

Results: In study 1, the median age (years) at symptom onset was 1.25 in childhood onset patients, and 47 years in adulthood onset patients. These numbers were 7 and 33 in study 2, respectively. RNA-seq revealed 569 differentially expressed genes (DEGs) between childhood and adult Pompe patients in study 1. A substantial number of these DEGs were linked to known metabolic and cellular processes. The number of DGEs identified in the replicate study 2 was only 7, but a core set of 3 disease related DEGs could be established between childhood and adult phenotypes: *MAOA*, *STC1*, and *NRK1*. When comparing expression profiles between Pompe patients and healthy controls, we found 6607 DEGs for adults and 4426 DEGs for children in study 1. For study 2 these numbers were 53 and 89, respectively. Importantly, only one gene overlapped when comparing genes that were differentially expressed in adults plus children with Pompe disease compared to age matched controls, which was *GAA*. This result strongly supports the validity of the chosen methodology and the data analysis.

Conclusion: Our study shows that fibroblasts can be used for detection of genome- wide transcriptional differences between patients with Pompe disease and unaffected individuals but also between Pompe disease phenotypes. However, large individual differences in gene expression within both healthy controls and patients prevents the reproducible identification of most disease-regulated genes. *MAOA*, *STC1*, and *NRK1* were the exceptions and might be associated with the onset of symptoms in Pompe disease. This should be confirmed in larger cohorts.

INTRODUCTION

Pompe disease (MIM# 232300) or glycogen storage type II (GSDII) is an autosomal recessive disorder caused by acid a-glucosidase (GAA; NP_000143.2) deficiency, a lysosomal enzyme degrading glycogen to glucose. The combination of disease-associated variants in the two copies of the *GAA* gene causes deficiency or total absence of GAA enzyme activity resulting in intra-lysosomal accumulation of glycogen with profound degenerative effects on muscle cells. Continuous accumulation of glycogen prohibits the normal functioning of lysosomes, disturbs the autophagic transport pathways, interferes with cellular metabolism, and may even lead to rupture of lysosomes, all together resulting in cellular damage and progressive degeneration of skeletal muscles. Pompe disease presents as a spectrum of phenotypes. Apart from the rapidly fatal classic infantile phenotype, characterized by death within the first year of life due to cardiorespiratory failure [1], there are more slowly progressive phenotypes exemplified by children and adults presenting with skeletal and/or respiratory muscles

weakness at older age [2-6].

In 2006, enzyme replacement therapy (ERT) using recombinant DNA technology was approved for the treatment of Pompe disease based on the outcome of clinical trials in infants [7, 8]. Currently, ERT is commercially available as the only treatment for Pompe disease. Studies in infants have demonstrated that human recombinant GAA (rhGAA) is capable of improving cardiac and skeletal functions and longer-term survival by reducing lysosomal glycogen accumulation [7, 9, 10]. Studies focusing on affected children and adults have shown that ERT, at the group level, improves muscle function (distance walked during 6 minutes), muscle strength and pulmonary function, and has long-term benefit [8, 11-13]. However, patients' response to ERT with rhGAA is highly variable [13, 14]. This variation could be due to a multitude of factors among which, obviously, the severity of symptoms at age of diagnosis and the disease duration at start of treatment. Also, other factors play a role such as antibody formation against the administered rhGAA, which is generally considered as a negative outcome predictor and in most cases but not all associated with a CRIM negative status [15-17].

Optimal management of patients with Pompe disease depends on early diagnosis, intelligent prediction of disease progression, coordinated follow up, and timely initiation of ERT. However, other than the patients' *GAA* genotype, their level of residual GAA activity, and their clinical signs, there are hardly any biochemical/ molecular disease-markers to be guided by. Currently, Glucose tetrasaccharide (Glc4) is used as biomarker in Pompe disease; patients with Pompe disease typically excrete increased amounts of Glc4 in urine, which can be analyzed by high performance liquid chromatography (HPLC). A decrease in the level of Glc4 has been reported in patients responding to ERT. However, raised levels of Glc4 have been observed also in patients with glycogen storage diseases (GSDs) type Ia and type III suggesting that Glc4 is not a highly sensitive marker for Pompe disease [18]. Myostatin and insulin- like growth factor 1 (*IGF-1*) have been proposed as potential therapeutic biomarkers in Pompe disease as well. Those molecules have shown lower serum levels in Pompe patients than in healthy controls. After ERT, the expression levels of both myostatin and IGF-1 reached normal levels, suggesting possible utility for disease monitoring [19].

Recent studies have suggested a group of miRNAs as potential biomarkers of Pompedisease. The validation of the results identified three miRNAs, suggesting that may represent additional biomarkers for the follow-up of adult onset Pompe disease [20, 21]. At present there are no established methods that can facilitate and provide a better understanding of decision-making on therapeutic intervention. This difficult task has prompted the need to develop methods for prediction of disease progression and start of therapeutic intervention.

In our search for genes and associated pathways that could potentially affect the clinical course of Pompe disease and serve as biochemical / molecular

disease-marker we analyzed the transcriptome of cultured skin fibroblasts from two groups of patientswith functionally 'the same' *GAA* genotype but quite different age of onset of clinical symptoms, patients with childhood onset and patients with adult onset. Age matched healthy individuals were analyzed as controls. To enable the collection of a sufficiently large patient cohort with functionally 'the same' *GAA* genotypes we have chosen in our study for affected children and adults with the common c.-32-13T>G/ disease- associated variant genotypes (-32-13T>G/ null, in 35/39 cases). Patients with the IVS1/ null genotype may present with symptoms and any age from early childhood to lateadulthood [22].

The c.-32-13T>G transversion upstream of the splice acceptor site of exon 2 results in aberrant splicing whereby only 5-15% of the normal amount of structurally and functionally normal acid a-glucosidase is produced. c.-32-13T>G is the most frequent *GAA* variant among Caucasians with Pompe disease and is observed in more than two-thirds of the patients worldwide [23-25]. In the Netherlands, about 95% of adults with Pompe disease and 68% of affected children under 18 years have the c.-32-13T>G / 'null' genotype [22, 23].

MATERIAL AND METHODS

Study population

Inclusion criteria and study design

An overview of the entire experimental setting performed in study 1 and study 2 is provided in **Figure 1**.

Patients had to be diagnosed with Pompe disease during childhood or adulthood. For genome-wide expression analysis, all patients (39) had to be compound heterozygote for the common c.-32-13T>G *GAA* gene variant in combination with a disease-associated variant on the second *GAA* allele (which was null in 35 cases and presumably null in 4 cases). Patients with substantially different ages of onset (<8 years vs >35 years) were selected for the study. Cohort 1 comprised 17 patients with childhood onset of Pompe disease (13 male, 4 female) with a median age at symptom onset of 2 years (range 0.5- 13). Cohort 2 consisted of 22 patients with adult onset Pompe disease (8 male, 14 female) with a median age at onset of symptoms of 41.5 years (range 20-62). In addition, a total of 37 unaffected individuals were included in the study: 19 children (12 males, 7 females) with a median age at time of skin biopsy of 9 years (range 1.2-14.2), and 18 adults (9 males, 9 females) with a median age at time of biopsy of 52.5 years (range 35-76.9). Biopsies were taken from either the skin of the proximal part of the leg (patients) or form the distal part of the leg or from the foreskin (controls). Characteristics of study subjects and study materials

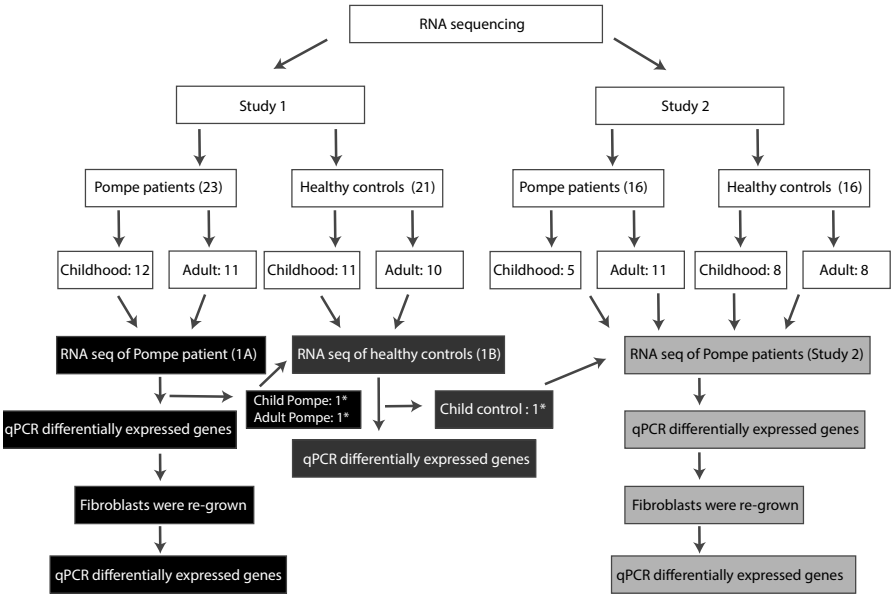


Figure 1. Workflow of the generic approach for the RNA sequencing analysis in Pompe patients and healthy controls included in studies 1 and 2. The differentially expressed genes (DEGs) were analyzed by qPCR. Afterwards, skin fibroblasts derived from Pompe patients were re-grown (1A and 2), and qPCR of DEGs as performed to confirm the RNAseq findings. *Samples used as internal control.

including ages at biopsy) are provided in **Table 1** and **Figure 2A**. To further validate the results obtained a skin biopsy, a third cohort of Pompe patients (3 childhood, 3 adult and 1 classic infantile) were analyzed. The Medical Ethical Committee at Erasmus University MC approved the study protocol, and all patients, or their parentsor legal guardians provided written informed consent.

For several reasons it was not possible to incorporate all RNAseq measurements in a single study. Therefore, the analyses were divided as follows: study 1 (all from cohort 1) was performed in two experiments, study 1A (12 children with evident symptoms under the age of 8 years and 11 adults with symptoms over 35 years); and study 1B (11 unaffected children and 10 unaffected adults with skin biopsies taken under the age of 16 or over the age of 35, respectively), 3 samples were chosen for measuring RNA profiles in both study 1A and study 1B to determine a potential batch effect, which was absent. Study 2 (all from cohort 2) included 8 unaffected and 5 affected children, and 8 unaffected and 11 affected adults with skin biopsies taken under the age of 16 (childhood) or over the age of 35 (adult), respectively (**Fig. 1** and **Fig. 2A**).

Table 1. Demographics of Patients and Controls in studies 1 and 2

Criteria	Childhood Pompe	Adult Pompe	Childhood control	Adult control
Age:				
at onset (y)	<8*	>35#		
at biopsy (y)	<16	>35	<16	>35
Gender:				
male	13	8	12	9
female	4	14	7	9
Biopsy:				
leg	17	22	13	18
foreskin			6	
Total	17	22	19	18

*Patients seen by doctors at indicated age (<8 years) with clear signs of Pompe disease. #Patients with first signs at indicated age (>35 years) and seen by doctors at that age.

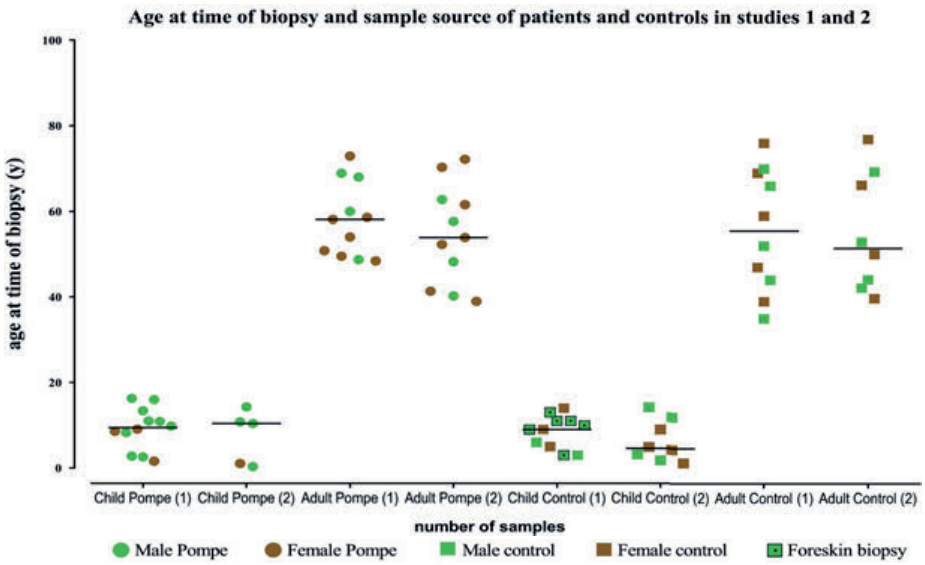


Figure 2A. Patient's age at time of biopsy and sample source. Skin biopsies were taken from the upper lower limb or the lower limb except for 6 samples whereby served as sample source (foreskin fibroblasts are represented as dotted green square). The study number is given within brackets. The horizontal line presents the median value.

Cell culture, GAA enzyme activity and glycogen content

Skin fibroblasts were cultured according to routine procedures in medium containing DMEM supplemented with 10% fetal calf serum and 1% penicillin/streptomycin, in a humidified atmosphere containing 5% CO_2/95% air, at 37° C. The cells were harvested with trypsin/EDTA (ethylenediaminetetraacetic acid) at day 3 after the last subculture

and stored as pellet at -80°C until use. Cell lysates were prepared and assayed for GAA activity and protein, as mentioned below, and for measuring the glycogen content according to Umapathysivan et al, 2005; whereby the cells were kept for 3 to 5 days on glucose-free medium before harvesting [26].

The standard GAA enzyme activity assay mixture consisted of 10 µL of cell homogenate (10 µg of protein) and 20 µL of 2.2 mmol/L 4-methylumbelliferone (4-MU)-a-D-glucopyranoside (4MUG, Sigma) in 0.2 mol/L Na-acetate buffer (pH 4.0) with 0.02% (wt / vol) sodium azide. The reaction mixtures were incubated for 1 hour at 37°C, and the reaction was stopped by the addition of 200 µL 0.5 mol/L Na_2CO_3/$NaHCO_3$ (pH 10.7). The fluorescence of released 4MU was measured with a fluorimeter (Thermo Electron corporation). The protein concentration of the samples was determined using Pierce™ BCA Protein Assay Kit (Thermo Scientific™) following the instructions provided by manufacturer and previously described [27]. The GAA enzyme activity assays and the protein assays of study 1 (see Results) were performed at time of this study. The GAA enzyme activities per mg protein of study 2 were obtained using exactly the same routine procedures, but were derived from the patients' medical records.

RNA sample collection for RNA sequencing and qPCR

Skin fibroblasts were cultured according to routine procedures in DMEM medium supplemented with 10% fetal calf serum and 1% penicillin/streptomycin. Fully confluent tissue culture flasks were split and harvested 3 days later for RNA isolation using the RNAeasy miniprep kit with on-column DNase treatment according to the manufacturer's recommendations (Qiagen). For gene expression analysis, 800 ng of RNA was used for generation of cDNA using the iScript cDNA synthesis kit (Biorad) and real time qPCR. The cDNA solution was diluted 10 times before use. To determine the relative concentration of each sample, 4 µl of each cDNA sample (10 times diluted in H_2O) 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 using Q-actin.

RNA isolation, cDNA library construction, Illumina deep sequencing and Data processing of RNA-Seq experiments

Total RNA was extracted using the RNAeasy miniprep kit following the manufacturer's protocol (Qiagen). A total of ≥ 100 ng RNA per sample was sent for library preparation using the TruSeq RNA sample preparation kit and sequenced according to the Illumina TruSeq v3 protocol on the HiSeq2000 with a single read 43 bp and 7bp index, and

mapped against the requested reference using Tophat (version 2.0.10). We called gene expression values using Cufflinks (version 2.1.1). More than 10 million aligning reads per sample were generated for each sample. The gene/transcript annotation hg19 transcriptome was used. Raw data in the fastq format were processed. We applied three methods for differential gene expression analysis: T-test per gene on cufflinks FPKM (fragments per kilobase of transcript per million mapped reads) values,Limma on read counts, and EdgeR with a generalized linear model. RNA sequencing was performed by the Erasmus Center for Biomics at Erasmus MC (www.biomics.nl, Rotterdam, The Netherlands).

Differential Analysis: Pompe patients Child vs Adult and Controls

Differential expression analysis was performed using edgeR (R package) with a generalized linear mode and we called gene expression values using Cufflinks (version 2.1.1). Hierarchical cluster analysis and multidimensional scaling (MDS) plots were used to identify DEGs. DEGs were filtered with false discovery rate (FDR) <0.05, fold change (FC) >1.5 or <-1.5 in each pairwise comparison.

GO analysis

Gene ontology (GO) analysis was performed to determine the biological implications of the expression of unique genes in significant or representative profiles of genes that were differentially expressed using The PANTHER database (Protein Analysis Through Evolutionary Relationships) at http://pantherdb.org). The Bonferroni correction for multiple testing was applied to identify significant GO categories, and p-values of <0.05 were considered significant. The categories are described as GO molecular function (GOTERM_MF), GO biological process (GOTERM_BP), and GO cellular component (GOTERM_CC), GO panther pathways, and Reactome pathways. All enriched biological processes were ranked from top to bottom according to the p-values for each GO term.

Pathway analysis was used to determine the significant pathway(s) of the DEGs according to the PANTHER database and Ingenuity software (IPA; Qiagen, Valencia, CA; http://www.ingenuity.com). The analysis was determined by adjusting p-values for multiple testing, including the Bonferroni and Benjamini-Hochberg correction methods (P<0.05), in order to identify the most significant pathways and causal relationships associated with the experimental data.

Statistical Analysis

We called gene expression values using Cufflinks (version 2.1.1). The data were analyzed using IBM SPSS (version 25), and GraphPad prism (version 8.0.0). The category percentage and the basic statistics value were calculated for each group and/or study. Linear regression analysis was used to modeling the relationship between gene expression level and age of onset/biopsy. We performed the one-way and 2-way ANOVA: Turkey's multiple comparisons test for testing groups to see if there is a difference among and between them. Differences were regarded to be significant in all tests at a significance level of "$P < 0.05$".

The pathway and GO terms associated with the obtained DEGs were tested for significance using the Bonferroni correction and Benjamini-Hochberg correction for multiple testing ($P < 0.05$) and were subjected to gene ontology analysis.

RESULTS

Figure 1 provides an overview of the experimental setting.

Table 2. Characteristics of patients reported in studies 1 and 2

	Childhood Pompe	Adult Pompe	all patients
	Total	Total	Total
Patients with Pompe disease	17	22	39
Gender: males (%)[a]	13 (76.5)	8 (36.4)	21 (54)
Median age at first symptoms (y)[b]	2 (0.5-13)	41.5 (20-62)	22 (0.5-62)
Median diagnosis age (y)[b]	2.5 (0-14.1)	46.5 (26-72)	34 (0-72)
Median age at time of biopsy (y)[b]	9.8 (0.35-16.3)	55.8 (39-72.9)	41 (0.35-72.9)
Mobility (%)[a]			
Ambulant	16 (94)	14 (63.6)	30 (76.9)
Wheelchair dependent		2 (9.1)	2 (5.1)
Partially wheelchair dependent	1(5.9)	5 (22.7)	6 (15.4)
Walking aid		1 (4.5)	1 (2.6)
Ventilation dependent (%)[a]	1 (5.9)	6 (27.3)	7 (17.9)
GAA activity[b,c]	9.1 (6.2-29)	12 (5.2-18.7)	11.5 (5.2-29)
Patients and genotypes (%)[a]			
c.-32-13T>G/c.525del p.(Glu176Argfs*45)	6 (35)	12 (55)	18 (46)
c.-32-13T>G/ other 'null' *GAA* variant	9 (53)	8 (36)	17 (44)
c.-32-13T>G/ 'non-null', c.2135T>C p.(Leu712Pro)	2 (12)		2 (5)
c.-32-13T>G/ 'non null', c.1076-22T>G p.?		2 (9)	2 (5)

The figures are a percentages or b medians with ranges. c The control range for GAA activity in fibroblast is: 40-180 nmol MU/mg.h.

Clinical features and disease severity

Table 2 provides the characteristics of the patients in *cohort 1* (onset of symptoms under the age of 16 years) and *2* (onset of symptoms over the age of 35 years). The majority (76.5%) of affected children i n cohort 1 was male. The median age at first symptoms was 2 years (0.5-13), and the median age at time of biopsy was 9.8 years (0.35-16.3). Among the adult patients in cohort 2, 36.4% are males. The median age at first symptoms was 41.5 years (20-62), and the median age at time of biopsy was 55.8 years (38.9-72.9). **Table 2** also lists the number of *GAA* genotypes of all patients. With regard to the latter, all patients carry the common c.-32-T>G variant that combines in 6 (35%) of the childhood cases and in 12 (55%) of the adult cases with the c.525del allele and in all other cases of childhood (53%) and adult onset (36%) with an equally deleterious "null" other disease-associated *GAA* variant, except for 4 patients (2 child (12%) and 2 adult (9%) that contain a 'non-null' *GAA* variant.

Figures 2B and **2C** present the age at first symptoms, the age at diagnosis, the age at time of biopsy, and the disease duration for each individual patient, and **Suppl.**

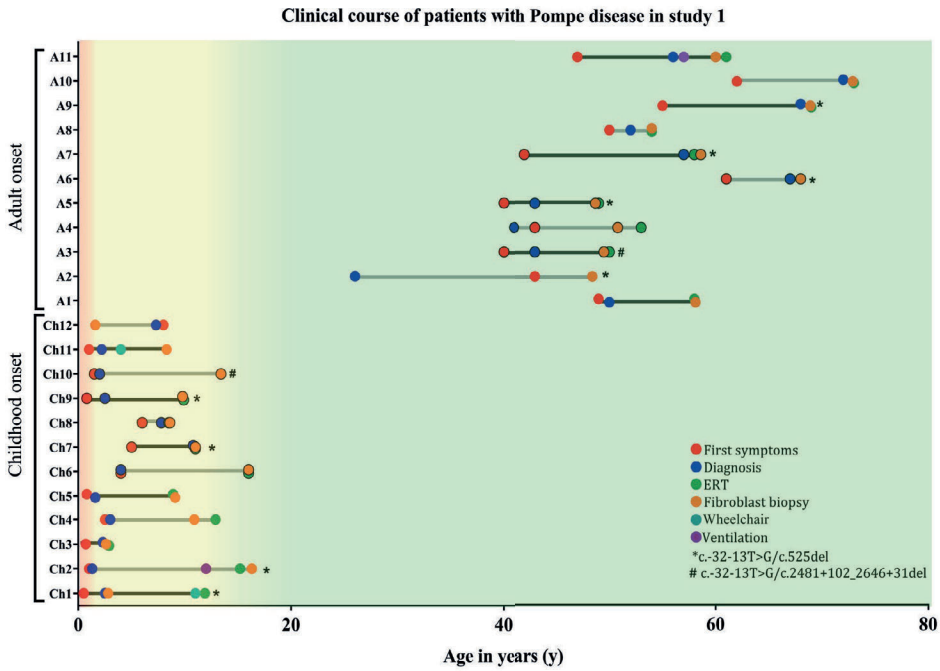


Figure 2B. Clinical course of patients in study 1. C1-C12 are patients with childhood onset Pompe patients and A1-A11 are adult onset Pompe patients. Dots represent an event during the course of the disease. Patients marked with * or # share the same GAA genotype. Patient C11 is partly wheelchair dependent, but can walk for 200 m. Patient Ch1 was wheelchair dependent but acquired walking ability during ERT.

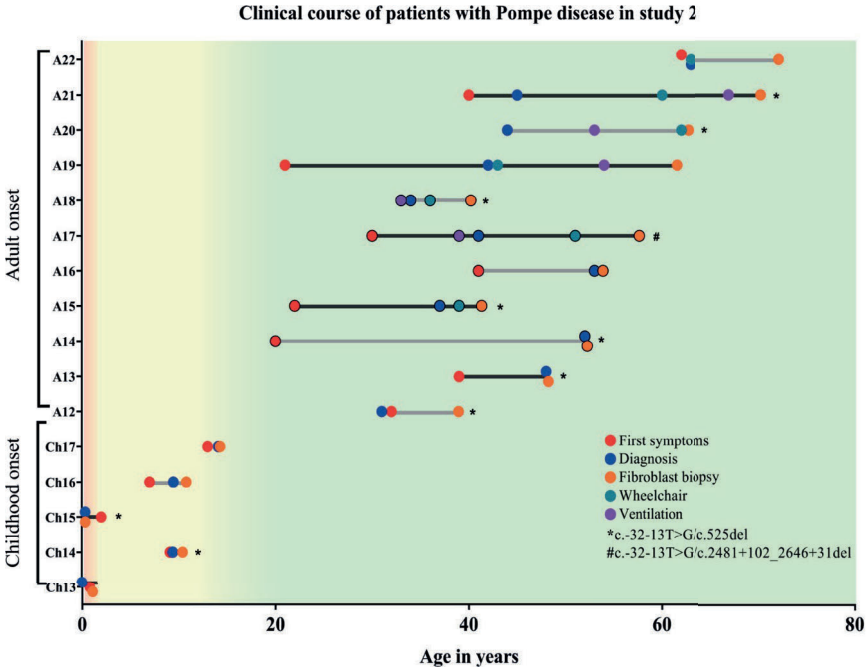


Figure 2C. Clinical course of patients in study 2. Patients Ch13-Ch17 had childhood onset and A12-A22 during adulthood. Dots represent an event during the course of the disease. Patients marked with * or # share the same GAA genotype. Patients A15, A18- A20 and A22 are partially wheelchair dependent.

Figure 12 provides the ages at biopsy of controls. If known, the figures also include the age at which the patients became wheelchair or ventilator dependent and/or started to receive ERT. Patients sharing exactly the same *GAA* genotype are marked. The youngest of all patients developed symptoms in the first year of life, while the oldest developed symptoms at the age of 62 years. The age at diagnosis varies between 0 and 72 years (median 34 years). The age at time of skin biopsy varies between 0.35 and 72.9 years (median 41 years). **Table 3** shows the median age at first symptoms of the childhood and adulthood onset patients with Pompe disease per study.

GAA enzyme activity and glycogen accumulation

The GAA enzyme activity and glycogen accumulation assays were performed in fibroblasts of Pompe patients and healthy controls to investigate potential associations between residual GAA activity, glycogen accumulation, and clinical phenotype. **Figure 2D** showed no statistically significant differences between the GAA enzyme activity in fibroblasts from childhood vs adult patients that could explain differences in phenotype ($P>0.05$). For healthy controls a significant difference was observed between GAA activities

Table 3. Minimum and maximum ages of onset among all patients in studies 1-3

Study number	Study 1		Study 2		Study 3	
Phenotype	Childhood	Adult	Childhood	Adult	Childhood	Adult
Patients (n)	12	11	5	11	6	6
Median age at first symptoms (y)	1.25	47	7	33	8.75	39.5
Minimum age at first symptoms (y)	0.5	40	0.75	20	1	36
Maximum at first symptoms (y)	8	62	13	62	15	57
Range (y)	7.5	22	12.25	42	14	21
Mean age at first symptoms (y)	2.7	48.4	6.4	34.9	8.3	43.2
Std. Deviation	2.5	8.0	5.1	12.3	6.4	8.1

The median, mean and std. deviation of age of symptoms onset is indicated per phenotype and study number

of children and adults (P < 0.001). As expected, patients showed GAA deficiency compared to controls (P<0.001), with the following values: childhood patients: 9.1 [6.2-29], adult patients: 12 [5.2-18.7], childhood control: 98.95 [79.1-168], adult control: 70.3 [30.2-116] (control range of GAA activity in fibroblasts: 40-180 nmol MU/mg.h) [22].

In order to measure the lysosomal glycogen content of the fibroblast, a culture

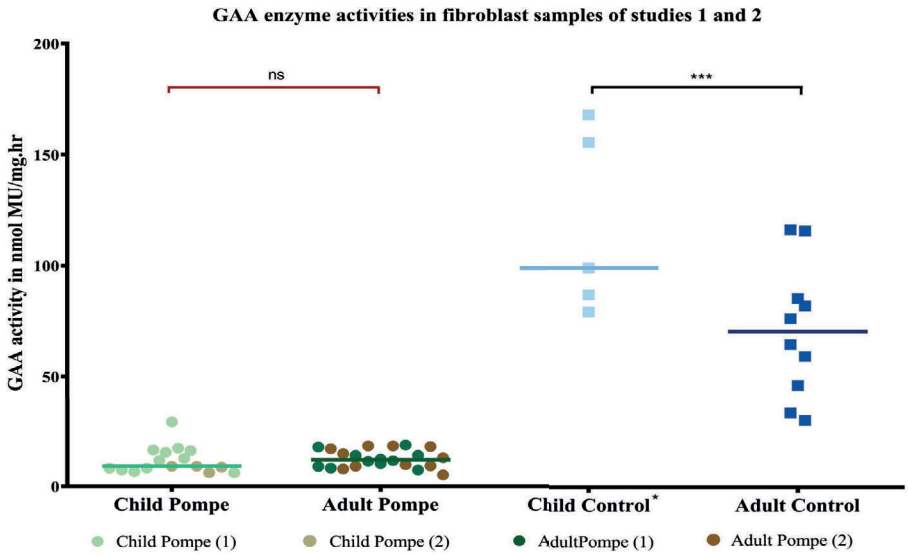


Figure 2D. GAA enzyme activities in fibroblast samples of studies 1 and 2. The horizontal line presents the median value. The study number is given within brackets. The difference between the childhood and adult onset groups is non-significant (ns; (P>0.05), but the difference between the patients (childhood plus adult onset) and controls were significantly different P<0.05 (**** <0.0001, *** indicate p 0.0001). Normal range of GAA activity in fibroblasts is 40-180 nmol MU/mg.h (4MUG as substrate was used). *Foreskin fibroblasts are not included in this figure. Enzyme activities for controls in study 2 were not performed. The one-way ANOVA test was used for the analysis.

condition without glucose had to be applied for 3-5 days to deplete cytoplasmic glycogen. Before embarking on this assay, we investigated whether changing the culture conditions had any effect on the GAA activity (**Suppl. Fig S1**). There were no significant differences between the GAA activity of fibroblasts from affected children and adults (cohorts 1 and 2), neither between childhood controls and adulthood controls cultured in media with or without glucose for either 3 or 5 days' time (P>0.05). We measured the glycogen content of the fibroblasts from children with Pompe disease and unaffected children and adults on day

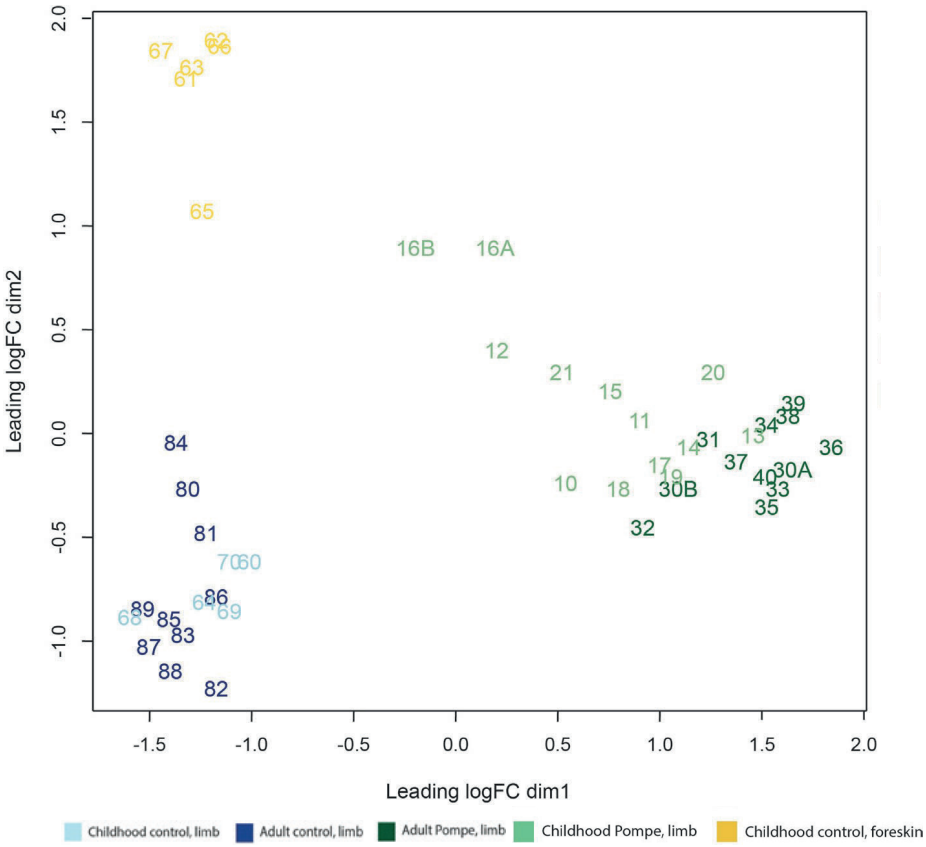


Figure 3A. Multidimensional Scaling Plot (2-D MDS) of the RNA-seq data for the patients and controls of study 1. The MDS plot shows different gene expression profiles between Pompe patients and healthy controls. The profiles of adult onset patients (dark green, n= 11: 30A, 30B, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40) differ from those of childhood onset patients (light green, n=12: 10, 11, 12, 13, 14, 15 ,16A, 16B, 17, 18, 19, 20, 21). In contrast, unaffected children (light blue, n=5: 60, 64, 68, 69, 70) and adult controls (dark blue, n=10: 80, 81,82, 83, 84, 85, 86, 87, 88, 89) controls show similar expression profiles except for samples from 6 unaffected children (gold: 61, 62, 63, 65, 66, 67) that revealed a completely different profile. These w ere fibroblasts derived from foreskin. The stress for this plot is 13.5%. Of note, the patient numbers shown in this figure do not correspond with those in the tables (see Suppl. Table S5. for comparison).

3 after culture in glucose free media and did not find statistically significant differences between the four sample types (**Suppl. Fig. S2**). The cells did not store glycogen, as opposed to fibroblasts from a single patient with classic infantile Pompe disease, which was included in the experiment as positive control and which did store glycogen with a 5-10 fold increase. Supplementary **Figure S2** summarizes the results per individual sample and shows the median glycogen content per sample type in µg glycogen per mg protein (childhood patient: 37 [8.1-143], adult patient: 25 [9.5-60], childhood control: 58.4 [15-86], adult control: 56.5 [11-88]).

Searching for DEGs: RNA-Seq analysis

Differentially expressed genes in childhood vs adult onset Pompe disease compared to childhood versus adult controls

RNA-Seq analyses were conducted to generate transcriptome profiles from fibroblasts of Pompe patients and healthy controls in order to identify genetic denominators that could possibly explain the difference between childhood and adult onset Pompe disease. We conducted pairwise comparison of materials from adult and childhood Pompe patients (study 1A) and adult and childhood controls (study 1B).

The Multidimensional scaling (2-D MDS) plot (**Fig. 3A**) revealed distinct expression profiles of Pompe patients and healthy controls included in the studies 1A-B. Adult Pompe patients (dark green) had different gene expression levels than childhood Pompe patients (light green), whereas adult (dark blue) compared to childhood (light blue) control samples did not. Quite unexpectedly, six childhood control samples (yellow) had completely different gene expression profiles than all other samples from either patients or controls. It then turned out that the deviant samples were from foreskin, while all others were derived from skin biopsies of the upper leg. Without clear understanding, we were confronted with the fact that foreskin may have other characteristics than other types of skin fibroblasts. This was reason to exclude the results obtained with foreskin fibroblasts were excluded from all further analyses. The DEGs were filtered based on level fold change (FC) >1.5 or < -1.5, and false discovery rate (FDR) <0.05, in each pairwise comparison. The outcome of hierarchical cluster analysis of study 1A plus 1B (foreskin excluded) is depicted in Supplementary **Figure S3A** and showed a group-wise clustering of patients in the left branches of the tree and controls in the right branches. Healthy children and adults were randomly represented in the various sub-branches. Within the patient branches were sub-branches with adults or children only, but there was no complete separation of phenotypes. Similar results were obtained in study 2 (Supp. **Fig. S3B**) albeit that the branching pattern was not precisely the same as in study 1, and the separation between adult and childhood cases of Pompe disease even less evident, which suggested a very large variation in study 2 compared to study 1.

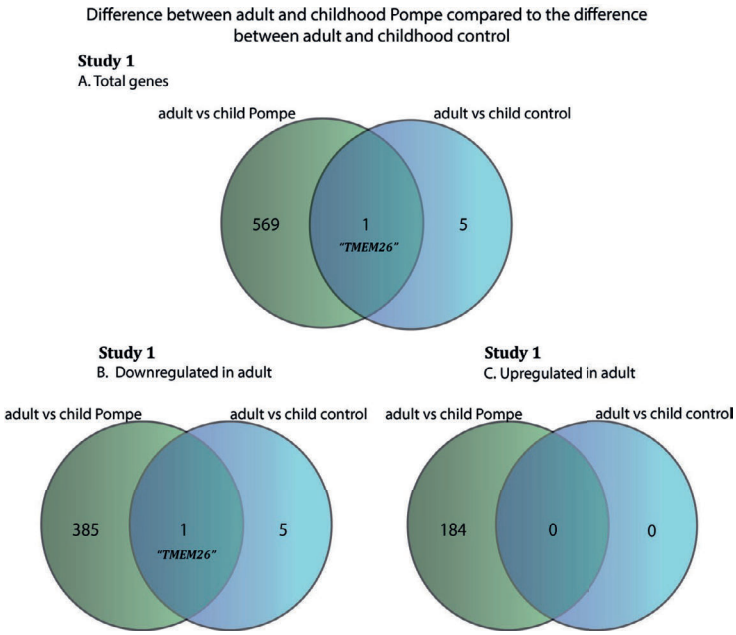


Figure 3B. Venn diagram showing the numerical difference between genes expressed in adult versus childhood onset Pompe disease (green) compared to the difference between adult versus childhood control (blue) of study 1. (A) A total of 569 genes are differentially expressed in childhood versus adult onset Pompe disease (corrected for controls), whereas only 5 in the comparison of unaffected children versus unaffected adults. (B) Of the 569, 385 are downregulated in adult onset Pompe disease, and (C) 184 upregulated. Foreskin fibroblasts were excluded. The DEGs were filtered using a cut off value of < -1.5 - >1.5 fold change (FC) (LogFC:+/- 0.584) and false discovery rate (FDR) of <0.05

Pairwise comparisons of adult vs childhood cases of Pompe disease (study 1A) compared to adult vs childhood healthy controls (study 1B) revealed that 569 genes were significantly differentially expressed between the childhood compared with the adult patient cohorts and 6 between the childhood and adult controls (**Fig. 3B[A]**). The Venn diagram shows that only 1 gene, *TMEM26*, was shared between patients and controls. Of the 569 DEG's between fibroblasts of affected children compared to affected adults, 385 were down regulated (**Fig. 3B[B]** and 184 upregulated in adults (**Fig. 3B-[C]**).

All DEG's that were identified in study 1 were subjected to further analysis using the Panther database and Ingenuity software. The pathway and GO terms associated with the obtained DEGs were subjected to gene ontology analysis. The GO enrichment of the GO terms suggested that the transcriptional levels of genes related to molecular function (such as aminoacyl-tRNA ligase activity catalyzing the

formation of aminoacyl-tRNA and related compounds), and biological processes (such as the type I interferon signaling pathway; cellular response to type I interferon, response to type I interferon, response to virus, defense response to virus) were altered (Supp. **Figs. S4A-B**). Similar results were obtained when analyzing the data in the context of GO pathway category. In fact, the results of the GO enriched pathway analysis supported our findings obtained with ingenuity pathway analysis (Supp. **Figs. S5A-B**). Pathway terms that came out as being significant included several major processes such as i) amino acid metabolism and biosynthesis (tRNA charging and the super- pathway of serine and glycine biosynthesis I), ii) cell and organ morphology (axonal guidance signaling, nitric oxide signaling in the cardiovascular system), iii) immune system response/processes (role of pattern recognition receptors in recognition of bacteria and viruses), and iv) hormone metabolism (thyronamine and iodothyronamine metabolism, thyroid hormone metabolism I via de-iodination).

qRT-PCR analysis was used to validate the quality of the RNA-Seq data. For this purpose, we selected 10 downregulated and 10 upregulated genes that were identified in study 1A-B. The relative gene expression levels of selected genes measured by qRT- PCR were in good agreement with the RNA-sequencing data (Supp. **Figs 6A-B**). The strong correlation between the RNA-Seq and qRT-PCR data indicates the reliability of our transcriptomic profiling data. Therefore, the culture condition dependency and reproducibility were analyzed by re-growing the fibroblasts of patients included in study 1A and the expression levels of down and upregulated genes were analyzed by qRT-PCR. The majority of those genes are in line with our previous results by qRT-PCR.

The expression of some genes seemed culture condition dependent (Supp. **Figs 6C-D**). The same approach was taken to evaluate the outcome of study 2. The Venn diagramof this experiment revealed only 7 genes differentially expressed between adult onsetand childhood onset Pompe patients, and 80 in childhood versus adult unaffected controls (**Fig. 4A[A]**). Of those 7 genes, controls and patients share 3, and all three are downregulated in adult compared to childhood Pompe disease (**Fig. 4A[B]**). Four genes are upregulated (**Fig. 4A[C]**). Although, the DEGs in study 2 are not in line with our previous results (study 1), we performed RT-qPCR of genes selected for study 1. Conversely, the majority of these genes showed the same trend of down and upregulation as in study 1. However, they were not statistically significant (data not shown).

Differentially expressed genes in childhood vs adult Pompe disease combining study 1 and 2

Venn diagrams were also made to identify the DEGs in adult vs childhood onset Pompe disease, corrected for childhood vs adult controls, identified in the combined studies 1 + 2. (**Fig. Supp. 10A**). By combining the results of all these studies only 3 DEGs came out

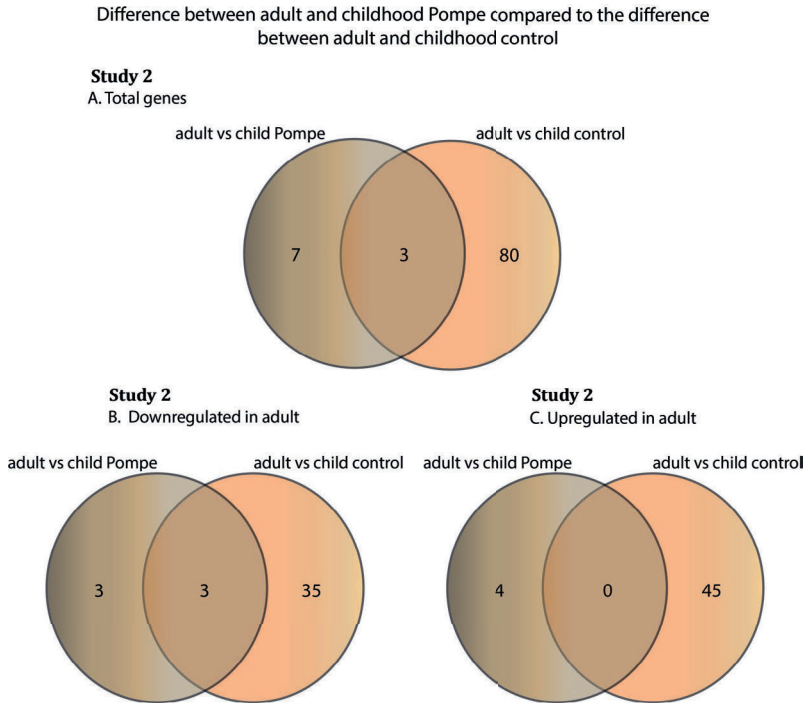


Figure 4A. Venn diagram showing the numerical difference between genes expressed in adult versus childhood onset Pompe disease (brown) compared to the difference between adult versus and childhood control (orange) of study 1. (A) A total of 7 genes are differentially expressed in childhood versus adult onset Pompe disease (corrected for controls), whereas 80 in the comparison of unaffected children versus unaffected adults. (B) Of the 7 genes, 3 are downregulated genes in adult onset disease and (C) 4 upregulated. Foreskin fibroblasts were excluded. The DEGs were filtered using a cut off value of 1 .5 fold change (FC) and a false discovery rate (FDR) of <0.05.

distinguishing between childhood and adult onset Pompe disease: one downregulated in adult onset compared to childhood onset disease and two upregulated (**Fig. 4B**) The nik related kinase (*NRK*) gene showed over 14-fold diminished expression in 13 of the 22 samples from Pompe patients with adult onset disease relative to childhood Pompe (slightly up regulated in 4 samples and slightly down regulated in 4 others. (**Fig. 5A**). Genes coding for a mitochondrial enzyme *MAOA* and a secreted, homodimeric glycoprotein: *STC1*, exhibited over two-fold increased expression in adult Pompe compared to the childhood Pompe patients (**Fig. 5B**). Interestingly, those genes have been implicated in various biologic processes including bone and muscle development, cellular calcium/phosphate homeostasis, cellular metabolism, and cellular biogenic amine metabolic process. (Supp. **Tables S3A-B**).

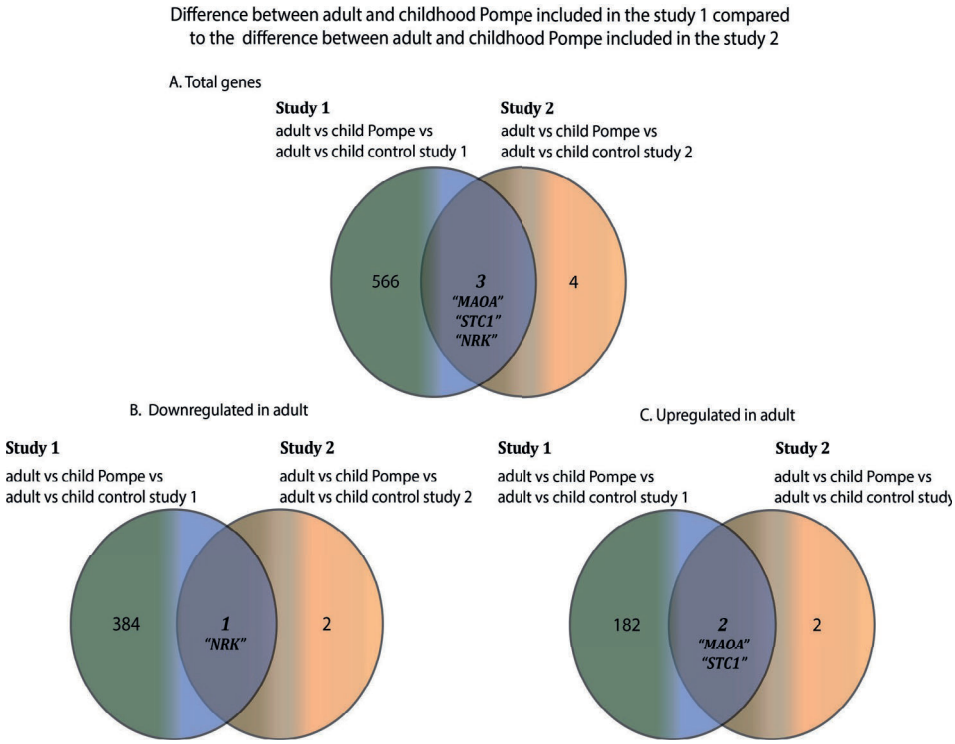


Figure 4B. Venn diagram showing the numerical difference between genes expressed in adult versus childhood onset Pompe disease combining study 1 and 2. The venn diagram is corrected for the differences between unaffected children and adults, and shows that (A) A total of 3 genes are differentially expressed in adult versus childhood onset Pompe disease, of which (B) one is downregulated in adult onset disease and (C) two upregulated.

Differentially expressed genes in adult Pompe patients vs adult Controls: Study 1

Apart from determining the difference between early and late onset disease among Pompe patients we also investigated whether particular genes would be differentially expressed in patients compared to unaffected controls.

In study 1 (A-B), pairwise comparisons were made between expression profiles of adult Pompe disease patients and adult controls. It turned out that 6607 genes were significantly differentially expressed. Of these 6607 genes, 3292 were significantly downregulated and 3315 significantly upregulated in adult Pompe patients compared to adult healthy controls (Supp. **Table S4A**).

All those DEGs were subjected to gene ontology and pathway analysis as described above. It appeared that over 600 GO terms were enriched among all DEGs and 390 GO terms among the downregulated DEGs (MF= 30, BP= 265 CC= 95). Enrichment of these GO terms suggests that genes involved in amino acid synthesis, RNA translation,

and protein binding play a role. These results were supported by GO pathway analysis terms (Supp. **Fig. S4A**). Furthermore, we found by pathway analysis (IPA) that most altered pathways are associated with protein synthesis, apoptosis, and cellular stress such as EIF2 signaling, mTOR signaling, regulation of eIF4 and p70S6K signaling and mitochondrial dysfunction (Supp. **Fig. S5A**). The mTOR signaling pathway was the principal core and key pathway involved. It plays an important role in protein synthesis and apoptosis, but particularly in regulation of the autophagy machinery. Interestingly, dysfunctional autophagy contributes greatly to the skeletal muscle pathology in Pompe

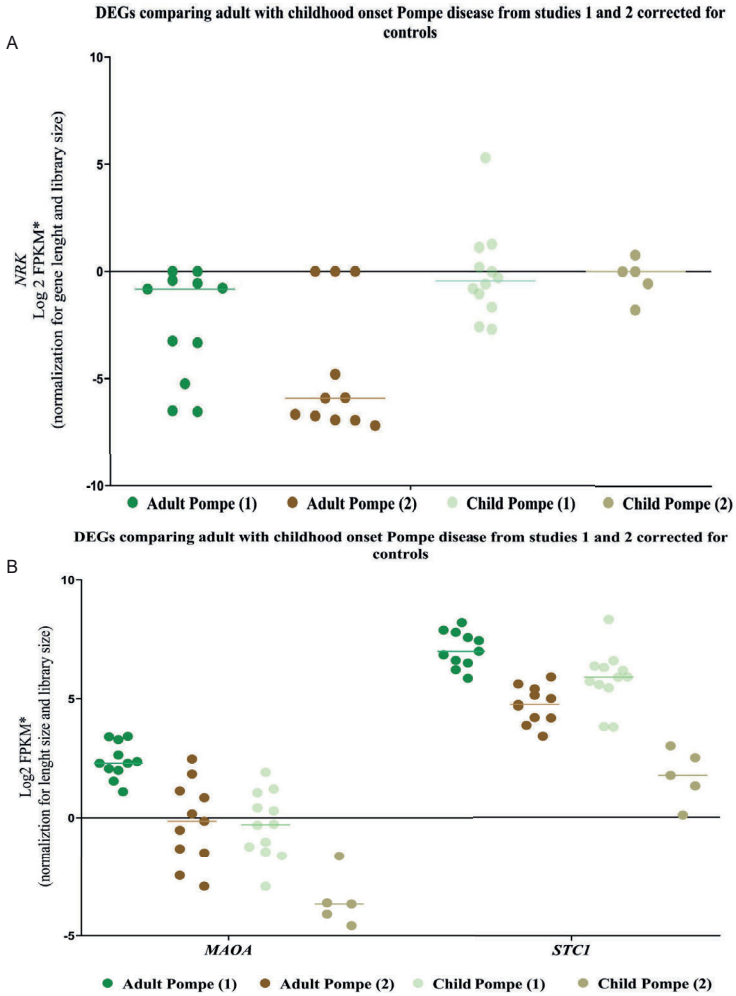


Figure 5. DEGs comparing adult with childhood onset Pompe disease from studies 1 and 2 (corrected for controls). The horizontal line represents the median value; (1/2)= study number. *Fragments per kilobase of exon per millon reads mapped. Fig. 5A. Downregulated DEGs in adult versus childhood onset disease. Fig. 5B. Same as Fig. 5A but upregulated genes.

disease [28]. In fact, autophagy has a vital role in maintaining the amino acid pool. Among the upregulated DEGs, 265 GO terms were enriched (MF= 26, BP= 190, CC= 49). Immune response genes and related genes were significantly over-represented in adult vs childhood onset Pompe disease relative to adult control (Supp. **Fig. S4B**), which agrees with the outcome of the most significant pathway analysis by IPA (Supp. **Fig. S5B**).

Differentially expressed genes in adult Pompe patients vs adult Controls: study 2

In study 2, pairwise comparisons of adulthood Pompe versus adulthood control showed that 53 genes were significantly differentially expressed. Among these, 28 genes were significantly downregulated and 25 genes were significantly upregulated in adult Pompe patients compared to the adult healthy controls (Supp. **Table S4A**).

Differentially expressed genes in childhood Pompe patients vs childhood Controls: study 1

Using the same filtering parameters and considering FDR significance of only less than or equal to 0.05, we identified 4426 DEGs in the cohort of childhood Pompe patients, of which 2002 were downregulated and 2424 were upregulated compared to the childhood controls (Supp. **Table S4B**). Those DGEs were subjected to the analysis of Gene Ontology (GO) terms and pathway analysis. More GO terms were enriched by genes having lower expression level as well as their relationships in the cell cycle process, and cellular division in childhood Pompe relative to the childhood control (Supp. **Fig. S4A**), which highly correlated with the pathways potentially affected (Supp. **Figs. S4A** and **S5A**). The poor expression of genes related to cell division might be caused by a cascade of events induced by the cell culture reaching confluence. The analysis including upregulated genes by Panther pathway term and Ingenuity pathway provided clear indication of over-expressed genes in childhood Pompe relative to childhood control play a role in immune response (Supp. **Figs. S4B** and **S5B**).

To identify the DEGs specific for Pompe disease, we employed Venn diagrams comparing both adult Pompe vs adult controls as well as childhood Pompe disease vs childhood control. They illustrate that 3857 genes were expressed at a significantly different level (Supp. **Fig. S5A**). Among these, 1748 genes were significantly downregulated (Supp. **Fig. S5B**) and 2107 genes were significantly upregulated in Pompe disease (Supp. **Fig. S7**). In addition, this Venn diagram revealed a total of 2750 genes, which were only expressed in adult Pompe patients and 569 genes only in childhood patients (Supp. **Fig. S7**, top). The number of representative genes in the adult and childhood Pompe patients for the up or downregulated DEGs are shown in (Supp. **Figs. S7[-B-C]**).

Differentially expressed genes in childhood Pompe patients vs childhood Controls: study 2

We identified 89 DEGs in childhood Pompe, of which 49 were downregulated and 40 were upregulated compared to childhood controls (Supp. **Table S4B**).

We performed the same approach in study 2 as in study 1 for identifying the DEGs in Pompe patients by comparing the difference between adult Pompe and adult controls compared to the difference between childhood Pompe and childhood control. As shown in Supplementary **Figure S8[A]**, only seven genes were differentially expressed. Among these, five genes were significantly downregulated (Supp. **Fig. S8[B]**) and two genes were significantly upregulated in Pompe patients (Supp. **Fig. S8[C]**). In addition, this Venn diagram revealed a total of 46 genes that were differentially expressed in adult Pompe patients and 82 genes in childhood patients (Supp. **Figs. S8[B-C]**).

With regard to the biological significance of DEGs identified in study 2, the GO and IPA analysis did not yield any conclusive results as no processes or pathways were foundto be significantly enriched, which can be explained by the much lower number of DEGs identified in study 2 compared to study 1.

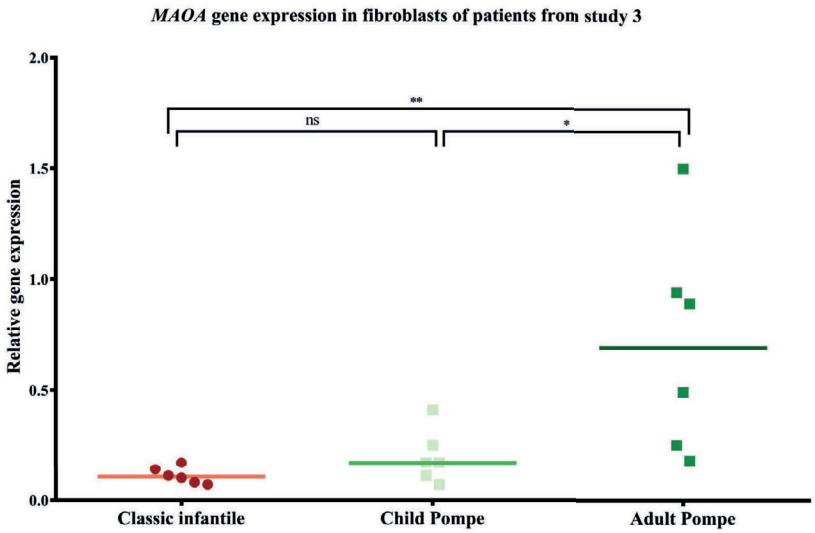


Figure 6A. *MAOA* gene expression in fibroblasts of patients from study 3. The gene expression level was measured by RT-qPCR in a third series of samples to validate the results obtained in study 1-2. The horizontal line presents the median value. ns, not significant (P>0.05), * significant (P<0.05); ** indicate P<0.01. The one- way ANOVA test was used for analysis.

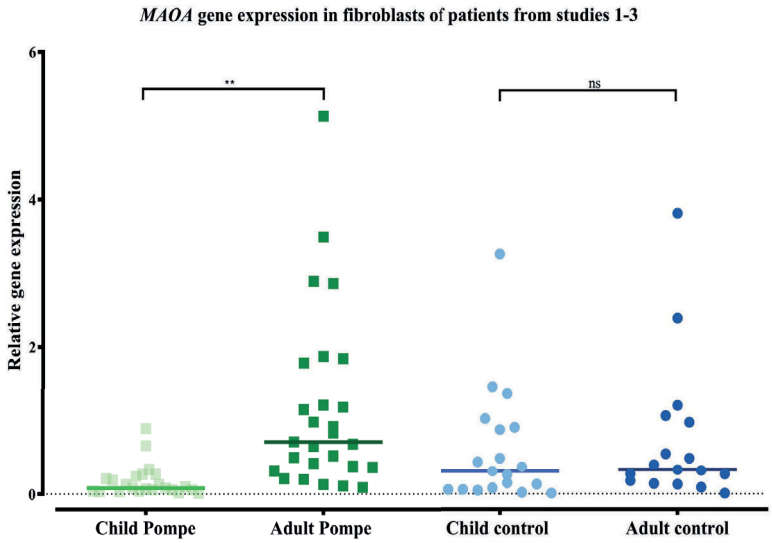


Figure 6B. *MAOA* gene expression in fibroblasts of patients from studies 1-3. The gene expression level was measured by RT-qPCR combining studies 1-3. The horizontal line presents the median value. ns, not significant (P>0.05). ** significant (P<0.01). The one-way ANOVA test was used for the analysis.

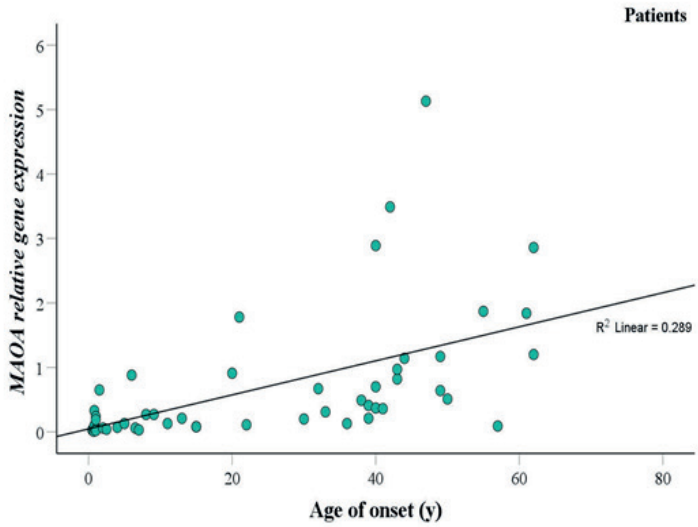


Figure 6C. *MAOA* gene expression by age of onset in studies 1-3. The expression level was measured by RT- qPCR. Linear regression analysis using SPSS was used.

Differentially expressed genes in Pompe disease compared to unaffected controls combining the results of study 1 + 2.

Apart from analyzing the differences in gene expression profiles between childhood and adult Pompe patients, we also compared the profiles of fibroblasts from affected versus unaffected individuals. The results are depicted in the Venn diagrams of Supplementary **Figures S9** (adults), **S10** (children) and **S11** (adults and children combined). In the comparison adult Pompe/ adult control 32 genes came out as differentially expressed (Supp. **Fig. S9[A]**), but only 17 genes were identified that had the same behavior (up or down) in both studies. Nine of these 17 genes were downregulated (Supp. **Fig. S9[B]**) and 8 upregulated in adult Pompe patients compared to adult controls (Supp. **Fig. S9[C]**). In the comparison childhood Pompe / childhood control 31 genes were identified (Supp. **Fig. S10[A]**), whereby 9 showing the same behavior came out in both studies. Six of these were downregulated (Supp. **Fig. S10[B]**) and 3 upregulated in childhood Pompe patients compared to childhood controls (Supp. **Fig. S10[C]**).

Finally, the DEGs identified in the childhood/childhood and adult/adult comparisons were analyzed in combination (Supp. **Fig. S11**) and this resulted in the identification of only one single gene that was downregulated in all Pompe disease patients compared to all controls. It turned out to be the *GAA* gene, and its lower expression in fibroblasts of all patients is explained by the disease-associated c.-32-13T>G variant that all patients in our study share and leads to 80-90% diminished GAA mRNA expression. The finding is comforting as it contributes to the validity of our approach and supports the additional findings.

Validation of the 3 common DEGs between adulthood and childhood patients in study 1 and 2

Again, to validate the accuracy of the RNA-Seq analysis, qRT-PCR was performed in study 3, which included fibroblasts of a new set of six childhood and six adult patients with Pompe disease. This new test confirmed the statistically significant higher expression level of *MAOA* in affected adults compared to affected children (**Fig. 6A**), also when applied to all patients and controls included in the studies 1-3 ($P<0.01$) (**Fig. 6B**). Figure 6C is suggestive for an upward trend of the *MAOA* gene expression level with age of onset. In contrast, in healthy controls age-related changes in gene expression of *MAOA* were not observed (Suppl. **Figure S12A**). Similar results were obtained for *STC1* gene expression in study 3 (**Fig. 7A**) as well as in studies 1-3 (**Figs. 7B-C**) albeit that 4 samples in particular had a strong effect on the upward trend. Supplementary Figure 12B does not show age-related gene expression changes of *STC1* in controls. Analysis of *NRK* expression by RT-qPCR revealed that this gene was not expressed in most individuals but that it was highly expressed in selected individuals, without preference for one group. Therefore, we did not

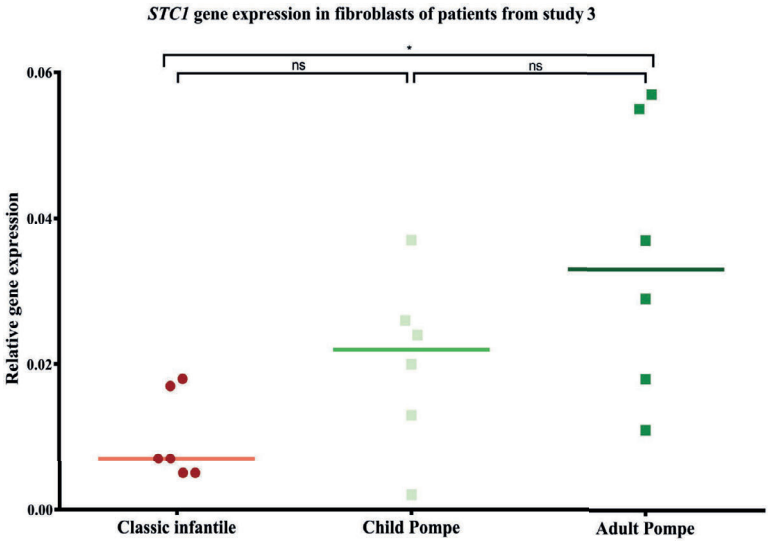


Figure 7A. *STC1* gene expression in fibroblasts of patients from study 3. The gene expression level was measured by RT-qPCR to validate the results obtained in study 1-2. The horizontal line presents the median value. The one-way ANOVA test was used for the analysis. ns, not significant (P>0.05), significant *P<0.05.

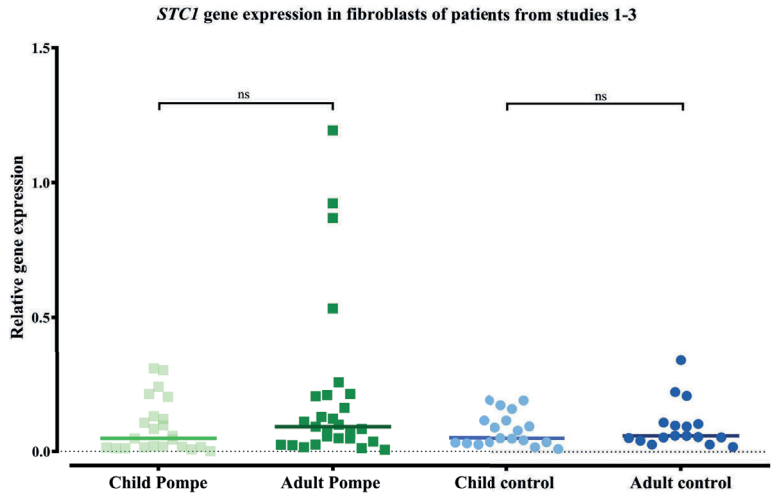


Figure 7B. *STC1* gene expression in fibroblasts of patients from studies 1-3. The gene expression level was studied by RT-qPCR. The horizontal line presents the median value. The one-way ANOVA test was used for the analysis. ns, not significant (P>0.05).

further pursue *NRK* expression.

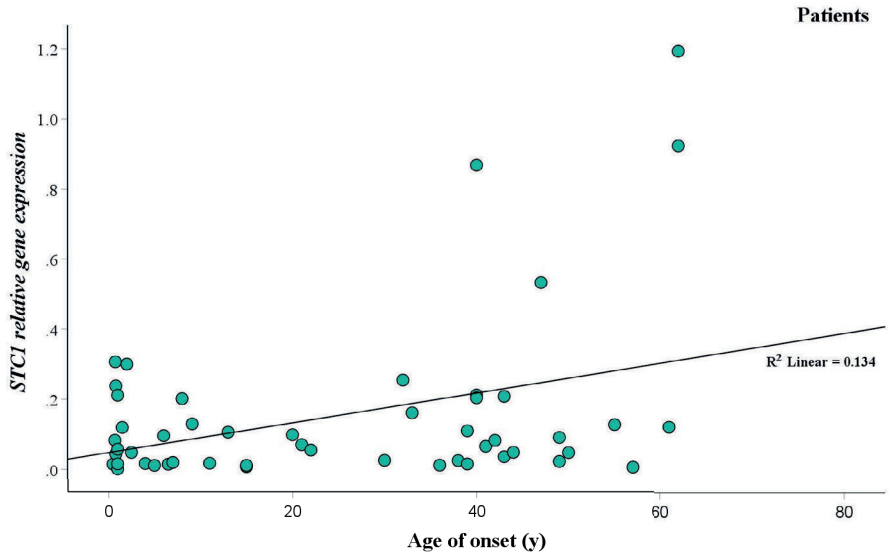


Figure 7C. *STC1* gene expression by age of onset in fibroblast of patients from studies 1-3. The expression level was measured by RT-qPCR. Linear regression analysis was performed using SPSS.

DISCUSSION

The transcriptomes of fibroblast cultures from patients with childhood and adult Pompe disease were investigated to gain insights in the DEGs pattern of various Pompe disease phenotypes. To the best of our knowledge this is a unique study for several reasons. Firstly, there is hardly any literature on markers and/or factors that can be used for monitoring disease progression in Pompe disease. Secondly, transcriptome analysis using Pompe patient fibroblasts has never been thoroughly described, although one paper has analyzed gene expression by using microarray, but this was performed on muscle biopsies of classic infantile Pompe patients that do not have the c.-32-13T>G variant that leads to residual GAA activity like the patients in our study [29]. Thirdly, cultured skin fibroblasts were specifically chosen for this purpose as they are by far the best sample source for accurate measurement of the residual GAA enzyme activity level. In fact, our results of the lysosomal glycogen content confirmed a previous finding that cultured fibroblasts from patients other than those with classic infantile Pompe disease do not store measurable amounts of glycogen [26]. Finally, our study utilizes the more powerful deep sequencing strategy and different approaches to identify DEGs instead of the more commonly used microarray datasets. Lastly, we compared the transcriptome of childhood and adult Pompe patients carrying the functionally similar c.-32-13T>G / 'null', genotype in 35/39 cases to understand the broad clinical variation observed in those patients.

The significantly downregulated genes related to the amino acid biosynthesis pathway in adult vs childhood Pompe disease compared to childhood versus adult controls discovered in study 1 might suggest a compensatory molecular mechanism in response to cellular stress and cellular metabolic adaptation. The overexpression of genes associated with immune response pathways might point to better cellular regeneration in fibroblasts from patients with adult compared to childhood Pompe disease.

Importantly, comparison between the analyses of study 1 and study 2 revealed only one common gene that showed altered expression between adult and childhood Pompe disease, which was the *GAA* gene. This can be taken as a clear indication of unbiased analysis. Despite the fact that study 2 identified a low number of DEGs, the trend of significantly up- or down-regulated genes in this study was the same but not significant (data not shown), which is likely due to a large variation between individuals. The large variation observed in the second study reflects reality but serves as a highly valuable 'stress test'. For this reason, we argue that the core dataset determined under this approach survived this stress test and may represent a more robust read- out for symptom onset in Pompe disease compared to either dataset from study 1 and 2 separately. The 3 genes *MAOA, STC-1*, and *NRK-1* have been implicated in various biologic processes including bone and muscle development, cellular calcium/ phosphate homeostasis, cellular metabolism, and cellular biogenic amine metabolic process. In addition, from the total of 24 DGEs identified in fibroblasts from Pompe patients compared to controls (Supp. **Fig. S11**), 16 were differentially expressed only in adult Pompe. The downregulated DEGs (in adulthood Pompe) were mainly involved in cell adhesion and cell-cell adhesion. Of particular interest, *WNT2* plays an important role in the Wnt and mTOR pathways, which are involved in several cellular processes such as proliferation, cell migration, apoptosis, protein biosynthesis, and autophagy [30, 31]. In fact, *WNT2* is an upstream signaling factor of the mTOR pathway that regulates glycogen synthase kinase (GSK-3) activity. GSK-3 has been shown to function in cellular division, proliferation, and survival, but furthermore, it is an important enzyme in glycogen metabolism. In addition the mTOR complex is involved in anabolic activities in the lysosome including autophagy, which has been shown to be dysfunctional in Pompe patients and consequently contributes to the pathological cascade in skeletal muscle [32, 33]. The down-regulation of *WNT2* expression might be explained as part of a compensatory molecular mechanism in response to cellular stress and cellular metabolic adaptation in adult Pompe patients.

On the other hand, patients with childhood Pompe disease showed up-regulation of differentially expressed genes such as *FNIP2* and *SLC9A7* that may be involved in energy and/or nutrient sensing through the AMPK and mTOR signaling pathways, transport of glucose plus other sugars, bile salts and organic acids, metal

ions and amine compounds [www.genecards.org/]. Notably, diseases associated with pathogenic *SLC9A7* gene variants include scapuloperoneal myopathy in which muscle fibers are also damaged or dysfunctional [34], as in Pompe disease [35-37]. In fact, scapular winging is a condition that has been reported in Pompe disease, but with low frequency. Also, three childhood Pompe patients (C2, C3 and C6) in this study (1A) presented scapular winging. However, as shown in Supplementary **Figure S10**, patient (C2) who's symptoms were so severe they interfered with daily life (moderate status) at the time of biopsy had a higher *SLC9A7* gene expression than the two other children (C3 and C6) who's symptoms barely interfered with daily life (mild status) at the time of biopsy. In fact, the gene expression of *SLC9A7* in C3 and C6 is almost the same. (Supp. **Fig. S13**). This argues against a causative role of *SLC9A7* in scapular winging.

Part of the explanation for the lack of consistency between study 1 and 2 may be technical, variations in control samples, small sample size, and possible batch effects because the patients and controls of the first study were handled in separate studies (1A and 1B). However, we included three identical RNA samples in study 1A in study 1B, and these gave similar results, making a batch effect unlikely. A more likely explanation can be deduced from **Figs. 2B** and **C**. In study 1, the patient cohorts are very well separated, while in study 2, the separation is far less defined with respect to age of onset and disease progression. Also in study 3, separation between the two cohorts is not as good as in study 1 (**Table 3**). This indicates that differences in gene expression profiles can be found at the group level provided that patient cohorts are sharply separated, but that it is difficult to link expression profiles to individual patients.

Pompe disease manifests with a broad clinical heterogeneity. Currently, the diagnosisand progression of Pompe disease relies predominately on GAA activity indices. Studies and development of factors that may modulate the underlying pathogenic etiologies of glycogen accumulation should offer the possibility to target those disease-markers that would monitor disease progression and could help in decision making about the timing of therapeutic intervention. Our study may also help to reveal new insight in potential disease-markers for Pompe Disease.

Our findings provide an argument to further investigate the impact of *MAOA* and *STC1* on the clinical course of Pompe disease, as they comprise the core set of DEGs that were consistently upregulated in patients with adulthood onset compared to patients with childhood onset. In light of these results, this core set of DEGs might be considered as potential disease-markers in patients with the IVS1/null genotype.

Funding information
Administrative Department of Science, Technology and Innovation (Colciencias, Colombia), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brasil; grant number: 234407/2014-0).

REFERENCES

1. Pompe, JC (1932). Over idiopathische hypertrofie van het hart. Ned Tijdsch Geneesk 76: 304-311.
2. Schoser, B, Laforet, P, Kruijshaar, ME, Toscano, A, van Doorn, PA, van der Ploeg, AT, et al. (2015). 208th ENMC International Workshop: Formation of a European Network to develop a European data sharing model and treatment guidelines for Pompe disease Naarden, The Netherlands, 26-28 September 2014. Neuromuscul Disord.
3. Engel, AG, Seybold, ME, Lambert, EH, and Gomez, MR (1970). Acid maltase deficiency: comparison of infantile, childhood, and adult types. Neurology 20: 382.
4. Engel, AG, Gomez, MR, Seybold, ME, and Lambert, EH (1973). The spectrum and diagnosis of acid maltase deficiency. Neurology 23: 95-106.
5. Reuser, AJJ, Hirschhorn, R, and Kroos, MA (2019). Pompe Disease: Glycogen Storage Disease Type II, Acid a-Glucosidase (Acid Maltase) Deficiency. In: Valle, D, S Antonarakis, A Ballabio, A Beaudet and GA Mitchell (eds). The Online Metabolic and Molecular Bases of Inherited Disease. McGraw-Hill Education: New York, NY.
6. van der Ploeg, AT, and Reuser, AJ (2008). Pompe's disease. Lancet 372: 1342-1353.
7. Kishnani, PS, Corzo, D, Nicolino, M, Byrne, B, Mandel, H, Hwu, WL, et al. (2007). Recombinant human acid [alpha]-glucosidase: major clinical benefits in infantile-onset Pompe disease. Neurology 68: 99-109.
8. van der Ploeg, AT, Clemens, PR, Corzo, D, Escolar, DM, Florence, J, Groeneveld, GJ, et al. (2010). A randomized study of alglucosidase alfa in late-onset Pompe's disease. N Engl J Med 362: 1396-1406.
9. V an den Hout, H, Reuser, AJ, Vulto, AG, Loonen, MC, Cromme-Dijkhuis, A, and Van der Ploeg, AT (2000). Recombinant human alpha-glucosidase from rabbit milk in Pompe patients. Lancet 356: 397-398.
10. Winkel, LP, Kamphoven, JH, van den Hout, HJ, Severijnen, LA, van Doorn, PA, Reuser, AJ, et al. (2003). Morphological changes in muscle tissue of patients with infantile Pompe's disease receiving enzyme replacement therapy. Muscle Nerve 27: 743-751.
11. de Vries, JM, van der Beek, NA, Hop, WC, Karstens, FP, Wokke, JH, de Visser, M, 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.
12. Toscano, A, and Schoser, B (2013). Enzyme replacement therapy in late-onset Pompe disease: a systematic literature review. J Neurol 260: 951-959.
13. Kuperus, E, Kruijshaar, ME, Wens, SCA, de Vries, JM, Favejee, MM, van der Meijden, JC, et al. (2017). Long- term benefit of enzyme replacement therapy in Pompe disease: A 5-year prospective study. Neurology 89: 2365-2373.
14. van der Ploeg, AT, Kruijshaar, ME, Toscano, A, Laforet, P, Angelini, C, Lachmann, RH, et al. (2017). European consensus for starting and stopping enzyme replacement therapy in adult patients with Pompe disease: a 10-year experience. Eur J Neurol 24: 768-e731.
15. de Vries, JM, Kuperus, E, Hoogeveen-Westerveld, M, Kroos, MA, Wens, SC, Stok, M, et al. (2017). Pompe disease in adulthood: effects of antibody formation on enzyme replacement therapy. Genet Med 19:90-97.
16. van Gelder, CM, Hoogeveen-Westerveld, M, Kroos, MA, Plug, I, van der Ploeg, AT, and Reuser, AJ (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.
17. van der Meijden, JC, Kruijshaar, ME, Harlaar, L, Rizopoulos, D, van der Beek, N, and van der Ploeg, AT (2018). Long-term follow-up of 17 patients with childhood Pompe disease treated with enzyme replacement therapy. J Inherit Metab Dis. 41: 1205-1214.
18. Manwaring, V, Prunty, H, Bainbridge, K, Burke, D, Finnegan, N, Franses, R, et al. (2012). Urine analysis of glucose tetrasaccharide by HPLC; a useful marker for the investigation of patients with Pompe and other glycogen storage diseases. J Inherit Metab Dis 35: 311-316.
19. Chien, YH, Han, DS, Hwu, WL, Thurberg, BL, and Yang, WS (2013). Myostatin and insulin-like growth factor I: potential therapeutic biomarkers for pompe disease. PLoS One 8: e71900.
20. Carrasco-Rozas, A, Fernandez-Simon, E, Lleixa, MC, Belmonte, I, Pedrosa-Hernandez,

I, Montiel-Morillo, E, et al. (2019). Identification of serum microRNAs as potential biomarkers in Pompe disease. Ann Clin Transl Neurol 6: 1214-1224.
21. Tarallo, A, Carissimo, A, Gatto, F, Nusco, E, Toscano, A, Musumeci, O, et al. (2019). microRNAs as biomarkers in Pompe disease. Genet Med 21: 591-600.
22. Wens, SC, van Gelder, CM, Kruijshaar, ME, de Vries, JM, van der Beek, NA, Reuser, AJ, et al. (2013). Phenotypical variation within 22 families with Pompe disease. Orphanet J Rare Dis 8: 182.
23. Kroos, M, Hoogeveen-Westerveld, M, van der Ploeg, A, and Reuser, AJ (2012). The genotype-phenotype correlation in Pompe disease. Am J Med Genet C Semin Med Genet 160C: 59-68.
24. Oba-Shinjo, SM, da Silva, R, Andrade, FG, Palmer, RE, Pomponio, RJ, Ciociola, KM, et al. (2009). Pompe disease in a Brazilian series: clinical and molecular analyses with identification of nine new mutations. J Neurol 256: 1881-1890.
25. Huie, ML, Chen, AS, Tsujino, S, Shanske, S, DiMauro, S, Engel, AG, et al. (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. Human Molecular Genetics 3: 2231-2236.
26. Umapathysivam, K, Hopwood, JJ, and Meikle, PJ (2005). Correlation of acid alpha-glucosidase and glycogen content in skin fibroblasts with age of onset in Pompe disease. Clin Chim Acta 361: 191-198.
27. Smith, PK, Krohn, RI, Hermanson, GT, Mallia, AK, Gartner, FH, Provenzano, MD, et al. (1985). Measurement of protein using bicinchoninic acid. Anal Biochem 150: 76-85.
28. Feeney, EJ, Spampanato, C, Puertollano, R, Ballabio, A, Parenti, G, and Raben, N (2013). What else is in store for autophagy? Exocytosis of autolysosomes as a mechanism of TFEB-mediated cellular clearance in Pompe disease. Autophagy 9: 1117-1118.
29. Palermo, AT, Palmer, RE, So, KS, Oba-Shinjo, SM, Zhang, M, Richards, B, et al. (2012). Transcriptional response to GAA deficiency (Pompe disease) in infantile-onset patients. Mol Genet Metab 106: 287-300.
30. Lim, JA, Li, L, Shirihai, OS, Trudeau, KM, Puertollano, R, and Raben, N (2017). Modulation of mTOR signaling as a strategy for the treatment of Pompe disease. EMBO Mol Med 9: 353-370.
31. Lyu, JW, Xu, XB, Ji, KQ, Zhang, N, Sun, Y, Zhao, DD, et al. (2019). Activated mTOR signaling pathway in myofibers with inherited metabolic defect might be an evidence for mTOR inhibition therapies. Chin Med J (Engl) 132: 805-810.
32. Lim, JA, Meena, NK, and Raben, N (2019). Pros and cons of different ways to address dysfunctional autophagy in Pompe disease. Ann Transl Med 7: 279.
33. Nascimbeni, AC, Fanin, M, Tasca, E, Angelini, C, and Sandri, M (2015). Impaired autophagy affects acid alpha-glucosidase processing and enzyme replacement therapy efficacy in late-onset glycogen storage disease type II. Neuropathol Appl Neurobiol 41: 672-675.
34. Tzschach, A, Ullmann, R, Ahmed, A, Martin, T, Weber, G, Decker-Schwering, O, et al. (2011). Christianson syndrome in a patient with an interstitial Xq26.3 deletion. Am J Med Genet A 155A: 2771-2774.
35. Herbert, M, Case, LE, Rairikar, M, Cope, H, Bailey, L, Austin, SL, et al. (2019). Early-onset of symptoms and clinical course of Pompe disease associated with the c.-32-13T>G variant. Mol Genet Metab 126: 106-116.
36. van der Beek, NA, de Vries, JM, Hagemans, ML, Hop, WC, Kroos, MA, Wokke, JH, 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.
37. Oya, Y, Morita, H, Ogawa, M, Nonaka, I, Tsujino, S, and Kawai, M (2001). [Adult form of acid maltase deficiency presenting with pattern of muscle weakness resembling facioscapulohumeral dystrophy]. Rinsho Shinkeigaku 41: 390-396.

SUPPLEMENTARY DATA / SUPPORTING INFORMATION

Supplementary text

Clinical features and disease severity

Table 2 provides demographic characteristics of the patients in cohort 1 and cohort 2. The majority (76.5%) of affected children in cohort 1 i s male. Only one of the 1 7 patients (5.9%) in cohort 1 was partially wheelchair and ventilation dependent at the time of study. All others (94%) were still ambulant. The most frequent first symptom was delayed motor development (35.3%) followed by elevated CK (29.4%). Among the adult patients in cohort 2, 36.4% are males, Five of these patients (22.7%) were partially wheelchair dependent, 2 (9.1%) wheelchair dependent, 1 (4.5%) using a walking aid, while the others (63.6%) were ambulant. Six (27.3%) patients were ventilation dependent. The most frequent first symptom in cohort 2 was fatigue, presenting in 6 (27.3%) patients, followed by difficulty climbing stairs in 4 (18.2%).

One childhood patient (C11) became wheelchair dependent at 4 years of age, but is able to walk a distance of 200 meter. Another child (C1) was wheelchair dependent at the age of 11 years, but regained walking ability during ERT. Two adult patients were fully wheelchair dependent at 51 (A17) and 60 years of age (A21), respectively. Five adult patients had become partially wheelchair dependent from 36 and 63 years onwards. One childhood patient (C2) became ventilator-dependent, using nightly bi- level positive airway pressure (BiPaP) at 12 years, and six adult patients were treated with either invasive or non-invasive ventilation support. The patients using non- invasive ventilation were: A11 at age 57 years, A19 at age 54 years and A20 at age 53 years. The patients that required invasive ventilator support are listed as follow: A17 at age 39 years, A18 at age 33 and A21 at age 67 years (Supp. Tables S1A-B).

As shown in Supplementary **Tables S1A-B** the currently most severely affected Pompe patients are two adults (patients A17 and A21). Patient A17 had first symptoms at 30 years, started invasive ventilation support at 39 years and became fully wheelchair dependent at the age of 51 years. Patient A21 with an age of onset at 40 years, was fully wheelchair dependent at the age of 60 years and was invasively ventilated at the age of 67 years. Of course, disease severity in Pompe disease is a matter of age of onset, rate of disease progression, and moment of evaluation. In this respect, the childhood patient (C2) that had first symptoms under the age of 1 year and started to use ventilation support at 12 years of age may in time become the most severely affected. The clinical course is hard to predict in most cases. The clinical features of all child and adult patients with Pompe disease and kind of first symptoms are described in Supplementary **Tables S1A-D**. The onset of symptoms and severity of disease progression varied substantially even between these patients with

the same set of disease-associated variants on both *GAA* alleles. Their *GAA* genotypes and complementary information of the second disease-associated variant are listed in Supplementary **Table S2**.

Supplementary Table S1A. Clinical features of patients with Pompe disease in study 1.

Clinical phenotype	Childhood Pompe patients												Adult Pompe patients										
Patient id	Ch1	Ch2	Ch3	Ch4	Ch5	Ch6	Ch7	Ch8	Ch9	Ch10	Ch11	Ch12	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
Age at first symptoms (y)	0.5	1	0.7	2.5	0.8	4	5	6	0.8	1.5	1	8	49	43	40	43	40	61	42	50	55	62	47
Age at time of biopsy (y)	2.8	16.3	2.6	10.9	9.1	16	11	8.6	9.8	13.4	8.3	1.6	58.1	48.4	49.5	50.8	48.7	68	58.6	54	68.9	72.9	60
Clinical status at biopsy	mild	mod	mild	mild	mild	mild	mild	very mild	very mild	mild/ mod?	?	unk-nown	mild/ mod	mild	mod	mod	mild	mild	mod	mild	mild/ mod	mild/ mod	sev
Diagnosis age (y)	2.5	1.3	2.3	3	1.6	4	10.8	7.8	2.5	2	2.2	7.3	50	26	43	41	43	67	57	52	68	72	56
Start ERT (y)	11.9	15.2	2.9	12.9	8.9	16	11	8.5	9.9	no	no	?	58	no	50	53	49	na	58	54	69	73	61
Gender	M	M	M	M	F	M	M	F	M	M	M	F	F	F	F	F	M	M	F	F	M	F	M
Wheelchair age (y)	11										<4												
Mobility (wheelchair)											+^												
Mobility (ambulant)	+*	+	+	+	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+	+
Ventilation age (y)		12																					57
Ventilation type		BP																					Non-I
Elevated CK			+			+		+	+			+											
FVC (sitting) % predicted	97	66	86	66	102	67	102	104	82	na	59	71	79	97	53	70	84	94	89	96	98	68	54
FVC (supine) % predicted	87	45	na	54	101	53	97	92	80	na	48	na	52	77	33	38	73	75	75	86	77	49	22
MRC score (%)	78	89	80	83	92	98	97	99	99	na	80	na	84	90	85	84	89	95	84	94	89	93	86

Abbreviations: Ch1-Ch12= childhood onset Pompe disease and A1-A11= adult onset Pompe disease included in study 1. BP = Nightly BiPaP; Non-I= non-invasive; na= not available/applicable; M= male, F= female; mild= symptoms barely interfere with daily life; mod= moderate: symptoms so severe that they interfere with daily life; sev= severe: in the need of a wheelchair and/or ventilator; +*= Ch1 starts walking during ERT; +^= Ch11 is partly wheelchair dependent, can walk 200m.

Supplementary Table S1B. Clinical features of patients with Pompe disease in study 2.

Clinical phenotype	Childhood Pompe patients					Adult Pompe patients										
Patient id	Ch13	Ch14	Ch15	Ch16	Ch17	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22
Age at first symptoms (y)	0.75	9.1	2	7	13	32	39	20	22	41	30	33	21	44	40	62
Age at time of biopsy (y)	1.06	10.4	0.35	10.8	14.3	39	48	52	41	53.9	57.6	40	61.5	62.7	70	72
Clinical status at biopsy	mild	mild	mild	sev	mild	mild	mild	mild/ mod	sev	mod	sev	sev	mod/ sev	sev	sev	mild
Diagnosis age (y)	0	9.4	0.33	9.5	14.1	31	48	52	37	53	41	34	42	44	45	63
Start ERT (y)	1	10.6	13.1	33	14.9	39	48.9	52.3	52	54.7	57.7	59	61.6	62.9	70.4	74
Gender	F	M	M	M	M	F	M	F	F	F	M	M	F	M	F	F
Wheelchair age (y)				22					39		51	36	43	62	60	63
Mobility (wheelchair)				+					+*		+	+*	+*	+*	+	+*
Mobility (ambulant)	+	+	+		+	+	+	+								
Mobility (walking aid)									+	+		-*	-*	-*		-*
Ventilation age (y)											39	33	54	53	67	
Ventilation type											INV	INV	Non-I	Non-I	INV	
FVC (sitting) % predicted	115	84	97	na	90	87	61	103	48	75	na	na	18	33	34	81
FVC (supine) % predicted	109	63	na	na	85	86	50	89	23	41	na	na	na	na	na	76
MRC score (%)	98	97	33	na	97	82	94	85	64	79	57	60	68	72	na	81

Abbreviations: Ch13-Ch17 are childhood onset Pompe patients, and A12-A22 are adult onset Pompe patients included in study 2. INV= invasive; Non-I= non-invasive; +*= partially wheelchair dependent; -*= no walking aid; na= not available; M= male, F= female; mild= symptoms barely interfere with daily life; mod= moderate: symptoms so severe that they interfere with daily life; sev= severe: in need of a wheelchair and/or ventilator; +=yes.

Supplementary Table S1C. Kind of first symptoms of patients with Pompe disease in study 1.

Clinical phenotype	Childhood onset Pompe patients												Adult onset Pompe patients										
patient ID	Ch1	Ch2	Ch3	Ch4	Ch5	Ch6	Ch7	Ch8	Ch9	Ch10	Ch11	Ch12	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
Kind of first symptoms																							
hypotonia		+	+		+				+														
delayed motor development		+			+				+	+	+												
delayed growth								+															
growth stagnation							+																
feeding problems	+									+													
difficulty swallowing																						+	
difficulty chewing																						+	
slipping through									+														
fatigue													+	+						+			+
pulmonary weakness																							+
lower back pain																		+					
muscle ache																		+					
mild muscle weakness													+										
falling				+																			
walking problems				+																			
abnormal gait										+													
difficulty walking																+							
difficulty running																					+		
difficulty with practicing sports															+								
difficulty climbing stairs																	+		+				
difficulty standing from squatting position														+									
difficulty changing posture																			+				
difficulty standing straight																				+			
difficulty stumbling																+							
stiffness															+								
dysarthria																						+	
diarrhea							+	+															

Supplementary Table S1D. Kind of first symptoms of patients with Pompe disease in study 2

Clinical phenotype	Childhood onset Pompe patients					Adult onset Pompe patients									
patient ID	Ch13	Ch14	Ch15	Ch16	Ch17	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22
Kind of first symptoms																
delayed motor development	+															
fatigue		+			+			+								+
Respiratory failure												+				
lower back pain							+									+
muscle ache		+				+				+						
mild muscle weakness of legs															+	
falling		+														
waddling gait														+		+
abnormal gait					+											
difficulty walking								+		+						
difficulty with practicing sports				+												
difficulty climbing stairs						+							+			
difficulty standing from supine position											+					
difficulty changing posture							+									
stumbling									+							
stiffness													+			
reduced force legs			+													
difficulty standing up									+							

Supplementary Table S2. *GAA* genotypes of patients with Pompe disease reported in studies 1 plus 2.

	pt ID	Location	DNA	Protein	Type of variant (DNA)	predicted severity	ACMG classification	phenotype with null allele	CR IM
CHILDHOOD ONSET PATIENT	**Ch1**	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	-*
	Ch2	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	-*
	Ch3	E15	c.2135T>C	p.(Leu712Pro)	substitution	less severe	likely pathogenic	unknown (disease-associated)	+
	Ch4	I16	c.2331+2T>A	p.?	substitution	very severe	pathogenic	unknown (disease-associated)	-*
	Ch5	E5	c.923A>C	p.(His308Pro)	substitution	potentially less severe	likely pathogenic	classic infantile	+
	Ch6	E10	c.1441T>C	p.(Trp481Arg)	substitution	potentially less severe	likely pathogenic	classic infantile	+
	Ch7	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	-*
	Ch8	I16	c.2331+2T>A	p.?	substitution	very severe	pathogenic	unknown (disease-associated)	-*

Supplementary Table S2 continued.

	pt ID	Location	DNA	Protein	Type of variant (DNA)	predicted severity	ACMG classification	phenotype with null allele	CR IM
CHILDHOOD ONSET PATIENT	**Ch9**	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	-*
	Ch10	I17	c.2481+102_2646+31del	p.(Gly828_Asn882del)	deletion	very severe	pathogenic	classic infantile	+
	Ch11	E6	c.1051del	p.(Val351Cysfs*41)	deletion	very severe	pathogenic	unknown (disease-associated)	-
	Ch12	E15	c.2066_2070dup	p.(Ala691Serfs*7)	duplication	very severe	pathogenic	unknown (disease-associated)	-
	Ch13	E15	c.2135T>C	p.(Leu712Pro)	substitution	less severe	likely pathogenic	unknown (disease-associated)	+
	Ch14	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	-*
	Ch15	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	-*
	Ch16	E10	c.1548G>A	p.(Trp516*)	substitution	very severe	pathogenic	classic infantile	-*
	Ch17	E14	c.1933G>A	p.(Asp645Asn)	substitution	potentially less severe	pathogenic	classic infantile	+ #
ADULT ONSET PATIENT	**A1**	E10	c.1548G>A	p.(Trp516*)	substitution	very severe	pathogenic	classic infantile	-*
	A2	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	-*
	A3	I17	c.2481+102_2646+31del	p.(Gly828_Asn882del)	deletion	very severe	pathogenic	classic infantile	+
	A4	E5	c.925G>A	p.(Gly309Arg)	substitution	potentially less severe	likely pathogenic	classic infantile	+
	A5	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	-*
	A6	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	-*
	A7	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	-*
	A8	I6	c.1076-22T>G	p.?	substitution	potentially mild	pathogenic	childhood	?
	A9	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	-*
	A10	E10	c.1548G>A	p.(Trp516*)	substitution	very severe	pathogenic	classic infantile	-*
	A11	E5	c.896T>C	p.(Leu299Pro)	substitution	potentially less severe	likely pathogenic	classic infantile	+
	A12	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	-*
	A13	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	-*
	A14	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	-*
	A15	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	-*
	A16	I6	c.1076-22T>G	p.?	substitution	potentially mild	pathogenic	childhood	?
	A17	I17	c.2481+102_2646+31del	p.(Gly828_Asn882del)	deletion	very severe	pathogenic	classic infantile	+
	A18	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	*
	A19	E5	c.925G>A	p.(Gly309Arg)	substitution	potentially less severe	likely pathogenic	classic infantile	+
	A20	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	-*
	A21	E2	c.525del	p.(Glu176Argfs*45)	deletion	very severe	pathogenic	classic infantile	-*
	A22	E5	c.925G>A	p.(Gly309Arg)	substitution	potentially less severe	likely pathogenic	classic infantile	+

All patients included in this study share the same variant on the first GAA allele, c.-32-13T>G, and a second variant on allele 2 as listed. Ch1-C1h7: patients with childhood onset Pompe disease; A1-A22: patients with adult onset Pompe disease. E= exon, I= intron, += positive; -= negative; ?= unknown. *CRIM, as published by Bali et al. 2012. *Predicting cross-reactive immunological material (CRIM) status in Pompe disease using GAA mutations: lessons learned from 10 years of clinical laboratory testing experience. Am J Med Genet C Semin Med Genet. 160C(1):40-9*; #= CRIM as published by Kishnani et al. 2006. *Chinese hamster ovary cell-derived recombinant human acid alpha-glucosidase in infantile-onset Pompe disease J Pediatr. 149(1):89-97.*

Supplementary Table S3. Description of genes significantly differentially expressed in adult onset compared to childhood onset Pompe disease. These genes were revealed by pairwise comparisons of differentially expressed genes in Pompe disease affected individuals from studies 1 plus 2 (corrected for controls).

Supp. Table S3A.

Gene symbol	Chromosomal location	Subcellular location	Molecular function	Tissue expression	Signaling pathway	Biomarker application(s)	Additional information
NRK	Xq22.3	cytosol	-receptor signaling protein -serine/threonine kinase activity -ATP binding	overexpressed in: -ovary -adrenal gland	-TNF receptor signaling pathway		The encoded protein may be involved in the induction of actin polymerization in late embryogenesis

Description of genes significantly downregulated in adult compared with childhood onset Pompe disease.

Supp. Table S3B.

Gene symbol	Chromosomal location	Subcellular location	Molecular function	Tissue expression	Signaling pathway	Biomarker application (s)	Additional information
STC1	8p21.2	-extracellular -nucleus -plasma membrane	-hormone activity	overexpressed in: -thyroid -pituitary	ND	Diagnosis (ovarian cancer)	The protein may play a role in the regulation of renal and intestinal calcium and phosphate transport, cell metabolism, or cellular calcium/phosphate homeostasis. Overexpression of human stanniocalcin 1 in mice produces high serum phosphate levels, dwarfism, and increased metabolic rate
MAOA	Xp11.3	-mitochondrion -cytosol	-primary amine oxidase activity -flavin adenine dinucleotide binding -serotonin binding	ubiquitous	-Enzymatic degradation of dopamine by monoamine oxidase -Tryptophan metabolism -Tyrosine metabolism Glycine, serine and threonine metabolism -Arginine and proline metabolism -Histidine metabolism Phenylalanine metabolism -Metabolic pathways		Catalyzes the oxidative deamination of biogenic and xenobiotic amines and has important functions in the metabolism of neuroactive and vasoactive amines in the central nervous system and peripheral tissues

Description of genes significantly upregulated in adult compared with childhood onset Pompe disease.

Supplementary Table S4. Number of genes differentially expressed in adult Pompe disease compared to adult controls and in childhood Pompe compared with childhood controls in the studies 1 plus 2.

Supp. Table S4A.

Comparison	Study	Down	Up	Total of Genes
adult patient vs adult control	1 (A-B)	3292	3315	6607
adult patient vs adult control	2	28	25	53

Number of genes differentially expressed in adult onset Pompe disease compare with adult controls in studies 1 plus 2.

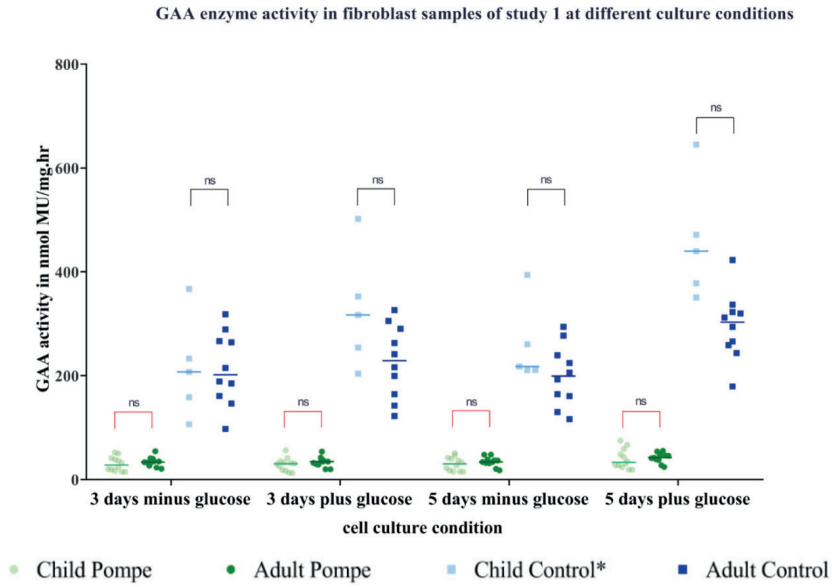
Suppl. Table S4B.

Comparison	Study	Down	Up	Total of Genes
childhood patient vs childhood control	1 (A-B)	2002	2424	4426
childhood patient vs childhood control	2	49	40	89

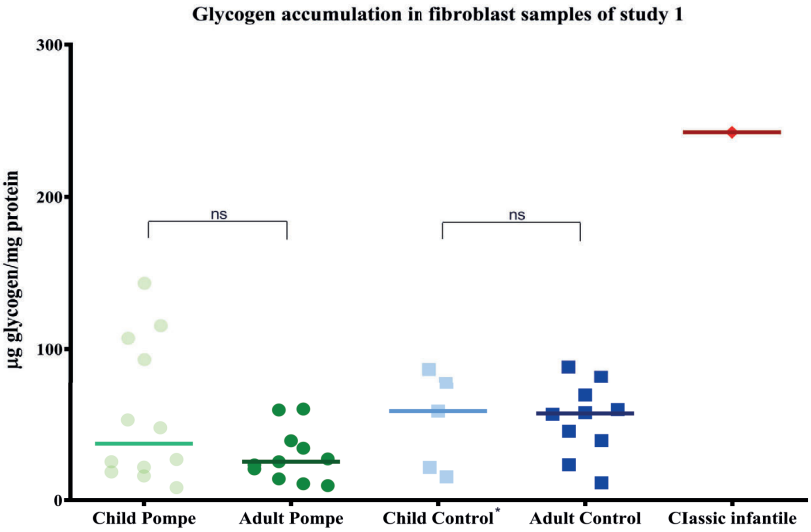
Number of genes differentially expressed in childhood onset Pompe disease compared with childhood controls in studies 1 plus 2.

Suppl. Table S5. The corresponding number identification of childhood and adult onset Pompe patients in the tables compared to the multidimensional scaling plot (MDS, Figure 3A)

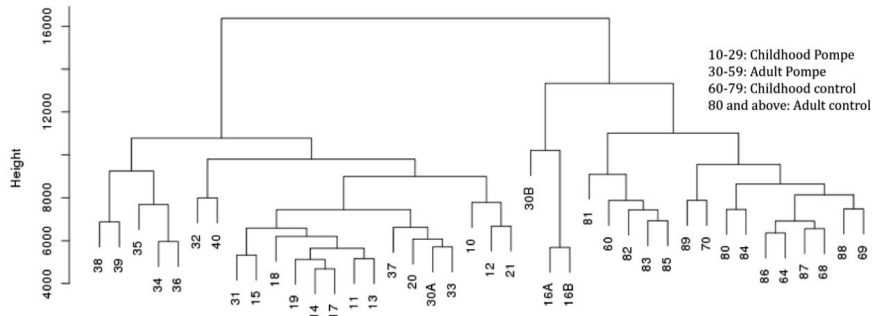
Clinical phenotype	Tables Patient ID	MDS patient ID	Clinical phenotype	Tables Patient ID	MDS patient ID
Childhood onset Pompe patients	CH1	12	Adult onset Pompe patients	O1	30A, 30B
	CH2	10		O2	37
	CH3	21		O3	40
	CH4	11		O4	38
	CH5	14		O5	39
	CH6	15		O6	34
	CH7	16A, 16B		O7	36
	CH8	17		O8	31
	CH9	13		O9	32
	CH10	20		O10	33
	CH11	18		O11	35
	CH12	19			



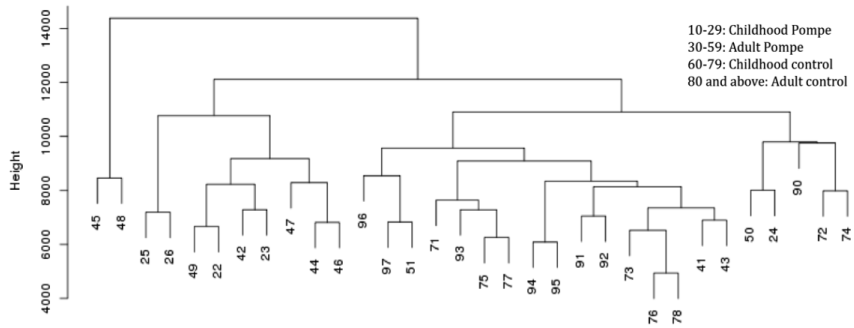
Supplementary Figure S1. GAA enzyme activity in fibroblast samples of study 1 at different culture conditions. The cells were routinely grown in DMEM as described in Material and Methods. For measurement of the lysosomal glycogen content the medium was replaced by glucose free medium for either 3 or 5 days where after the cells were harvested and the GAA activity measured. The horizontal line presents the median value. As in Fig.4A, ns = non-significant (P>0.05). The differences between the GAA activities in patients *versus* controls were in all instances significant (P<0.05). The same methodology was used as in Fig 4A. *Fibroblasts from foreskin were not included. The 2way ANOVA test was used for the analysis



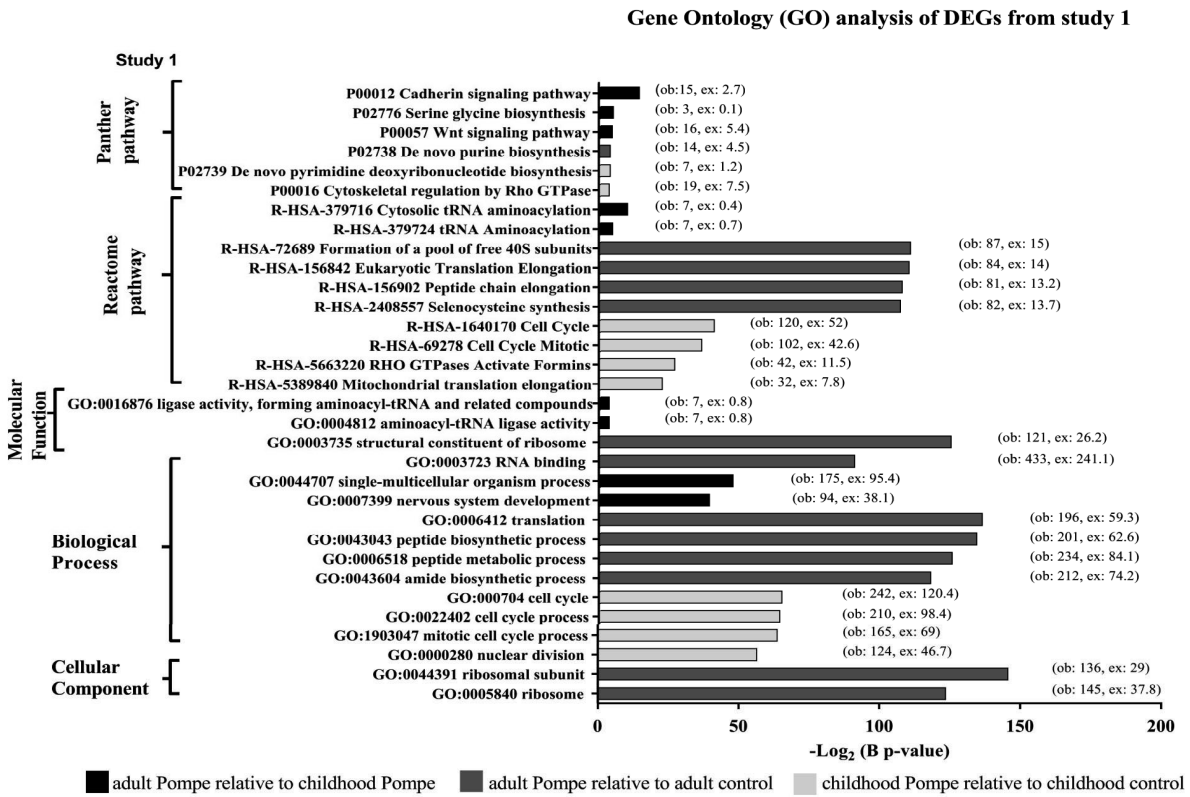
Supplementary Figure S2. Glycogen accumulation in fibroblast samples of study 1. The horizontal line presents the median value. There were no significant differences between the groups (ns; P>0.05). Only the single cell line from a patient with classic infantile Pompe disease that was included as a control in this study showed significant accumulation of glycogen when compared with all other samples. The same methodology and statistical analysis was used as in Fig 4A. *Fibroblasts from foreskin were not included.



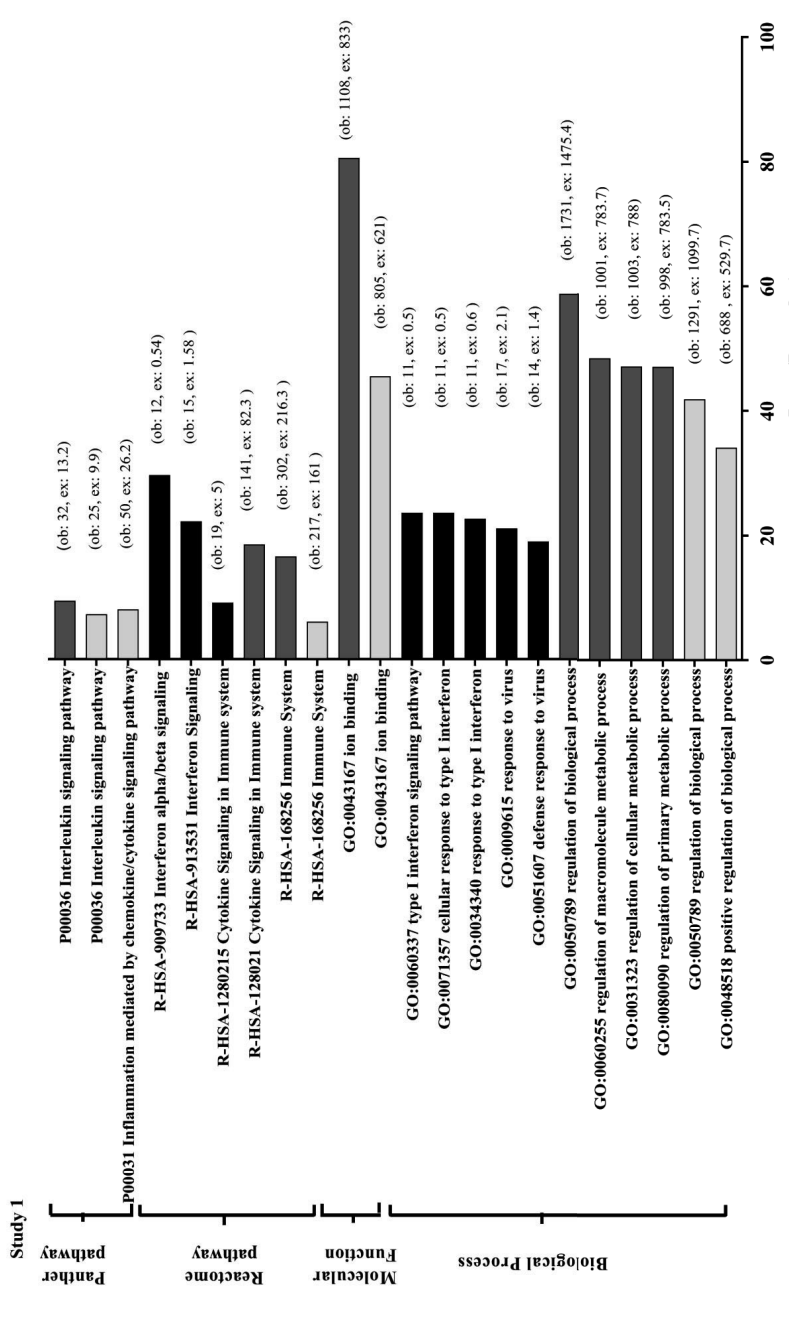
Supplementary Figure S3A. Hierarchical cluster analysis by RNA sequencing comparing Pompe patients and healthy controls from studies 1A plus B. The patients cluster together in the left branch, and the controls in the right sub-branch of the top-right branch. The left sub-branch of the top-right branch includes one sample of a child and one of an adult onset of Pompe disease from study 1B and one sample of a child from study 1A. These three samples are quite different from the controls however, as expected. Within the patients' tree, the left branch is all patients, whereas the right branch is all childhood patients, together with 4 adult patients. Within the controls tree, there is a clear separation between a branch with only child controls, and a branch with a mixture of child and adult.



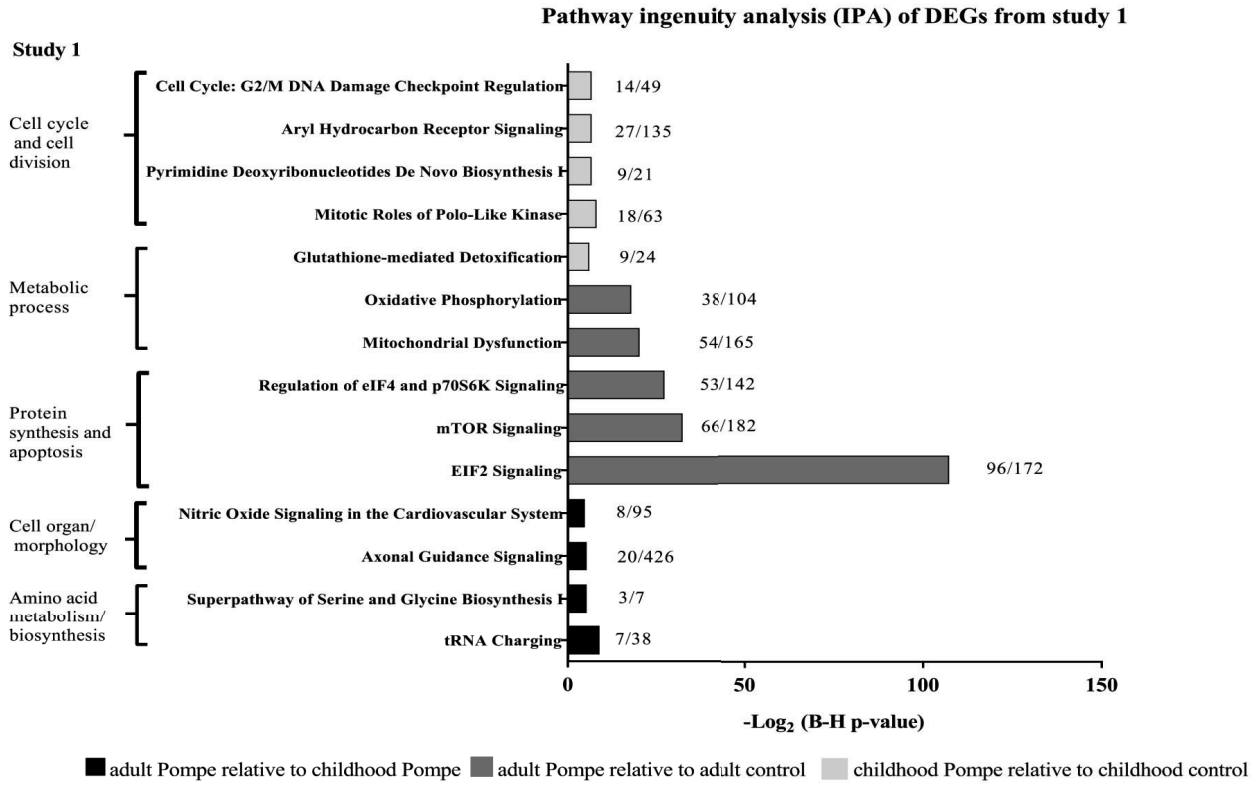
Supplementary Figure S3B. Hierarchical cluster analysis by RNA sequencing comparing Pompe patients and healthy controls from study 2. Most patient samples cluster apart from the controls except for five (24, 41, 43, 50, 51) that cluster with the controls. Two samples from (adult) patients (45 and 48) cluster together apart from all other.



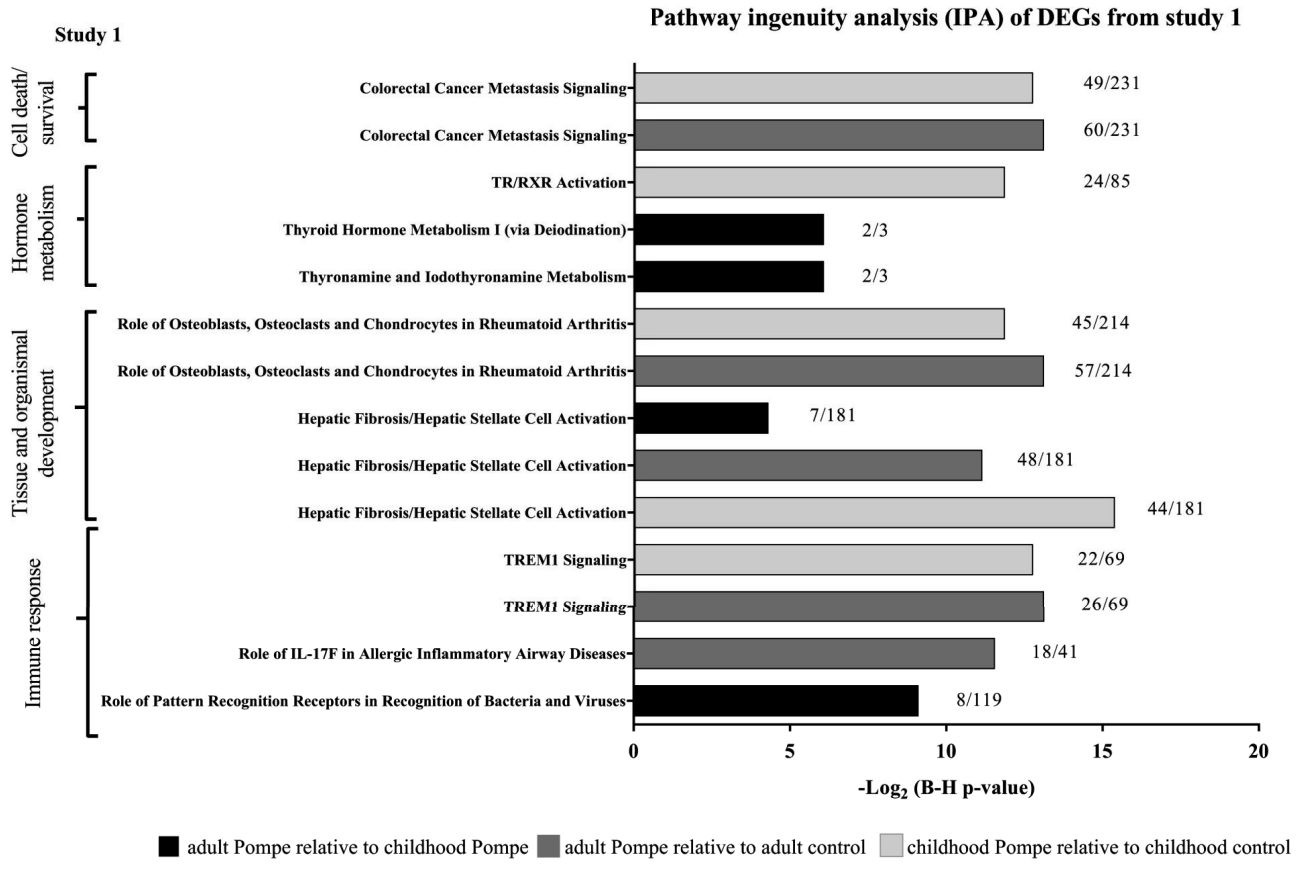
Supplementary Figure S4. Gene Ontology (GO) analysis of DEGs from study 1. The-DEGs detected by RNA-Seq were investigated for significantly enriched GO terms (B p-value <0.05) using the Panther classification system. All significant GO terms were ranked according to the Bonferroni correction p-values. Bars represent p-values adjusted by Bonferroni (B) correction multiple test (-Log_2 B p-value) of the most significant GO terms. Colors represent the GO enriched terms in adult *versus* childhood onset Pompe disease, relative to healthy controls. The GO terms are listed by category: Panther pathways, Reactome pathways, molecular function (MF), biological process (BP) and cellular component (CC). GOID and the term are included. Number of observed genes (ob), and number of expected genes (ex) are given at the end of the bars. **Supp. Fig. S4A**, depicts the set of downregulated genes, **Supp. Fig. S4B** the set of upregulated genes.



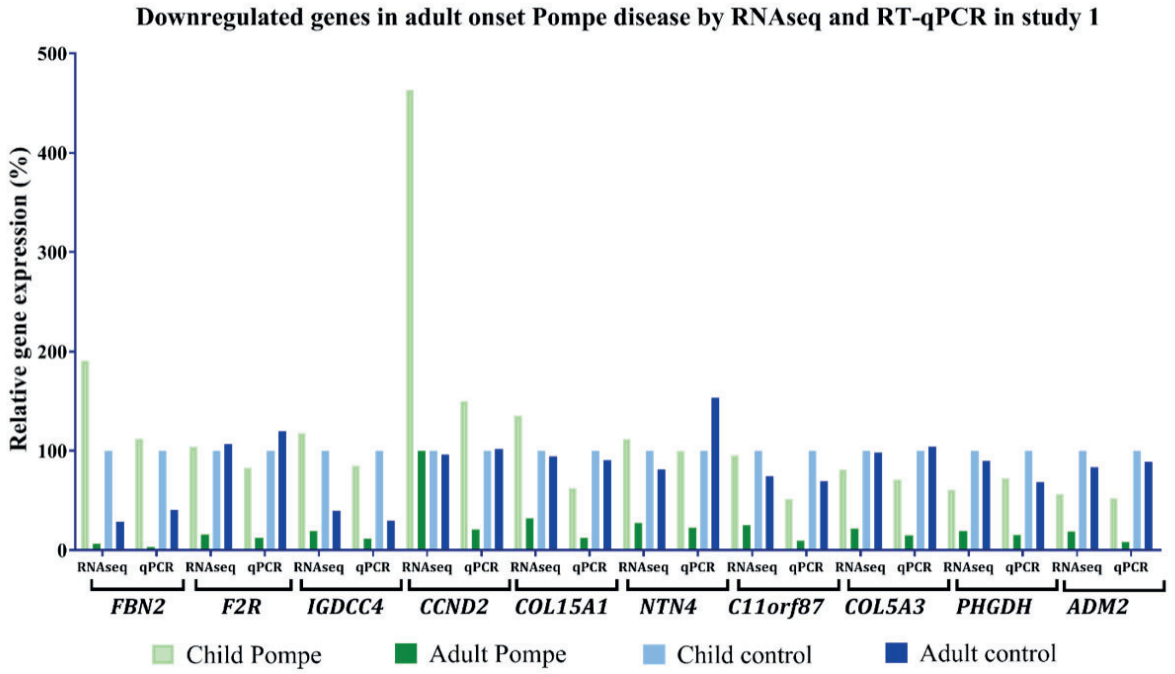
Supplementary Figure S4B



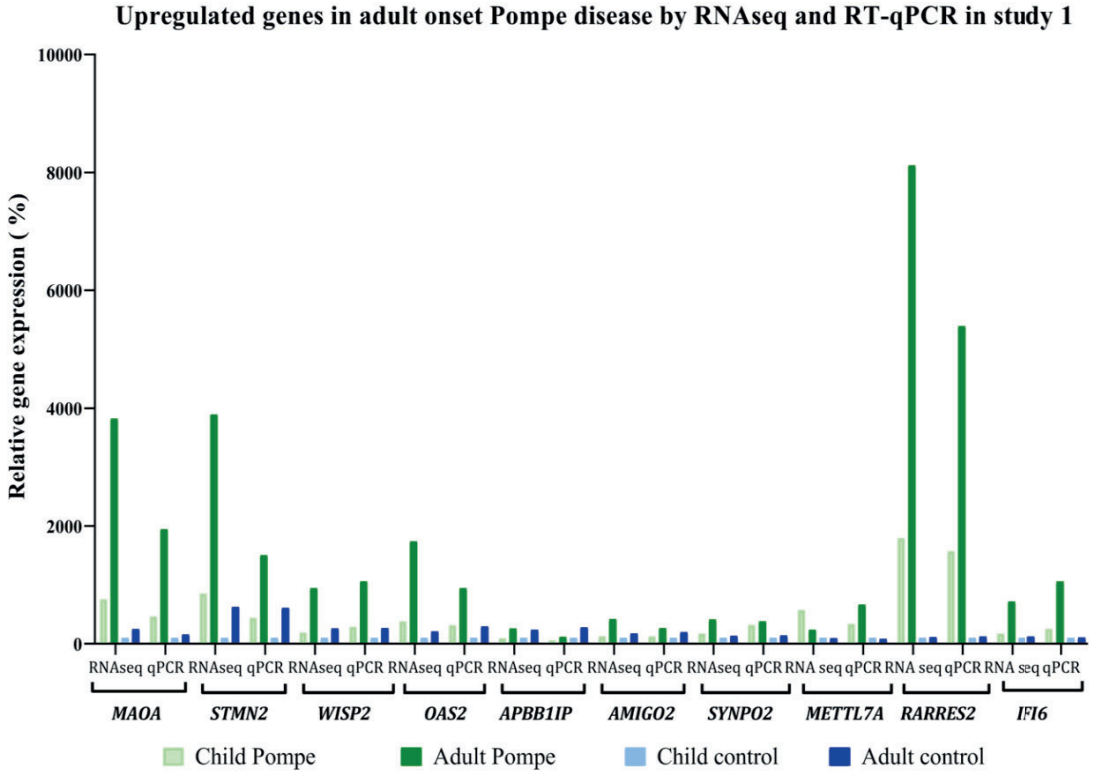
Supplementary Figure S5. Pathway ingenuity analysis (IPA) of DEGs from study 1. The DEGs detected by RNA-Seq were investigated for identification of pathways (B-H p-value <0.05) using Ingenuity software and ranked accordingly. Bars represent adjusted p-values (-Log_2 B-H p-value) of the most significant pathways. Colors represent the pathways identified in adult *versus* childhood onset disease relative to healthy controls. Top functions of the pathways are listed at the right. The figures at the end of the bars represent the number of DEGs (in patients) per pathway relative to the total number of genes assigned to that pathway. **Supp. Fig. S5A**, depicts the set of downregulated genes by pathway, **Supp. Fig. S5B** the set of upregulated genes by pathway.



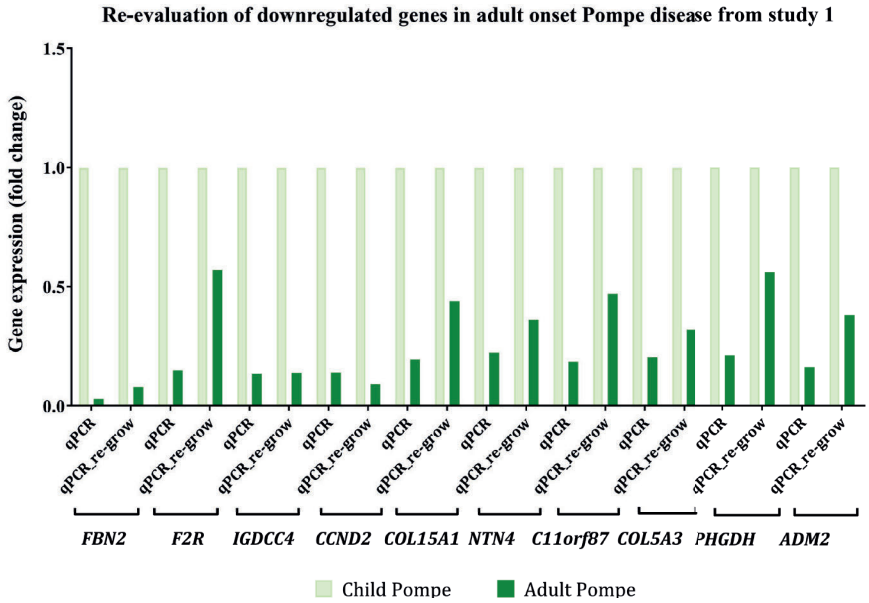
Supplementary Figure S5B



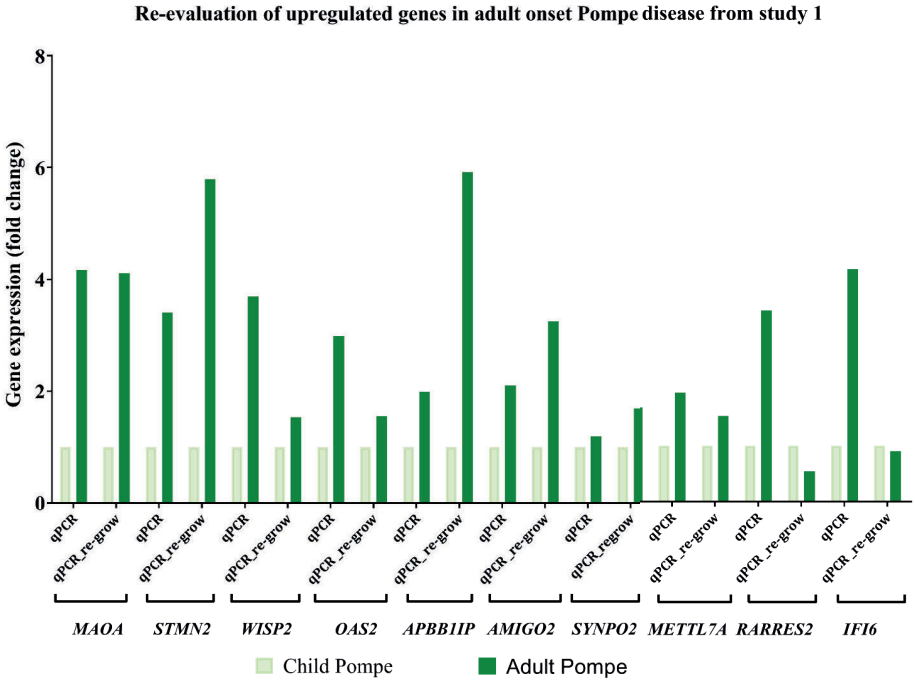
Supplementary Figure S6A. Downregulated genes in adult onset Pompe disease by RNAseq and RT-qPCR (study 1). Ten downregulated genes identified by RNAseq were analyzed by qPCR to validate the quality of the RNAseq data. The relative gene expression levels of selected genes measured by RT-qPCR were in good agreement with the RNA-sequencing data



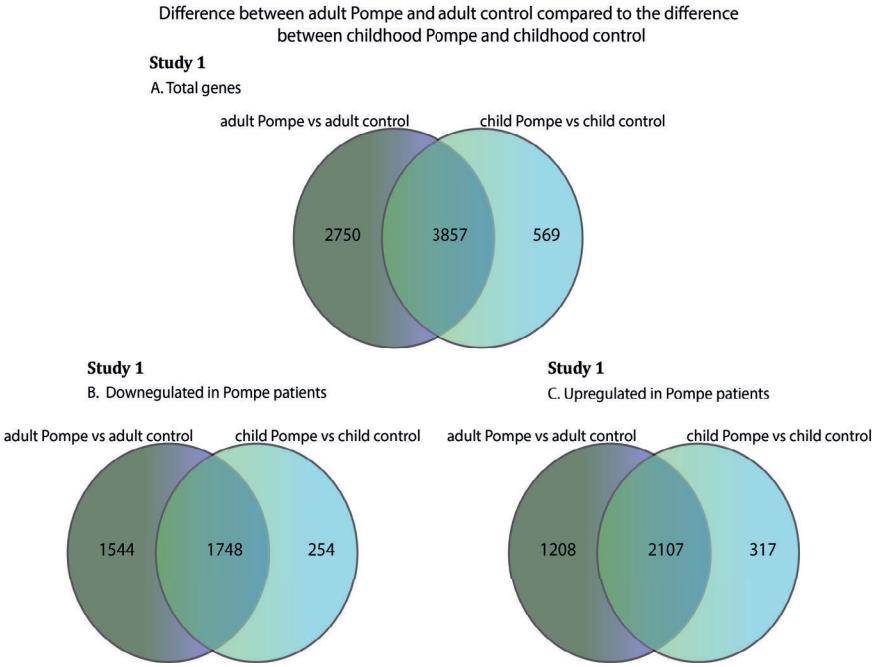
Supplementary Figure S6B. Upregulated genes in adult onset Pompe disease by RNAseq and RT-qPCR (study 1). Ten upregulated genes identified by RNAseq were analyzed by qPCR to validate the quality of the RNAseq data. The relative gene expression levels of most selected genes measured by RT-qPCR were in good agreement with the RNA-sequencing data.



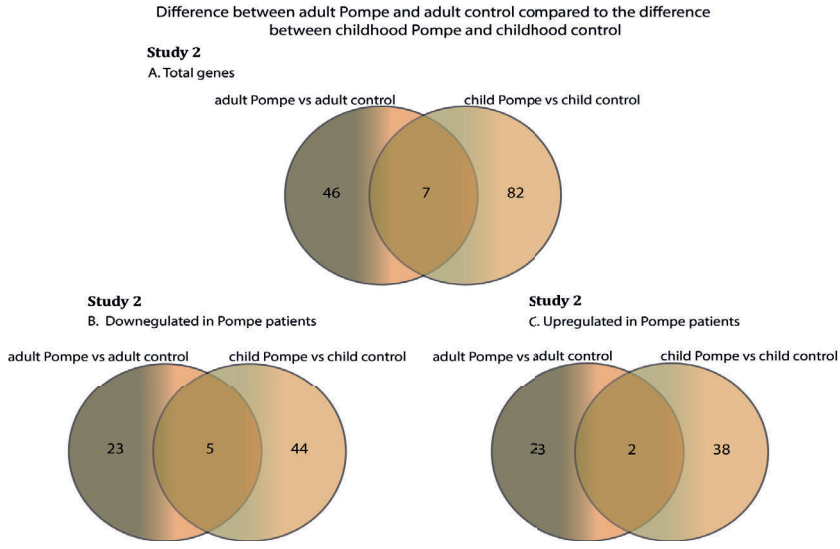
Supplementary Figure S6C. Re-evaluation of downregulated genes in adult onset Pompe disease from study 1. The fibroblasts of patients included in study 1 were re-grown and gene expression levels of downregulated genes were re-evaluated by RT-qPCR in order to analyze dependency on culture conditions and reproducibility.



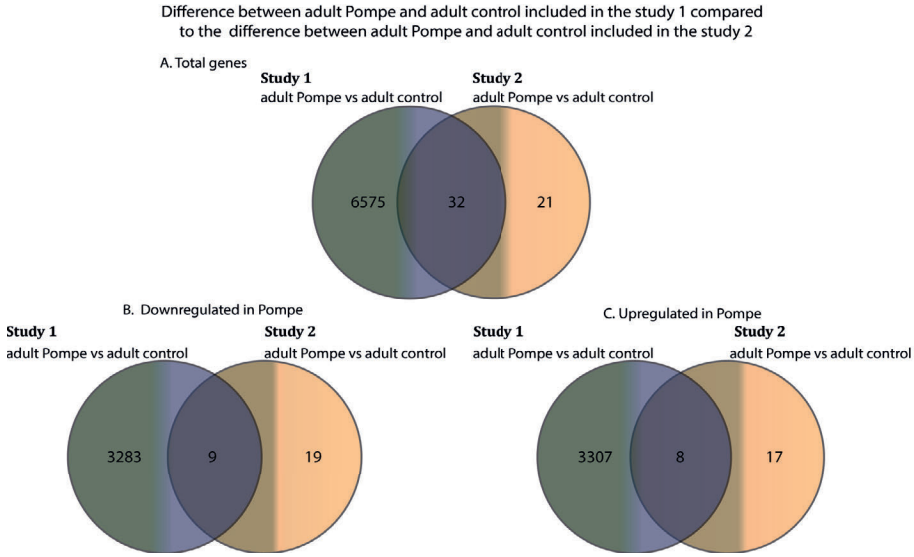
Supplementary Figure S6D. Re-evaluation of upregulated genes in adult onset Pompe disease from study 1. The fibroblasts of patients included in study 1 were re-grown and gene expression levels of upregulated genes were re-evaluated by RT-qPCR in order to analyze dependency on culture conditions and reproducibility.



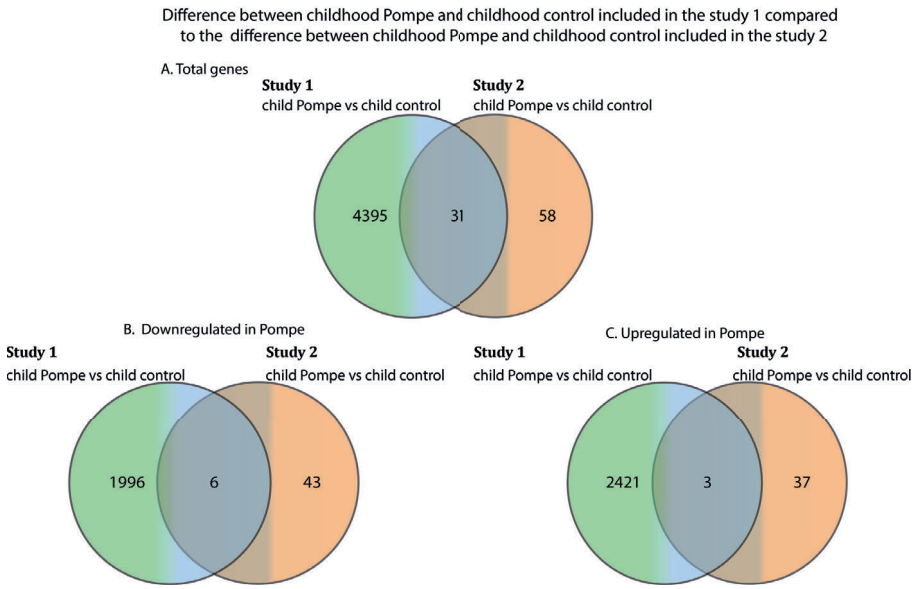
Supplementary Figure S7. Venn diagram showing the difference between affected adults and children compared to unaffected controls from study 1. (A) A total of 3857 genes are differentially expressed in fibroblasts of the patients. **(B)** 1748 genes are downregulated and **(C)** 2107 are up-regulated in the patients. The DEGs were filtered using a cut off value of fold change (FC) < -1.5/>1.5, and false discovery rate (FDR) <0.05.



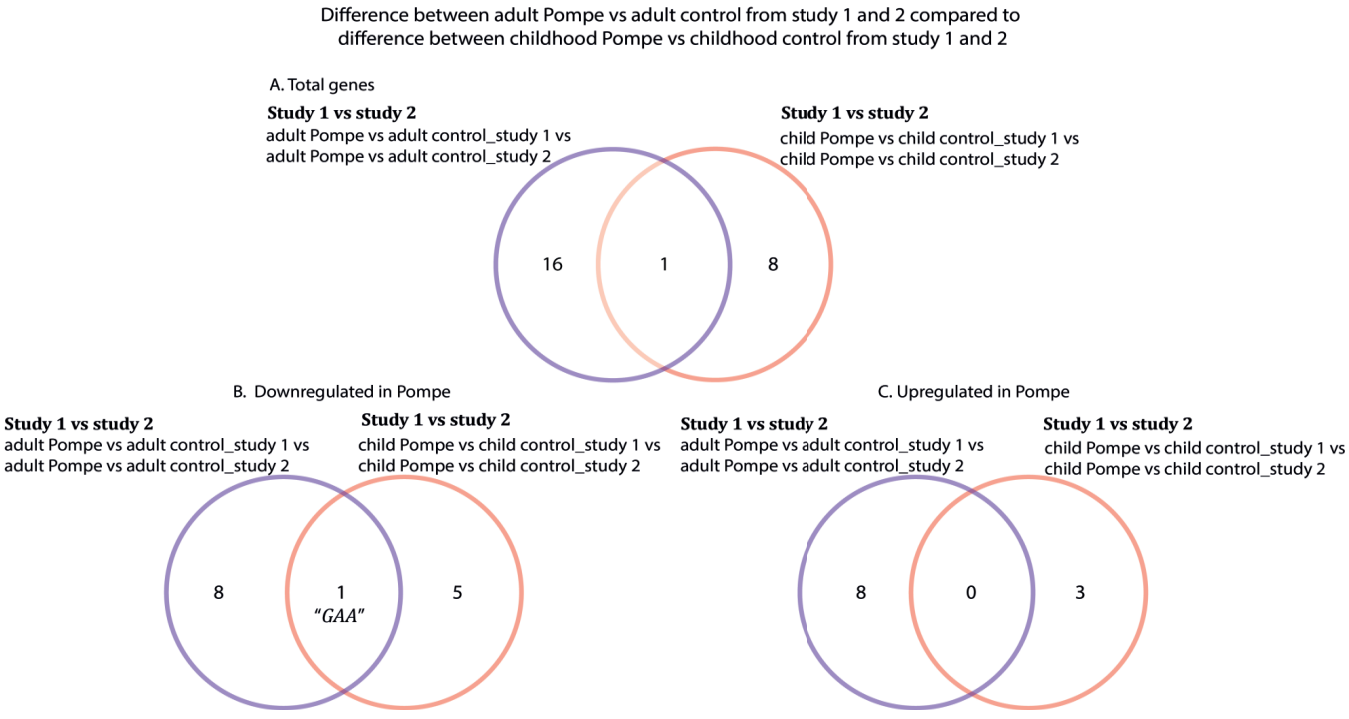
Supplementary Figure S8. Venn diagram showing the difference between affected adults and children compared to unaffected controls from study 2. (A) A total of 7 genes are differentially expressed in Pompe patients. **(B)** 5 genes are downregulated and **(C)** 2 are upregulated genes in the patients. The DEGs were filtered using a cut off value of fold change (FC) < -1.5/>1.5, and false discovery rate (FDR) <0.05.



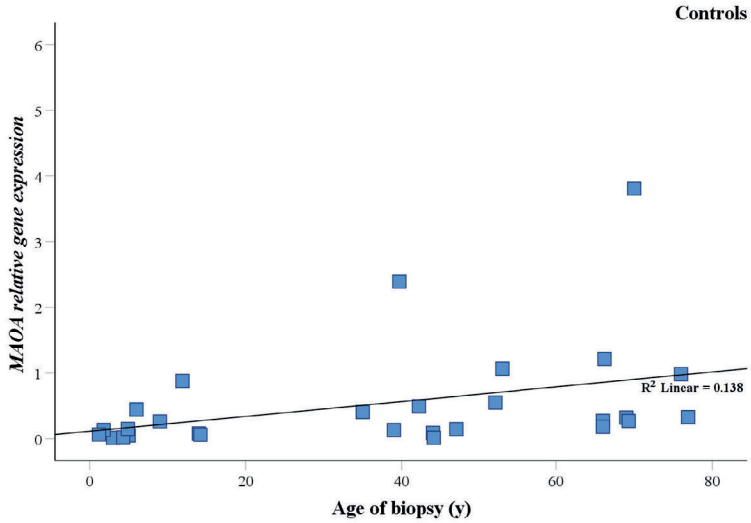
Supplementary Figure S9. Venn diagram showing the difference between affected adults and unaffected adult controls combining studies 1 and 2. (A) A total of 32 genes are differentially expressed in adult patients compared to adult controls **(B)** 9 genes are downregulated in the patients and **(C)** 8 are upregulated.



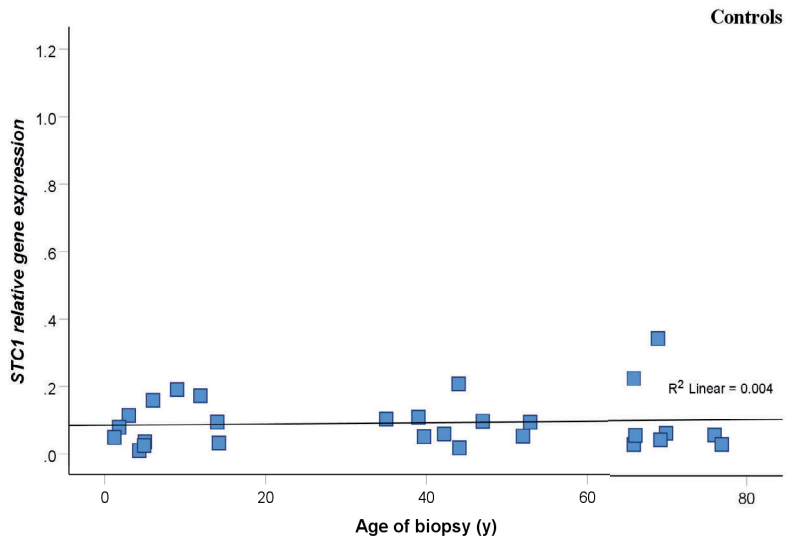
Supplementary Figure S10. Venn diagram showing the difference between affected children and unaffected childhood controls combining studies 1 and 2. (A) A total of 31 genes are differentially expressed in childhood patients compared to controls. **(B)** 6 genes are downregulated in the patients and **(C)** 3 are upregulated.



Supplementary Figure S11. Venn diagram showing the number of DEGs comparing both adult patients with adult controls as well and affected children with childhood controls combining the data sets from studies 1 and 2 (**A)** A total of 17 genes are differentially expressed in adult patients compared to controls, and 9 are differentially expressed in childhood patients compared to controls. **(B)** 9 are downregulated in adult onset Pompe disease and 6 in childhood onset disease. Notably, the single one gene downregulated in both adult as well as childhood patients, compared with age matched controls, is the *GAA* gene, which showed over two-fold decreased in Pompe patients. **(C)** 8 genes are upregulated in affected adults and 3 in affected children compared to age matched controls.



Supplementary Figure S12A. *MAOA* gene expression by age in controls of studies 1- 3. The expression level was measured by RT-qPCR. Samples from foreskin were not included. Linear regression analysis was performed using SPSS.



Supplementary Figure S12B. *STC1* gene expression by age in controls of studies 1- 3. The expression level was measured by RT-qPCR. Samples from foreskin were not included. Linear regression analysis was performed using SPSS.

CHAPTER 8

General Discussion

GENERAL DISCUSSION

Since the advancement of technologies for DNA analysis, the field of rare diseases has developed substantially. Inherited metabolic disorders represent a vast and heterogeneous collection of more than 1000 different rare genetic diseases [1]. Pompe disease is a metabolic disorder classified as glycogen storage disease type II, a lysosomal storage disorder. Enzyme replacement therapy (ERT) with alglucosidase alfa (Genzyme, Cambridge, USA) has been available since 2006 for patients with this disease. The therapy has improved the survival rate in classic infantile Pompe disease by several years, beyond the natural life expectancy [2], [3]. Additionally, patients with later-onset phenotypes have experienced improvement with respect to walking abilities, and stabilisation of respiratory function [4], [5].

POMPE DISEASE EPIDEMIOLOGY AND GEOGRAPHIC DISTRIBUTION

The availability of ERT for Pompe disease has stimulated the identification of new patients. Worldwide implementation of technologies and facilities for screening symptomatic patients by biochemical and molecular analysis has led to increasing numbers of diagnosed patients. Differences in prevalence among countries suggest that the number of cases is either underestimated and/or determined by genetic background. For example, Pompe disease has been reported to be less common in Australia (1:146.000) and Portugal (1:600.000) than in the Netherlands (1:40.000) [6]–[8]. The Australian study evaluated diagnosed cases over a period of 16 years (1980-1996) and the Portuguese study over 19 years (1982-2001). In the Dutch study, conducted in 1999, 3000 random blood spots of newborns were screened, and DNA analysis was used for the identification of the three most common variants, c.-32-13T>G (IVS1), c.525del and c.2481+102_2646+31del (delex 18). The populations in these three countries are all of European descent. However, the Dutch study identified more patients, suggesting that the applied method using screening of DNA variants brought the results closer to today's reality.

A reference study for screening Pompe patients is derived from Taiwan, in which newborn screening (NBS) officially started in 1985. The overall incidence of Pompe disease in that country is 1:17.000, with an incidence of 1:52.000 for classic infantile disease and 1:25.000 for late-onset forms [9], [10]. In Taiwan, NBS involves a two-step procedure: the measuring of GAA enzyme activity in dried blood spots (DBS) followed by *GAA* genotyping [11]. Recently, an NBS study performed in Pennsylvania, USA, screened 531,139 newborns, in which the overall incidence was 1:16.095. The screening algorithm was similar to the one

used in Taiwan [12]. Additionally, a recent study that analysed carrier frequency of disease-associated variants, based on information from several databases, e.g. gnomAD and Pompe variant database, predicted the genetic prevalence in seven different populations, whereby the prevalence in East Asians and non-Finnish Europeans was highly similar (approximately 1:14,000) [13]. Therefore, countries with different genetic backgrounds but with similar genetic prevalence's for Pompe disease, like Taiwan and the USA, point to the need for a similar approach in other countries, which may decrease the discrepancies among reported prevalence's.

THE ACTUAL SCENARIO FOR SCREENING POMPE PATIENTS

NBS is a public health initiative with the purpose to detect and identify affected newborns for which effective therapy is available [14]. The introduction of NBS in some countries, using DBS for first-tier testing, has contributed significantly to the early detection and treatment of Pompe disease patients [10], [11], [15]. Most studies performed fluorometry or tandem mass spectrometry (MS/MS) to analyse GAA enzyme activity in blood specimens using DBS as a sample [16]. However, NBS in Pompe disease using DBS has obstacles due to relatively high false-positive and false-negative rates, requiring a second assay for confirmation. **Chapter 2** showed that 16.7% of the patients came out with false-negative results using the DBS assay and needed another method to correct the diagnosis. Measuring a second (lysosomal) enzyme activity (e.g. BGAL) is strongly advised for judging the DBS quality. At our center, the DBS assay is not used routinely. A leukocyte-based assay is preferred, and fibroblasts are used for confirmation using 4MUG as a substrate. The latter is the most reliable assay for biochemical diagnosis. The assays using fibroblasts as sample source with 4MUG as an artificial substrate reveals distinct residual GAA enzyme activities between classic infantile, childhood, and adult forms of Pompe disease.

The use of leukocytes and fibroblasts is common practice in the Netherlands; however, these sample sources can be challenging to obtain in certain countries. Usually, the diagnosis of rare diseases such as Pompe disease is done in specialized centers, and the shipping of samples over long distances is a limitation in many countries. Taking a skin biopsy and the follow up culturing of fibroblasts is time-consuming, and culturing of the cells needs a strict procedure. Therefore, the advantage of using DBSs in diagnosis and newborn screening is that these samples can be transported safely over long distances, they can even be mailed in envelopes because of the stability of enzymes -when dried- for months at room temperature [17], [18]. In addition, the ability to block the interfering enzyme glucoamylase with acarbose makes the DBS assay a useful method for large scale

screening [19]. However, the issue of false-negative and false-positive outcomes, in part caused by pseudodeficiency alleles, is a drawback.

As concluded in **Chapter 2**, the analysis of a second lysosomal enzyme as a sample control in DBS increases the assay's reliability. Regarding false-positive cases, a study performed in Taiwan analysed a large series of patients' samples (all phenotypes included), carriers and carriers of pseudodeficiency alleles using a fluorometric assay. This method failed to distinguish between infantile and late-onset cases of Pompe disease, and carriers of a pseudoallele. However, when performing tandem mass spectrometry, the successful distinction between these individuals was claimed [20]. Nevertheless, mass spectrometry has high costs, requiring a significant budget for materials and equipment.

The most used assay for NBS worldwide is the fluorometric method using DBS but differences in the detection of Pompe patients are noticed even within a continent with a short geographic distance. For example, it can be expected to see slight genetic differences in European subpopulations that depend on the relative geographic distances, except for the Finnish population [21], [22]. However, screening of Danish, Slovenian and Belgian myopathic patients has not led to the identification of Pompe patients [23]–[25]. In the Danish study, enzyme activity was analyzed in DBSs from 47 patients, but the method applied was not clearly described, being either fluorometry or tandem mass spectrometry [23]. The Slovenian and Belgium studies analysed 90 and 100 patients, respectively, using the fluorometric assay in DBSs. In 5 Belgium patients, low enzyme activity was found, but *GAA* sequencing did not detect deleterious variants [24], [25]. A similar situation was reported in Brazil, where the proportion of cases that were diagnosed was much lower than the expected incidence at birth, suggesting that many cases could be underdiagnosed [26].

In summary, most screening programs so far relied on measuring the enzyme activity in DBS samples using a fluorometric assay as a first-tier test. We conclude that the implementation of a protocol that considers false-negatives and false-positives would decrease the differences among countries in the number of patients detected by NBS.

STRATEGIES TO IMPROVE THE EFFICIENCY OF ENZYMATIC AND MOLECULAR DIAGNOSIS OF POMPE DISEASE

The immediate concern in the diagnosis of Pompe disease is the early detection of classic infantile patients because the time spent to conclude the diagnostic outcome can negatively affect the patient's survival. In a patient with the classic infantile form, symptoms usually start shortly after birth and, without ERT, the median age of death is six months [27]. Taking Brazil as an example, the diagnosis of rare diseases can take months to years [28]; the same may apply to other countries. Several issues can influence the diagnostic process:

- General practitioners (GP) are not aware of rare diseases;
- The reference centers are usually located in big cities, forcing people to consult a local GP. In addition, the GP needs to refer the patient to the right specialist, which can take time. The travel from the city of the patient to the city with the reference center is also a limitation, especially for patients with a low income;
- After confirming the diagnosis, receiving ERT may require a judicial process [29].

For Pompe patients to receive ERT, it is required to be appropriately diagnosed by enzymatic and molecular analyses, and the identification of two disease-associated variants in the *GAA* is required in several countries [30]. As mentioned before, the enzymatic diagnosis can be a challenge due to potential false-negative or false-positive results. In many cases, DNA analysis will be performed after a positive enzymatic result. Routine DNA analysis usually focuses on coding regions using PCR and Sanger sequencing. Next-Generation sequencing (NGS) is becoming an alternative when disease-associated variants are not found. We reported in **Chapter 3** the case of a symptomatic patient with low enzyme activity but only one disease-associated variant in heterozygosity. By performing an SNP array on different tissues, a 39.5MB allele imbalance was detected that was more profound in fibroblasts compared to leukocytes. Additional analysis showed genetic mosaicism for a segment of Chr 17. This illustrates that extended analysis can resolve inconclusive cases, which can be crucial to start or continue ERT.

Extended analysis may need additional samples such as fibroblasts, and/or require RNA for cDNA synthesis. On the other hand, quick results are essential for the survival of classic infantile Pompe patients. Most of the samples available in many countries are DBS on filter paper or FTA® Cards. In my experience -when I was working on the molecular analysis of patients from Latin America in São Paulo, Brazil- patient's samples were mostly received on filter paper by post. After the Biochemical laboratory had taken a punch from the filter paper for enzymatic analysis, the rest of the paper was used to isolate DNA for the performance of PCR and Sanger sequencing in case the enzymatic analysis turned out to be positive. In a few cases, the sample quality was low, in which case we had to solicit for a new one. It was challenging to analyse gross deletions and insertions using samples on filter paper because the quality of the DNA might be too low for performing fragment analysis like Multiplex probe amplification (MLPA). Some laboratories already developed technologies to improve the filter paper quality. For instance, by using CentoCard®, it is possible to perform both enzymatic as well as DNA analysis (including MLPA, exome sequencing, genome sequencing and biomarker assays).

Due to the limited sample source and shipping opportunities, efficient use of filter paper samples needs to be implemented. Here, I argue for the implementation of NGS as a second-tier assay in NBS, in which delay and misdiagnosis can be avoided [31]–[33].

NGS technologies include gene panels, such as Motorplex, whereby panels can

be pre-designed to screen many different genes [34], to perform whole exome sequencing (WES) and whole genome sequencing (WGS). WES has the advantage to screen for rare variants, however, intronic variants will be missed. WGS is an alternative to screen the entire gene, although the chance to detect variants with a minor allele frequency below 5% decreases compared to WES. We have shown in **Chapter 4** that the number of listed *GAA* variations in the Pompe variant database has increased by 61% in a period of 4 years. More registries are still appearing because of the increased frequency of NBS for Pompe disease and the use of NGS. The compilation of data in the Pompe variant database is a useful tool to guide medical doctors in how to diagnose and manage Pompe patients, as it links disease-associated variants to clinical severities [35].

The treatment with ERT can be influenced by antibodies to the recombinant GAA enzyme. For this reason, cross-reactive immunological material (CRIM) negative patients receive immunomodulation therapy prior to initiation of ERT. In many instances, the *GAA* genotype can predict the CRIM status pointing to the importance of the Pompe variant database and regular updates (**Chapter 4**).

NGS can generate a huge collection of uninterpretable variants, known as variants of unknown significance (VUS). In **Chapter 4** we discussed the issue of VUS in putative patients diagnosed by NBS. VUS can lead to underestimation (disease-associated variants interpreted as benign) and overestimation (benign variants are misinterpreted as disease-associated) of Pompe disease. Another issue that became apparent is that asymptomatic individuals with a VUS are identified by NBS with the possibility of an emerging disease and doubt on when starting ERT. Such situation leads to emotional stress [36], [37]. Therefore, an up-to-date database, the proper training of GPs, and awareness of rare diseases would optimize the diagnostic process. The use of NGS as a routine procedure combined with NBS has the potential to prioritize patients at high risk and ensure adequate genetic counselling of parents.

A FUTURE PERSPECTIVE ON THE USE OF NGS FOR SCREENING POMPE PATIENTS

We are entering the 'big data' era. For example, genomic data from NBS/NGS-based patient screening programs are becoming available. Predictions for 2030 see a significant increase in human genomics based on NGS since the human genome project was released in 2003 [38]. This includes: the association of genomic and phenotypic information; the prediction of pathogenicity of genomic variants, rendering the designation VUS obsolete; and access to our own genomic data via our smartphone [38]. Prices of WES and WGS analyses are decreasing every year. Although prices can vary substantially per laboratory depending

on whether bioinformatics' analysis is included, high coverage analysis can be performed for a total of 1000 euros per sample. Analysis using artificial intelligence (AI) will decrease the price even more, by using an AI algorithm known as deep learning [39]. Recent studies suggest how deep learning can revolutionize variant calling for nanopore-based sequencing technologies, a sequencing technology that enables real-time analysis of long DNA and RNA fragments in a short time [40]. Genotype-phenotype prediction provides diagnoses and forecasts future disease risks. The AI-based methods can accurately aid in the diagnosis and prediction of disease risks by integrating data. Therefore, for Pompe disease, in the future integrating AI with the Pompe variant database (www.pompevariantdatabase.nl) (integrated into the LOVD) may efficiently predict the genotype-phenotype outcome highlighting the importance to keep the existing databases updated, as presented in **Chapter 4**. Additionally, the use of NGS as a first-tier test seems a future possibility, turning the diagnostic and screening processes into a homogenous and inclusive procedure worldwide.

HOW TO DEAL WITH UNKNOWN MOLECULAR RESULTS?

As stated above, to properly diagnose a Pompe patient the outcomes of the enzymatic analysis and the DNA analysis need to be consistent and compatible with the clinical diagnosis of Pompe disease. In certain countries, these results need to be available before the judicialization process can start that is required for a patient to receive ERT. Depending on the country, the costs of ERT can be covered by the health insurance system or the government [41], [42]. In **Chapter 2**, we discussed how to manage and improve the enzymatic analysis. However, how to handle in putative case of Pompe disease wherein no variants are found by Sanger sequencing?

The main concern is putative cases of classic infantile Pompe disease, in which the survival rate is proportional to an early start of ERT. In most cases, the only available sample will be a DBS. In **Figure 1**, we depict two possible scenarios when DBS is the only available sample. In scenario 1, after the laboratory has received the sample, enzymatic and DNA analysis can be performed simultaneously. The first-tier analysis of the enzymatic analysis shows low enzyme activity, and with DNA analysis, two disease-associated variants are found, of which the (predicted) CRIM status is known. ERT as well as the subsequent judicialization process can be initiated. In scenario 2, low enzyme activity is found in DBS in the first/second tier; however, Sanger sequencing of the flanking exons does not reveal a disease-associated variant. If a VUS is found in the coding or flanking region, an *in-silico* analysis can be performed and used to predict its potential impact. If a VUS is not found but when the clinical and enzymatic diagnostic outcomes are indisputable, ERT is recommended, the judicialization process is started.

In many cases, molecular results are still required to continue treatment with ERT.

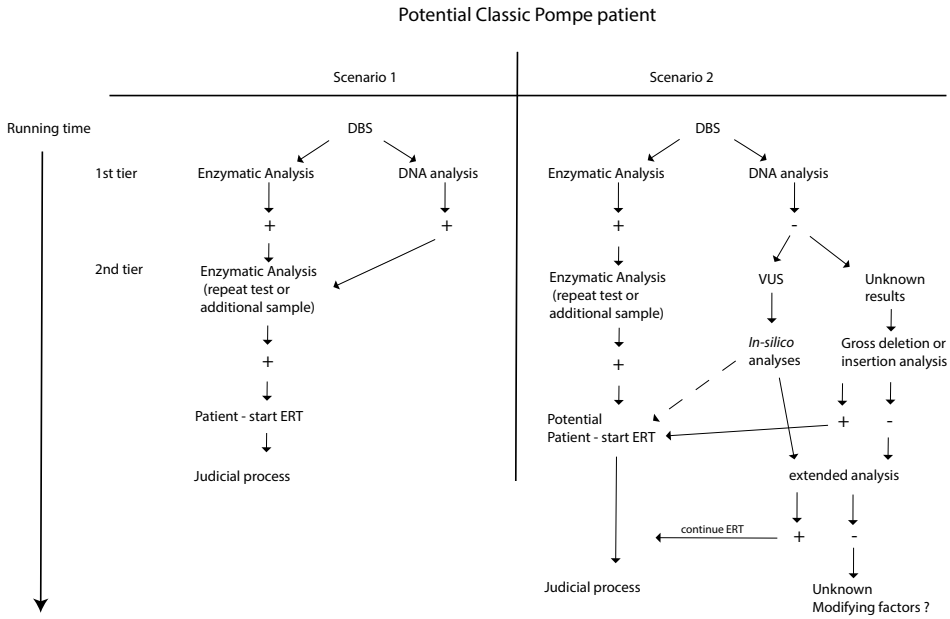


Figure 1. Two possible scenarios when DBS (Dried blood spot) is the only sample available for a potential classic infantile Pompe patient. (+) Results compatible with the diagnosis Pompe disease; (-) Results are insufficient/negative to diagnose Pompe disease; (dashed line) *in-silico* analyses for the prediction of pathogenicity and CRIM status. VUS (Variant of unknown significance); ERT (Enzyme replacement therapy).

When no variants are found, the next step is to search for large deletions or insertions using, for example, MLPA, quantitative qPCR or genome sequencing. Some laboratories have already developed techniques to perform this analysis in DBS [43]-[47]. If the results remain inconclusive, extended analyses and further investigations are recommended. **Chapter 3** describes six patients in which the molecular diagnosis was successfully performed using a generic-splicing assay, minigene analysis, SNP array analysis, and targeted Sanger sequencing; however, for these analyses, additional samples such as fibroblasts were required.

MODIFYING FACTORS IN POMPE DISEASE

Modifying factors may have an impact on the clinical course of disease severity. For example, the variant c.510C>T (p.=), was disregarded for many years because it is silent, i.e. it does not change the amino acid. However, recently our group found that in compound heterozygous patients - with IVS1 on one allele and another disease-associated variant on the other allele - c.510C>T can co-occur on the same allele as IVS1 and aggravate the

extent of aberrant splicing caused by the IVS1 variant. The combination c.510C>T plus IVS1 was demonstrated to be associated with an early presentation of symptoms [48]. However, it is still poorly understood why some patients without c.510C>T manifest symptoms much earlier than others, even within families.

Several publications reported the broad spectrum of symptom onset in patients with the IVS1 variant [48]-[51]. In **Chapter 5**, we analysed the information included in the most recent update of the Pompe variant database and found that - besides the IVS1 variant - nine other disease-associated variants (c.2647-7G>A; c.546G>T; c. 2238G>C; c.2173C>T; c.1076-22T>G; c.1309C>T; c.-32-3C>A; c.1875C>G; c.1064T>C) are associated with a broad spectrum of disease onset. This also counted for variants at homozygous state. The c.510C>T variant is linked to the IVS1 allele *in cis* and was shown to be a modifier that accelerates the age at symptom onset. However, in the other nine alleles, the variant c.510C>T was not found, suggesting an influence of additional modifying factors in disease progression.

Environmental factors, epigenetic factors, genetic background, and even gut microbiota can in theory be involved in the genotype-phenotype relationship. We can speculate that factors involved may include differences in muscle fiber types, or up versus downregulation of genes involved in relevant pathways such as glucose metabolism, 'lysosomal biology', or muscle function; all areas on which future research can be focused. *GAA* haplotypes can act at another level as modifying factor. A haplotype is an uninterrupted series of DNA variations that tend to be inherited together without recombination. A DNA locus that correlates with a quantitative trait variation in the phenotype is named a quantitative trait locus (QTL). Furthermore, a DNA locus or a haplotype that influences expression levels of mRNA transcripts is known as eQTLs.

It is already known that eQTLs acting *in cis* or *in trans* can influence the expression of genes across human tissues [52]-[54] (**Figure 2**). A large project has been to collect samples from almost 1000 *postmortem* individuals to understand which *cis*-eQTLs can

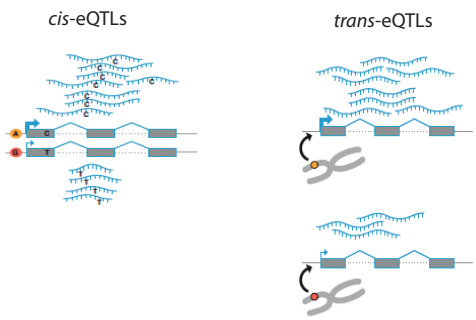


Figure 2. Local and distal genetic effects (*cis*-eQTLs and *trans*-eQTLs, respectively) influencing gene expression (Modified from [58]).

affect gene expression across 54 different non-diseased tissues. This effort is called the Genotype-Tissue Expression (GTEx) project (www.gtexportal.org) and aims to characterise genetic effects on the human transcriptome and link regulatory mechanisms to trait and disease associations. Interestingly, when we analysed genes involved in glycogen metabolism, we found that the *GAA* is particularly enriched in *cis*-eQTLs compared to other genes (**Figure 3**). This could explain the broad spectrum of GAA activity found in healthy individuals (**Chapter 2**). By analogy, the broad range in age of onset among patients with Pompe disease and the same *GAA* genotype could be explained by *cis*-eQTLs in the *GAA*.

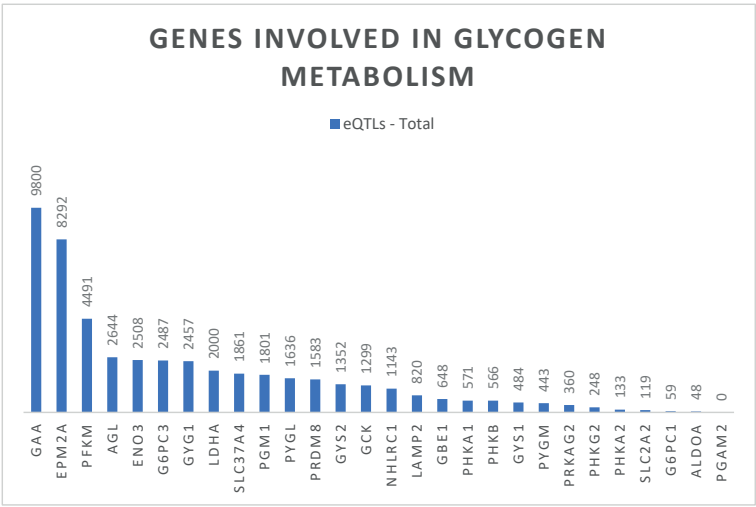


Figure 3. Number of *cis*-eQTLs present in genes involved in glycogen metabolism, based on 54 different tissues. Source: GTEx (www.gtexportal.org , analysis on 27-03-2021).

eQTLs AND HAPLOTYPES INFLUENCE IN GAA EXPRESSION

In **Chapter 6**, we defined the *GAA* haplotypes in a general population with Northern and Western European genetic ancestry. In short, we found two major haplotypes, one with many common variants (MAF ≥ 5%) and another without those variants. IVS1 is present in a haplotype without common variants. Therefore, special attention should be given to rare variants that are present *in cis* with IVS1 as these could be modifiers, like we found for c.510C>T. In non-IVS1 alleles, the common variants are *cis*-eQTLs registered in GTEX, and they could account for the variability of GAA activity because all those common variants are downregulating GAA expression.

Applying the same effort to another population such as the Asian population, we expect to find different haplotypes. Labrousse *et al*. 2010 found that the Asian

pseudodeficiency harbours distinct haplotypes for Pompe patients and unaffected individuals. In the Taiwanese population, a haplotype that negatively modulates GAA expression was found in 14% of individuals [55]. A more in-depth study of the Asian haplotypes and their possible effect on GAA expression is warranted.

The IVS1 variant was hypothesized to propagate as a founder variant through inheritance and migration [50]. A founder variant usually arises in a tribe or group of people that are geographically or culturally isolated. However, the IVS1 variant is widespread in people with a European genetic background. This argues against a founder effect as the single mechanism, as it would imply that all Europeans are derived from a single founder in which IVS1 was present, which is unlikely. Interestingly, the IVS1 variant was found in a haplotype with a neutral effect on GAA expression, whereas the great majority of variants in common haplotypes had a negative impact on GAA expression. In general, disease-associated variants are often found in haplotypes that negatively affect the expression of the gene in question. This co-occurrence increases the chance that the haplotype incorporating the pathogenic variant will be depleted - in the long run - by natural purification selection [56]. This phenomenon was indeed seen for most other disease-associated variants for Pompe disease (**Chapter 6**).

The IVS1 variant has a higher MAF (European non-Finnish MAF-0.5%) than other disease-associated variants. Therefore, we speculate that the IVS1 allele could have had a beneficial influence during human evolution, being subject to positive selection. The IVS1 allele causes leaky splicing that decreases GAA activity. We speculate that during human evolution and in episodes of starvation, the presence of IVS1 has been advantageous. The presence of IVS1 results in decreased degradation of lysosomal glycogen, thereby preserving a glycogen source for the cell. A similar selective advantage triggered by 'protective haplotypes' has been seen in genes related to type 2 diabetes [57]. The protective haplotype dating from the beginning of the Neolithic era did arise by purifying selection in response to specific pastoral dietary habits in individuals from Bishkek, Kyrgyzstan and Bukhara, Uzbekistan [57]. Thus, haplotypes that had a selective advantage in past environments may have become deleterious in the modern world due to changes in diet, behaviour, and/or external stimuli, which could explain the high prevalence of some polygenic diseases such as diabetes. One may hypothesis that a similar selective procedure has operated for the IVS1 variant in Pompe disease, however, its low frequency limits further investigation. Efforts to collect data from more IVS1 patients – for instance via the European Pompe Consortium - would bring opportunities to perform such analysis and to obtain solid evidence.

CONCLUSIONS

Despite progress in the last 15 years with collecting data from patients and their disease-associated variants in the *GAA*, new information is still to come. Patients from certain geographic areas are underrepresented in our Pompe variant database, resulting in a lack of important information for investigating genotype-phenotype relationships. Advances in NGS technology could enable it to become a first-tier test in NBS, making the diagnosis for Pompe disease inclusive in many countries. There is increasing evidence for phenotypic diversity among patients with the same *GAA* genotype. Investigating larger patient cohorts will facilitate the search for modifiers, either using a candidate or an unbiased approach in genome-wide association analysis and help to understanding how eQTLs in haplotypes can influence GAA expression in healthy subjects and in Pompe patients. Further research is required to find out whether the pathogenic IVS1 variant is inserted in a protective haplotype that has been preserved during evolution, driven by positive selection.

REFERENCES

[1] C. M. Mak, H. C. H. Lee, A. Y. W. Chan, and C. W. Lam, "Inborn errors of metabolism and expanded newborn screening: Review and update," Critical Reviews in Clinical Laboratory Sciences, vol. 50, no. 6. Taylor & Francis, pp. 142–162, Nov. 2013, doi: 10.3109/10408363.2013.847896.
[2] A. A. Khan et al., "Higher dosing of alglucosidase alfa improves outcomes in children with Pompe disease: a clinical study and review of the literature," Genet. Med., vol. 22, no. 5, pp. 898–907, May 2020, doi: 10.1038/s41436-019-0738-0.
[3] E. Poelman et al., "Effects of higher and more frequent dosing of alglucosidase alfa and immunomodulation on long-term clinical outcome of classic infantile Pompe patients," J. Inherit. Metab. Dis., 2020, doi: 10.1002/jimd.12268.
[4] A. van der Ploeg et al., "Prospective exploratory muscle biopsy, imaging, and functional assessment in patients with late-onset Pompe disease treated with alglucosidase alfa: The EMBASSY Study," Mol. Genet. Metab., vol. 119, no. 1–2, pp. 115–123, Sep. 2016, doi: 10.1016/j.ymgme.2016.05.013.
[5] C. Semplicini et al., "Long-term benefit of enzyme replacement therapy with alglucosidase alfa in adults with Pompe disease: Prospective analysis from the French Pompe Registry," J. Inherit. Metab. Dis., vol. 43, no. 6, 2020, doi: 10.1002/jimd.12272.
[6] P. J. Meikle, J. J. Hopwood, A. E. Clague, and W. F. Carey, "Prevalence of lysosomal storage disorders," J. Am. Med. Assoc., vol. 281, no. 3, pp. 249–254, Jan. 1999, doi: 10.1001/jama.281.3.249.
[7] R. Pinto et al., "Prevalence of lysosomal storage diseases in Portugal," Eur. J. Hum. Genet., vol. 12, no. 2, pp. 87–92, Feb. 2004, doi: 10.1038/sj.ejhg.5201044.
[8] M. G. E. M. Ausems et al., "Frequency of glycogen storage disease type II in The Netherlands: Implications for diagnosis and genetic counselling," Eur. J. Hum. Genet., vol. 7, no. 6, pp. 713–716, 1999, doi: 10.1038/sj.ejhg.5200367.
[9] S. C. Chiang, W. L. Hwu, N. C. Lee, L. W. Hsu, and Y. H. Chien, "Algorithm for Pompe disease newborn screening: Results from the Taiwan screening program," Mol. Genet. Metab., vol. 106, no. 3, pp. 281–286, Jul. 2012, doi: 10.1016/j.ymgme.2012.04.013.
[10] Y. H. Chien et al., "Long-term prognosis of patients with infantile-onset pompe disease diagnosed by newborn screening and treated since birth," J. Pediatr., vol. 166, no. 4, pp. 985-991.e2, 2015, doi: 10.1016/j.jpeds.2014.10.068.
[11] Y.-H. Chien, W.-L. Hwu, and N.-C. Lee, "Newborn screening: Taiwanese experience," Ann.

Transl. Med., vol. 7, no. 13, pp. 281-281, Jul. 2019, doi: 10.21037/atm.2019.05.47.
[12] C. Ficicioglu et al., "Newborn Screening for Pompe Disease: Pennsylvania Experience," Int. J. Neonatal Screen., vol. 6, no. 4, p. 89, Nov. 2020, doi: 10.3390/ijns6040089.
[13] K. S. Park, "Carrier frequency and predicted genetic prevalence of Pompe disease based on a general population database," Mol. Genet. Metab. Reports, vol. 27, p. 100734, Jun. 2021, doi: 10.1016/j.ymgmr.2021.100734.
[14] O. A. Bodamer, C. Ronald Scott, and R. Giugliani, "Newborn screening for pompe disease," Pediatrics, vol. 140, no. Suppl 1. American Academy of Pediatrics, Jul. 01, 2017, doi: 10.1542/peds.2016-0280C.
[15] P. V. Hopkins, C. Campbell, T. Klug, S. Rogers, J. Raburn-Miller, and J. Kiesling, "Lysosomal storage disorder screening implementation: Findings from the first six months of full population pilot testing in Missouri," J. Pediatr., vol. 166, no. 1, pp. 172-177, Jan. 2015, doi: 10.1016/j.jpeds.2014.09.023.
[16] M. H. Gelb, Z. Lukacs, E. Ranieri, and P. C. J. I. Schielen, "Newborn screening for lysosomal storage disorders: Methodologies for measurement of enzymatic activities in dried blood spots," International Journal of Neonatal Screening, vol. 5, no. 1. MDPI Multidisciplinary Digital Publishing Institute, 2019, doi: 10.3390/ijns5010001.
[17] N. A. Chamoles, G. Niizawa, M. Blanco, D. Gaggioli, and C. Casentini, "Glycogen storage disease type II: Enzymatic screening in dried blood spots on filter paper," Clin. Chim. Acta, vol. 347, no. 1-2, pp. 97-102, Sep. 2004, doi: 10.1016/j.cccn.2004.04.009.
[18] K. B. Müller, M. D. B. Rodrigues, V. G. Pereira, A. M. Martins, and V. D'Almeida, Reference values for lysosomal enzymes activities using dried blood spots samples - A Brazilian experience," Diagn. Pathol., vol. 5, no. 1, p. 65, Aug. 2010, doi: 10.1186/1746-1596-5-65.
[19] H. Zhang et al., "Comparison of maltose and acarbose as inhibitors of maltase-glucoamylase activity in assaying acid α-glucosidase activity in dried blood spots for the diagnosis of infantile Pompe disease," Genet. Med., vol. 8, no. 5, pp. 302-306, May 2006, doi: 10.1097/01.gim.0000217781.66786.9b.
[20] H. C. Liao et al., "Mass spectrometry but not fluorimetry distinguishes affected and pseudodeficiency patients in newborn screening for pompe disease," Clin. Chem., vol. 63, no. 7, pp. 1271-1277, Jul. 2017, doi: 10.1373/clinchem.2016.269027.
[21] O. Lao et al., "Correlation between Genetic and Geographic Structure in Europe," Curr. Biol., vol. 18, no. 16, pp. 1241-1248, Aug. 2008, doi: 10.1016/j.cub.2008.07.049.
[22] K. J. Karczewski et al., "The mutational constraint spectrum quantified from variation in 141,456 humans," Nature, vol. 581, no. 7809, pp. 434-443, May 2020, doi: 10.1038/s41586-020-2308-7.
[23] J. S. Hansen et al., "Screening for late-onset Pompe disease in western Denmark," Acta Neurol. Scand., vol. 137, no. 1, pp. 85-90, Jan. 2018, doi: 10.1111/ane.12811.
[24] M. Meznaric, K. Fumic, and L. Leonardis, "Selective screening of late-onset Pompe disease (LOPD) in patients with non-diagnostic muscle biopsies," J. Clin. Pathol., vol. 72, no. 7, pp. 468-472, Jul. 2019, doi: 10.1136/jclinpath-2018-205446.
[25] G. Remiche, Z. Lukacs, D. C. Kasper, M. Abramowicz, and M. Pandolfo, "Low prevalence estimates of late-onset glycogen storage disease type II in French-speaking Belgium are not due to missed diagnoses," J. Neuromuscul. Dis., vol. 5, no. 4, pp. 471-480, 2018, doi: 10.3233/JND-180336.
[26] R. Giugliani, A. Federhen, K. Michelin-Tirelli, M. Riegel, and M. Burin, "Relative frequency and estimated minimal frequency of lysosomal storage diseases in Brazil: Report from a reference laboratory," Genet. Mol. Biol., vol. 40, no. 1, pp. 31-39, Jan. 2017, doi: 10.1590/1678-4685-GMB-2016-0268.
[27] H. M. P. Van den Hout et al., "The natural course of infantile Pompe's disease: 20 Original cases compared with 133 cases from the literature," Pediatrics, vol. 112, no. 2 I, pp. 332-340, Aug. 2003, doi: 10.1542/peds.112.2.332.
[28] J. A. B. Iriart, M. F. Nucci, T. P. Muniz, G. B. Viana, W. de A. Aureliano, and S. Gibbon, "From the search for diagnosis to treatment uncertainties: Challenges of care for rare genetic diseases in Brazil," Cienc. e Saude Coletiva, vol. 24, no. 10, pp. 3637-3650, Oct. 2019, doi: 10.1590/1413-812320182410.01612019.
[29] L. C. Pereira et al., "Challenges in the Regulation of High-Cost Treatments—An Overview From Brazil," Value Heal. Reg. Issues, vol. 20, pp. 191-195, Dec. 2019, doi: 10.1016/j.vhri.2019.07.006.
[30] A. T. van der Ploeg et al., "European consensus for starting and stopping enzyme replacement therapy in adult patients with Pompe disease: a 10-year experience," Eur. J. Neurol., vol. 24, no. 6, pp. 768-e31, Jun. 2017, doi: 10.1111/ene.13285.

[31] K. Johnson et al., “Identification of GAA variants through whole exome sequencing targeted to a cohort of 606 patients with unexplained limb-girdle muscle weakness,” Orphanet J. Rare Dis., vol. 12, no. 1, Nov. 2017, doi: 10.1186/s13023-017-0722-1.
[32] L. D. Smith, M. N. Bainbridge, R. B. Parad, and A. Bhattacharjee, “Second tier molecular genetic testing in newborn screening for pompe disease: Landscape and challenges,” International Journal of Neonatal Screening, vol. 6, no. 2. MDPI Multidisciplinary Digital Publishing Institute, Jun. 01, 2020, doi: 10.3390/ijns6020032.
[33] E. G. Ames, R. Fisher, M. Kleyn, and A. Ahmad, “Current Practices for U.S. Newborn Screening of Pompe Disease and MPSI,” Int. J. Neonatal Screen., vol. 6, no. 3, p. 72, Sep. 2020, doi: 10.3390/ijns6030072.
[34] M. Savarese et al., “MotorPlex provides accurate variant detection across large muscle genes both in single myopathic patients and in pools of DNA samples,” Acta Neuropathol. Commun., vol. 2, no. 1, p. 100, 2014, doi: 10.1186/preaccept-1020679654138560.
[35] M. Y. Niño et al., “Extension of the Pompe mutation database by linking disease-associated variants to clinical severity,” Hum. Mutat., vol. 40, no. 11, pp. 1954–1967, Nov. 2019, doi: 10.1002/humu.23854.
[36] D. R. Morrison and E. W. Clayton, “False positive newborn screening results are not always benign,” Public Health Genomics, vol. 14, no. 3, pp. 173–177, May 2011, doi: 10.1159/000322527.
[37] C. G. van El, T. Rigter, A. J. J. Reuser, A. T. van der Ploeg, S. S. Weinreich, and M. C. Cornel, “Newborn screening for pompe disease? A qualitative study exploring professional views,” BMC Pediatr., vol. 14, no. 1, pp. 1–10, Aug. 2014, doi: 10.1186/1471-2431-14-203.
[38] E. D. Green et al., “Strategic vision for improving human health at The Forefront of Genomics,” Nature, vol. 586, no. 7831, pp. 683–692, Oct. 2020, doi: 10.1038/s41586-020-2817-4.
[39] R. Dias and A. Torkamani, “Artificial intelligence in clinical and genomic diagnostics,” Genome Medicine, vol. 11, no. 1. BioMed Central Ltd., Nov. 19, 2019, doi: 10.1186/s13073-019-0689-8.
[40] R. R. Wick, L. M. Judd, and K. E. Holt, “Performance of neural network basecalling tools for Oxford Nanopore sequencing,” Genome Biol., vol. 20, no. 1, Jun. 2019, doi: 10.1186/s13059-019-1727-y.
[41] A. Detiček, I. Locatelli, and M. Kos, “Patient Access to Medicines for Rare Diseases in European Countries,” Value Heal., vol. 21, no. 5, pp. 553–560, May 2018, doi: 10.1016/j.jval.2018.01.007.
[42] P. I. M. C. D’Ippolito and C. A. G. Gadelha, “O tratamento de doenças raras no Brasil: a judicialização e o Complexo Econômico-Industrial da Saúde,” Saúde em Debate, vol. 43, no. spe4, pp. 219–231, 2019, doi: 10.1590/0103-11042019s418.
[43] E. H. Choi, S. K. Lee, C. Ihm, and Y. H. Sohn, “Rapid DNA Extraction from Dried Blood Spots on Filter Paper: Potential Applications in Biobanking,” Osong Public Heal. Res. Perspect., vol. 5, no. 6, pp. 351–357, Dec. 2014, doi: 10.1016/j.phrp.2014.09.005.
[44] K. Malsagova, A. Kopylov, A. Stepanov, T. Butkova, A. Izotov, and A. Kaysheva, “Dried blood spot in laboratory: Directions and prospects,” Diagnostics, vol. 10, no. 4. MDPI AG, Apr. 01, 2020, doi: 10.3390/diagnostics10040248.
[45] Y. O. S. Wijaya et al., “Assessment of Spinal Muscular Atrophy Carrier Status by Determining SMN1 Copy Number Using Dried Blood Spots.,” Int. J. neonatal Screen., vol. 6, no. 2, p. 43, Jun. 2020, doi: 10.3390/ijns6020043.
[46] L. Czibere et al., “High-throughput genetic newborn screening for spinal muscular atrophy by rapid nucleic acid extraction from dried blood spots and 384-well qPCR,” Eur. J. Hum. Genet., vol. 28, no. 1, pp. 23–30, Jan. 2020, doi: 10.1038/s41431-019-0476-4.
[47] A. M. Bertoli-Avella et al., “Successful application of genome sequencing in a diagnostic setting: 1007 index cases from a clinically heterogeneous cohort,” Eur. J. Hum. Genet., vol. 29, no. 1, pp. 141–153, Jan. 2021, doi: 10.1038/s41431-020-00713-9.
[48] A. J. Bergsma et al., “A genetic modifier of symptom onset in Pompe disease,” EBioMedicine, vol. 43, pp. 553–561, May 2019, doi: 10.1016/j.ebiom.2019.03.048.
[49] M. Kroos, M. Hoogeveen-Westerveld, A. van der Ploeg, and A. J. J. Reuser, “The genotype-phenotype correlation in Pompe disease,” Am. J. Med. Genet. Part C Semin. Med. Genet., vol. 160 C, no. 1, pp. 59–68, Feb. 2012, doi: 10.1002/ajmg.c.31318.
[50] M. A. Kroos et al., “Broad spectrum of Pompe disease in patients with the same c.-32-13T→G haplotype,” Neurology, vol. 68, no. 2, pp. 110–115, Jan. 2007, doi: 10.1212/01.wnl.0000252798.25690.76.
[51] S. C. A. Wens et al., “Phenotypical variation within 22 families with Pompe disease,” Orphanet J. Rare Dis., vol. 8, no. 1, 2013, doi: 10.1186/1750-1172-8-182.

[52] H. J. Westra and L. Franke, “From genome to function by studying eQTLs,” Biochimica et Biophysica Acta - Molecular Basis of Disease, vol. 1842, no. 10. Elsevier B.V., pp. 1896–1902, Oct. 01, 2014, doi: 10.1016/j.bbadis.2014.04.024.
[53] F. Aguet et al., “The GTEx Consortium atlas of genetic regulatory effects across human tissues,” Science (80-.)., vol. 369, no. 6509, pp. 1318–1330, Sep. 2020, doi: 10.1126/SCIENCE.AAZ1776.
[54] S. Kim-Hellmuth et al., “Cell type-specific genetic regulation of gene expression across human tissues,” Science (80-.)., vol. 369, no. 6509, Sep. 2020, doi: 10.1126/SCIENCE.AAZ8528.
[55] P. Labrousse et al., “Genetic heterozygosity and pseudodeficiency in the Pompe disease newborn screening pilot program,” Mol. Genet. Metab., vol. 99, no. 4, pp. 379–383, Apr. 2010, doi: 10.1016/j.ymgme.2009.12.014.
[56] S. E. Castel et al., “Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk,” Nat. Genet., vol. 50, no. 9, pp. 1327–1334, Sep. 2018, doi: 10.1038/s41588-018-0192-y.
[57] L. Ségurel et al., “Positive selection of protective variants for type 2 diabetes from the Neolithic onward: A case study in Central Asia,” Eur. J. Hum. Genet., vol. 21, no. 10, pp. 1146–1151, Oct. 2013, doi: 10.1038/ejhg.2012.295.

APPENDIX

SUMMARY

Pompe disease is a rare, autosomal recessive disorder caused by disease-associated variants (mutations) in the gene coding for the enzyme acid-α-glucosidase (GAA). The function of GAA is to degrade glycogen to glucose in the lysosome. Deficiency of GAA will result in glycogen accumulation in the lysosome causing damage to the cells and tissues, leading to loss of function. Complete lack of GAA enzyme activity (0%) causes classic infantile Pompe disease, affecting newborns, mainly because of cardiac hypertrophy and skeletal muscle wasting. Childhood/adult Pompe patients have a milder phenotype because they have a residual enzyme activity (1-20%), leading to progressive muscle weakness, mobility problems, and respiratory insufficiency.

Since 2006 a treatment is available for Pompe disease: enzyme replacement therapy (ERT), which consists of intravenously administered human recombinant GAA enzyme. The therapy has improved the survival rate for classic infantile patients and showed improvements such as better walking abilities and stabilisation of respiratory function in children and adult patients. For Pompe patients to receive ERT, an appropriate diagnosis is required based on GAA enzyme deficiency, and preferably two disease-associated variants should be detected in the *GAA* gene. A general introduction to Pompe disease and an explanation of the process to perform enzymatic and molecular diagnosis is provided in **Chapter 1**.

Several methods and samples can be used to determine the enzyme activity for GAA. However, enzymatic diagnosis can face some issues such as the existence of false-positive and false-negative results. **Chapter 2** reports methods applied for the enzymatic diagnosis of Pompe disease and reviewed the outcome of a series of 2591 assays in 1709 individuals suspected of having Pompe disease. It shows the results of different types of samples and different methods that were used over the past 28 years in the department of Clinical Genetics of the Erasmus MC University Medical Center in Rotterdam. In order to determine the performance of each assay, GAA activity was measured in white blood cells (leukocytes) using glycogen or 4-methylumbelliferyl-α-D-glucopyranoside (4MUG) as substrate, dried blood spots (DBS) using 4MUG as substrate, and skin fibroblast using 4MUG as substrate. Mixed leukocytes provided the fastest diagnostic method, whereby the reaction mixture needed to contain acarbose to inhibit glucoamylase. DBS-based assays were considered valuable tools for screening, but an additional assay is required to confirm the diagnosis to avoid false-negative results. Cultured skin fibroblasts provided the best results to differentiate between classic infantile, childhood, and adult Pompe disease forms. Glycogen and 4MUG were suitable substrates, but glycogen discriminated best between affected and unaffected individuals. The disadvantage of using glycogen is that carriers of the GAA2 pseudo deficiency may be misdiagnosed as patients. In such cases, additional

use of 4MUG provided the most reliable assay and validated results from leukocytes and/or DBS.

Molecular analysis is also an essential step to conclude the diagnostic procedure of a Pompe patient. Traditional Sanger sequencing covering the coding exon flanking areas is the most commonly used method, however, in some cases, this method will not be enough to conclude a case. In **Chapter 3** we showed extended molecular diagnostic analyses and identified novel disease-associated variants in patients with Pompe disease where conventional analysis was insufficient. Additional assays, such as a generic-splicing assay, minigene analysis, SNP array analysis, and targeted Sanger sequencing, allowed the identification of an exonic deletion, a promoter deletion, and a novel splicing variant located in the 5′ UTR. This extended analysis could also identify an atypical phenotype for Pompe disease presented in a patient in which was explained by genetic mosaicism for a segment of chromosome 17 that includes *GAA*.

We maintain a public database on the understanding of the genotype-phenotype correlation in Pompe disease (www.pompevariantdatabase.nl), linking disease-associated variants to clinical severity. This database is important to guide clinicians and clinical geneticists to establish an accurate molecular diagnosis. In the recent update (**Chapter 4**), the actual database includes 648 disease-associated variants, 26 variants from newborn screening and 237 variants with unknown severity. Common sequence variants (minor allele frequency ≥ 1%) were added in response to the misreporting of such variants as the principal cause of disease in several patients. The data added to the Pompe variant database contains important information regarding genotypes and phenotypes from many publications on Pompe disease. In **Chapter 5**, we analysed all data present in the database for the most common *GAA* genotypes to analyse the phenotype-genotype correlation. Patients with the common c.-32-13T>G (IVS1)/null *GAA* genotype have a broad-spectrum symptom onset ranging from early childhood to late adulthood. Nine distinct *GAA* disease-associated variants (c.2647-7G>A; c.546G>T; c. 2238G>C; c.2173C>T; c.1076-22T>G; c.1309C>T; c.-32-3C>A; c.1875C>G; c.1064T>C) also showed considerable variation in age at symptom onset when present at compound heterozygous state in combination with a null allele or at homozygous state. The silent c.510C>T variant is linked to c.-32-13T>G in some cases and is associated with accelerated onset of symptoms. For many patients it is still unclear why phenotypic variation happens for these common *GAA* genotypes, which implies the influence of modifying factors.

Environmental factors, as well as the epigenetic and genetic background, may influence the genotype-phenotype correlation. A potential genomic modifier is the haplotype configuration. Usually, haplotypes have variants that influence the expression of mRNA transcripts, known as eQTLs. To analyse the possible impact of eQTLs on GAA expression, in **Chapter 6**, we analysed the disease-associated genetic environment of *GAA*. We defined European *GAA* haplotypes, searching for haplotype-based genotype-phenotype

correlations and the possible impact on GAA expression. We could see that most disease-associated variants are present in haplotypes that negatively modulate GAA expression, which may increase their penetrance. However, the IVS1 variant was found in a haplotype with a neutral effect on GAA expression, possibly suggesting a positive selection scenario.

We investigated the gene expression profile in fibroblasts to extrapolate our understanding of patients with the common IVS1/null *GAA* genotype (**Chapter 7**). Our analysis pointed to the conclusion that several factors are possibly associated with the clinical course of Pompe patients. Comparing the expression profile of two groups of patients, one with the onset of first symptoms under the age of 16 years and one with the onset of symptoms over the age of 35 years, we could see that three genes stood out (*MAOA*, *STC-1* and *NRK-1*), suggesting that these could be disease-markers. MAOA (Monoamine Oxidase Type A) catalyses the oxidation of monoamines, such as dopamine, serotonin and adrenalin; STC1 (Stanniocalcin-1) plays a role in regulating renal and intestinal calcium and phosphate transport, cell metabolism, or cellular calcium/phosphate homeostasis; and NRK (Nik Related Kinase) has been associated with the induction of actin polymerisation in late embryogenesis. The overall implication of these genes in biological processes includes bone and muscle development, cellular calcium/phosphate homeostasis, cellular metabolism, and cellular biogenic amine metabolic process suggesting that these genes might be considered potential disease-markers in patients with the IVS1/null genotype.

Finally, **Chapter 8** discusses the main findings, clarifies the methods for enzymatic diagnosis, and provides ways to improve the molecular analysis and to increase our knowledge on the genotype-phenotype correlation in Pompe disease.

SAMENVATTING

De *ziekte van Pompe* is een zeldzame, autosomaal recessieve aandoening die veroorzaakt wordt door ziekte geassocieerde varianten (mutaties) in het gen dat codeert voor het enzym zure-α-glucosidase (GAA). GAA is verantwoordelijk voor de omzetting van glycogeen naar glucose, en een gebrek aan actief GAA zal resulteren in de opstapeling van glycogeen in het lysosoom waardoor schade aan de cellen en weefsels kan ontstaan. De vroeg optredende (klassiek infantiele) vorm van de ziekte van Pompe, die pasgeborenen treft, ontstaat door een volledig gebrek aan GAA enzym activiteit (0%). Deze vorm wordt voornamelijk gekenmerkt door hypertrofie van het hart en afname van de functie van de skeletspieren. Kinderen en volwassen waarbij de ziekte van Pompe op een later moment tot uiting komt (kindertijd of op volwassen leeftijd) hebben mildere symptomen als gevolg van resterende enzymactiviteit (1-20%), wat leidt tot langzaam progressieve spierverzwakking, mobiliteitsproblemen en ademhalingsproblemen.

Sinds 2006 is een behandeling beschikbaar voor de ziekte van Pompe: enzymvervangingende therapie (ERT), bestaande uit intraveneus toegediend humaan recombinant GAA-enzym. Deze behandeling vergroot de overlevingskans voor klassieke infantiele patiënten en zorgt voor een verbetering van verschillende symptomen bij de later optredende vormen, zoals betere loopvaardigheden en stabilisatie van de ademhalingsfunctie. Voordat patiënten met de ziekte van Pompe ERT kunnen ontvangen is een diagnose vereist op basis van GAA enzym deficiëntie, en zouden bij voorkeur twee ziektegerelateerde varianten moeten worden gedetecteerd in het *GAA* gen. Een algemene inleiding tot de ziekte van Pompe en een uitleg van het proces voor het uitvoeren van de enzymatische en moleculaire diagnose is beschreven in **hoofdstuk 1**.

Verschillende methoden kunnen gebruikt worden om de enzymactiviteit van GAA te bepalen. Echter kan de enzymatische diagnose geconfronteerd worden met enkele problemen, zoals het bestaan van vals-positieve en vals-negatieve resultaten. **Hoofdstuk 2** behandelt methoden die toegepast worden voor de enzymatische diagnose van de ziekte van Pompe, en bespreekt de uitkomst van een reeks van 2591 assays bij 1709 personen die ervan verdacht werden de ziekte van Pompe te hebben. Het toont de resultaten van verschillende typen monsters en verschillende methoden om enzymactiviteit te bepalen die de afgelopen 28 jaar zijn gebruikt op de afdeling Klinische Genetica van het Erasmus MC Medisch Centrum in Rotterdam. Om de prestaties van elke test te bepalen werd GAA-activiteit gemeten in witte bloedcellen (leukocyten) met behulp van glycogeen of 4-methylumbelliferyl-α-D-glucopyranoside (4MUG) als substraat, in gedroogde bloedvlekken (DBS) met 4MUG als substraat en in huidfibroblasten met 4MUG als substraat. Gemengde leukocyten bieden de snelste diagnostische methode, waarbij het reactiemengsel acarbose moet bevatten om glucoamylase te remmen. DBS-gebaseerde assays werden beschouwd als waardevolle

instrumenten voor screening, maar aanvullende testen zijn vereist om vals-negatieve resultaten te voorkomen en de diagnose te bevestigen. Gekweekte huidfibroblasten bieden de beste resolutie om onderscheid te maken tussen de klassiek infantiele, en de later optredende vormen van de ziekte. Glycogeen en 4MUG waren geschikte substraten, maar glycogeen discrimineert het beste tussen aangedane en niet-aangedane personen. Het nadeel van het gebruik van glycogeen is dat dragers van de GAA2 pseudo-deficiëntie verkeerd geïdentificeerd kunnen worden als patiënten. In dergelijke gevallen biedt aanvullend gebruik van 4MUG in leukocyten en/of DBS de meest betrouwbare resultaten.

Naast enzymatische analyse is ook de moleculaire analyse een essentiële stap om de ziekte van Pompe te diagnosticeren. De meest gebruikte methode is traditionele Sanger-sequencing die de coderende exon flankerende gebieden bestrijkt, maar in sommige gevallen zal deze methode niet voldoende zijn om een diagnose vast te stellen. In **hoofdstuk 3** toonden we uitgebreide moleculaire diagnostische analyses en identificeerden we nieuwe ziektegerelateerde varianten bij patiënten met de ziekte van Pompe, waarbij de conventionele analyse onvoldoende was. Met behulp van aanvullende assays zoals generieke splicing assays, minigen analyse, SNP array analyse, en gerichte Sanger sequencing, werden zowel een exonische deletie, een promotor verwijdering, en een nieuwe splicing variant gelegen in de 5′ UTR geïndentificeerd. Deze uitgebreide analyse kon ook een atypisch fenotype voor de ziekte van Pompe bij een patiënt identificeren, wat verklaard kan worden door een genetisch mozaïek voor een segment van chromosoom 17 dat *GAA* omvat.

We houden een openbare database bij over het begrip van de genotype-fenotype correlatie bij de ziekte van Pompe (www.pompevariantdatabase.nl), waarbij ziekte geassocieerde varianten worden gekoppeld aan het fenotype van de patiënten. Deze database is belangrijk om clinici en klinisch genetici te begeleiden bij het vaststellen van een nauwkeurige en accurate diagnose. In de recente update (**hoofdstuk 4**) bevat de database 648 ziektegerelateerde varianten, 26 varianten van newborn screening en 237 varianten van onbekende mate van ernst. Veel voorkomende sequentievarianten (minor allele frequency ≥ 1%) werden toegevoegd als reactie op het onjuist rapporteren van zulke varianten als de de voornaamste oorzaak van de ziekte bij verschillende patiënten. De gegevens die zijn toegevoegd aan de Pompe variant database bevatten belangrijke informatie over genotypen en fenotypen uit vele publicaties. In **hoofdstuk 5** analyseerden we alle gegevens in de database voor de meest voorkomende *GAA*-genotypen om de fenotype-genotype correlatie te bekijken. Patiënten met het gemeenschappelijke c.-32-13T>G (IVS1)/null *GAA*-genotype hebben een breed spectrum van het moment van ontstaan van de ziekte, variërend van de kindertijd tot op latere volwassen leeftijd. Negen verschillende *GAA*-ziektegerelateerde varianten (c.2647-7G>A; c.546G>T; c. 2238G>C; c.2173C>T; c.1076-22T>G; c.1309C>T; c.-32-3C>A; c.1875C>G; c.1064T>C) vertoonden ook een aanzienlijke variatie in de leeftijd waarop de symptomen begonnen, wanneer deze heterozygoot voorkwamen in combinatie

met een nul-allel of wanneer deze homozygoot aanwezig waren. De stille c.510C>T variant is in sommige gevallen gekoppeld aan c.-32-13T>G en wordt geassocieerd met een versneld begin van de symptomen. Voor veel patiënten is het nog steeds onduidelijk waarom een variatie in het fenotype optreedt voor deze veel voorkomende *GAA*-genotypen, wat impliceert dat er modificerende factoren moeten zijn.

Omgevingsfactoren en de epigenetische en genetische achtergrond van de patiënt kunnen de correlatie tussen genotype en fenotype beïnvloeden. Een potentiële genomische modificator is de haplotype configuratie. Meestal hebben haplotypen varianten die de expressie van mRNA-transcripties beïnvloeden, beter bekend als eQTL's. Om de mogelijke impact van eQTL's op de GAA-expressie te analyseren hebben we in **hoofdstuk 6** de ziekte gerelateerde genetische omgeving van *GAA* geanalyseerd. We definieerden Europese *GAA*-haplotypes die een mogelijke impact hebben op de GAA-expressie. De meeste ziekte geassocieerde varianten zijn aanwezig in haplotypes die de GAA-expressie negatief kunnen beïnvloeden, wat hun penetrantie kan verhogen. De IVS1-variant werd echter gevonden in een haplotype met een neutraal effect op de GAA-expressie, wat een positief selectie scenario suggereert.

We onderzochten het genexpressieprofiel in fibroblasten om ons begrip van patiënten met het gemeenschappelijke IVS1/null *GAA*-genotype te vergroten (**hoofdstuk 7**). Onze analyse suggereert dat verschillende factoren mogelijk verband houden met het klinische verloop van Pompe-patiënten. Door het expressieprofiel van twee groepen patiënten te vergelijken, één met het begin van de eerste symptomen onder de leeftijd van 16 jaar en één met het begin van symptomen ouder dan 35 jaar, kwamen drie genen naar boven (*MAOA*, *STC-1* en *NRK-1*), wat suggereert dat dit potentiële ziekte-markers kunnen zijn. MAOA (Monoamine Oxidase Type A) katalyseert de oxidatie van monoamines, zoals dopamine, serotonine en adrenaline; STC1 (Stanniocalcine-1) speelt een rol bij het reguleren van het renale en intestinale calcium- en fosfaattransport, het cel metabolisme of de cellulaire calcium/fosfaathomeostase; en NRK (Nik Related Kinase) is geassocieerd met de inductie van actinepolymerisatie bij late embryogenese. De algehele implicatie van deze genen in biologische processen omvat bot- en spierontwikkeling, cellulaire calcium/ fosfaathomeostase, cellulair metabolisme en het cellulair biogene aminemetabolische proces, wat aangeeft dat deze genen mogelijk kunnen worden beschouwd als potentiële ziekte-markers bij patiënten met het IVS1/null genotype.

Ten slotte bespreekt **hoofdstuk 8** de belangrijkste bevindingen, verduidelijkt het de methoden voor enzymatische diagnose en biedt het manieren om de moleculaire analyse te verbeteren en onze kennis over de correlatie tussen genotype en fenotype bij de ziekte van Pompe te verhogen.

RESUMO

A *doença de Pompe* é uma doença autossômica recessiva rara, causada por mutações no gene que codifica a enzima α-glucosidase ácida (GAA). A função da enzima GAA é degradar o glicogênio em glicose no lisossomo. Sua deficiência resulta no acúmulo de glicogênio no lisossomo, causando danos às células e tecidos, levando à perda de função celular. A ausência completa da atividade da enzima GAA causa doença de Pompe infantil (forma clássica), afetando recém-nascidos, levando principalmente a hipertrofia cardíaca e perda de músculo esquelético. Pacientes infantis ou adultos com Pompe têm um fenótipo mais brando porque apresentam uma atividade enzimática residual (1-20%), causando fraqueza muscular progressiva, problemas de mobilidade e insuficiência respiratória.

A terapia de reposição enzimática (TRE) é utilizada como tratamento para esses pacientes e está disponível desde 2006 e consiste na administração intravenosa da enzima GAA recombinante humana. Quando tratados, pacientes com a formas infantil apresentam aumentos substanciais nas taxas de sobrevivência. Crianças e pacientes adultos também se beneficiam melhorando suas habilidades de caminhada e estabilização da função respiratória. Entretanto, para que os pacientes com a doença de Pompe recebam TRE, é necessário um diagnóstico apropriado com base na deficiência da enzima GAA e, preferivelmente, que sejam detectadas duas variantes associadas à doença no gene *GAA*. O **Capítulo 1** da presente tese apresenta uma introdução geral à doença de Pompe bem como uma explicação do processo para realizar o diagnóstico enzimático e molecular.

Atualmente, diferentes métodos e amostras podem ser utilizados na determinação da atividade enzimática da GAA. No entanto, o diagnóstico enzimático enfrenta algumas limitações, como a existência de resultados falso-positivos e falso-negativos confiáveis. Assim, é objetivo do **Capítulo 2**, relatar os métodos atualmente utilizados no diagnóstico enzimático da doença de Pompe e revisar o resultado de uma série de 2.591 ensaios em 1.709 indivíduos com suspeita de ter a doença. Este capítulo apresenta os resultados nos quais diferentes tipos de amostras bem como diferentes métodos que foram utilizados nos últimos 28 anos no departamento de Genética Clínica do Erasmus Medical Center em Rotterdam. A fim de determinar o desempenho de cada ensaio, a atividade enzimática para GAA foi medida em glóbulos brancos (leucócitos) usando glicogênio ou 4-metilumbeliferil-α-D-glucopiranosídeo (4MUG) como substrato, gotas de sangue secas (DBS) usando 4MUG como substrato, e de fibroblastos obtidos de biópsia de pele usando 4MUG como substrato. As amostras de leucócitos apresentaram o método de diagnóstico mais rápido, em que a mistura de reação precisa conter acarbose para a inibição da glucoamilase. Os ensaios baseados em DBS foram considerados ferramentas valiosas para a triagem, mas um ensaio adicional, como por exemplo análise em leucocitos é necessário para confirmar o diagnóstico e evitar resultados falso-negativos. Fibroblastos de pele cultivados

apresentaram os melhores resultados na diferenciação entre a forma infantil, a forma juvenil e a forma adulta da doença de Pompe. Glicogênio e 4MUG são substratos adequados, mas o glicogênio é capaz de discriminar melhor indivíduos afetados de não afetados. A desvantagem de usar glicogênio é que os portadores da pseudo deficiência GAA2 podem ser diagnosticados erroneamente como pacientes de Pompe. Nesses casos, o uso adicional de 4MUG resultou em uma abordagem mais confiável e os resultados foram validados em leucócitos e / ou DBS.

A análise genética-molecular também é uma etapa essencial para concluir o diagnóstico de um paciente Pompe com maior precisão. O sequenciamento Sanger tradicional cobrindo as áreas flanqueadoras do exon de codificação é o método mais comumente usado, no entanto, em alguns casos, esse método não se mostra suficiente para concluir o diagnóstico. No **Capítulo 3**, apresentamos análises aprofundadas de diagnóstico molecular e identificamos novas variantes genéticas associadas à doença em pacientes com doença de Pompe, onde a análise convencional era insuficiente. Ensaios adicionais, como um ensaio padrão de *splicing*, análise por minigene, análise de SNP array e sequenciamento Sanger direcionado, permitiram a identificação de uma deleção exônica, uma deleção do promotor bem como uma nova variante de *splicing* localizada na região 5 'UTR do gene. Esta análise mais detalhada também é capaz de identificar um fenótipo atípico para a doença de Pompe detectado em um paciente no qual pôde ser explicado por mosaicismo genético de um segmento do cromossomo 17 que inclui *GAA*.

Nosso laboratório mantém um importante banco de dados público sobre a compreensão da correlação genótipo-fenótipo na doença de Pompe (www.pompevariantdatabase.nl), relacionando as variantes associadas à doença à gravidade clínica. Este banco de dados é fundamental para orientar os médicos e geneticistas clínicos a estabelecer um diagnóstico molecular preciso. Na atualização recente (**Capítulo 4**), o banco de dados inclui 648 variantes associadas à doença, 26 variantes da triagem neonatal e 237 variantes com gravidade desconhecida. Variantes comuns (frequência do alelo menor(MAF) ≥ 1%) foram adicionadas em resposta a interpretação incorreta de tais variantes como a principal causa da doença em vários pacientes. Os dados adicionados ao banco de dados de variantes de Pompe contêm informações importantes sobre genótipos e fenótipos de diversas publicações científicas sobre a doença de Pompe. No **Capítulo 5**, apresentamos a análise de todos os dados presentes no banco para os diferentes genótipos *GAA* mais comuns com o intuito de avaliar a correlação fenótipo-genótipo. Pacientes com o genótipo comum c.-32-13T> G (IVS1) / *GAA* nulo apresentam o início de sintomas em um amplo espectro que varia desde a primeira infância até a idade adulta tardia. Nove variantes *GAA* distintas associadas à doença (c.2647-7G> A; c.546G> T; c. 2238G> C; c.2173C> T; c.1076-22T> G; c.1309C> T; c. -32-3C> A; c.1875C> G; c.1064T> C) também mostraram variação considerável na idade no início dos sintomas quando presente em heterozigose composta em combinação com um alelo nulo ou em homozigose. A variante silenciosa

c.510C> T está associada a c.-32-13T> G em alguns casos e está relacionada ao início acelerado dos sintomas. Porém, para muitos pacientes ainda não está claro por que a variação fenotípica ocorre para esses genótipos *GAA* comuns, o que sugere a influência de fatores modificadores.

Fatores ambientais, bem como modificações epigenéticas e antecedentes genéticos podem influenciar a correlação genótipo-fenótipo. Um potencial modificador genômico é a configuração do haplótipo. Normalmente, os haplótipos possuem variantes que influenciam a expressão de transcritos de mRNA, conhecidos como eQTLs. Para analisar o possível impacto dos eQTLs na expressão de *GAA* analisamos o ambiente genético *GAA* de variantes associadas à doença, apresentado no **Capítulo 6**. Ao identificar correlações genótipo-fenótipo baseadas em haplótipos fomos capazes de definir haplótipos *GAA* mais frequentes em indivíduos europeus bem como um possível impacto na expressão de *GAA*. Apresentamos evidências de que a maioria das variantes associadas à doença estão presentes em haplótipos que modulam negativamente a expressão de *GAA*, o que pode levar ao aumento da sua penetrância. No entanto, a variante IVS1 foi encontrada em um haplótipo com efeito neutro na expressão de *GAA*, possivelmente sugerindo um cenário de seleção positiva.

Investigamos o perfil de expressão gênica em fibroblastos para extrapolar nossa compreensão genética de pacientes com o genótipo comum IVS1 / *GAA* nulo (**Capítulo 7**). Nossa análise apontou para a conclusão de que vários fatores estão possivelmente associados ao curso clínico dos pacientes com Pompe. Comparando o perfil de expressão de dois grupos de pacientes, um com início dos primeiros sintomas abaixo dos 16 anos e outro com início dos sintomas acima dos 35 anos, pudemos observar que três genes se destacaram (*MAOA*, *STC-1* e *NRK-1*), sugerindo que estes podem ser marcadores de doença. MAOA (Monoamina Oxidase Tipo A) catalisa a oxidação de monoaminas, como dopamina, serotonina e adrenalina; STC1 (Stanniocalcina-1) desempenha um papel na regulação renal e intestinal de cálcio e transporte de fosfato, metabolismo celular ou homeostase celular de cálcio / fosfato; e NRK (Nik Related Kinase) está associado à indução da polimerização da actina em estágios tardios da embriogênese.

Por estarem de um modo geral associados a processos biológicos que incluem desenvolvimento ósseo e muscular, homeostase celular de cálcio / fosfato, metabolismo celular e processo metabólico de amina biogênica celular, o padrão de expressão destes três genes pode ser considerado como marcadores para o diagnóstico da doença doença de Pompe possível para pacientes com o genótipo IVS1 / nulo.

Finalmente, o **Capítulo 8** discute as principais descobertas apresentadas nesta tese, esclarece os métodos de diagnóstico enzimático e fornece maneiras de melhorar a análise molecular e aumentar nosso conhecimento sobre a correlação genótipo-fenótipo na doença de Pompe.

LIST OF PUBLICATIONS

Douglas O. S. de Faria*, Stijn L. M. in 't Groen*, Marianne Hoogeveen-Westerveld, Monica Y. Niño Ans T. van der Ploeg Atze J. Bergsma W.W. M. Pim Pijnappel. Update of the Pompe variant database for the prediction of clinical phenotypes: Novel disease-associated variants, common sequence variants, and results from newborn screening. December 2020 - *Human Mutation* 42(2):119-134

Douglas O. S. de Faria, Stijn L.M. in 't Groen, Atze J. Bergsma, Monica Y. Nino, Ans T. van der Ploeg, Oscar Lao, W.W.M. Pim Pijnappel. The haplotype paradox in Pompe disease: differential effects of haplotypes on the penetrance of European ancestry disease-associated variants. *In preparation*

Stijn L M In 't Groen, **Douglas O S de Faria**, Alessandro Iuliano, Johanna M P van den Hout, Hannie Douben, Trijnie Dijkhuizen, David Cassiman, Peter Witters, Miguel-Ángel Barba Romero, Annelies de Klein, Galhana M Somers-Bolman, Jasper J Saris, Lies H Hoefsloot, Ans T van der Ploeg, Atze J Bergsma, W W M Pim Pijnappel. Novel *GAA* Variants and Mosaicism in Pompe Disease Identified by Extended Analyses of Patients with an incomplete DNA Diagnosis. January 2020- *Molecular Therapy- Method & Clinical Development* 17:337-348

Monica Y. Niño, Mark Wijgerde*, **Douglas Oliveira Soares de Faria***, Marianne Hoogeveen-Westerveld, Atze J. Bergsma, Mike Broeders, Nadine A. M. E. van der Beek, Hannerieke J. M. van den Hout, Ans T. van der Ploeg, Frans W. Verheijen & W. W. M. Pim Pijnappel. Enzymatic diagnosis of Pompe disease: lessons from 28 years of experience. November 2020 - *European Journal of Human Genetics* 29(3):434-446

Monica Y. Niño*, Stijn L.M. in 't Groen*, **Douglas O.S. de Faria**, Marianne Hoogeveen-Westerveld, Hannerieke J.M.P. van den Hout, Ans T. van der Ploeg, Atze J. Bergsma, and W.W.M. Pim Pijnappel. Broad variation in phenotypes for common *GAA* genotypes in Pompe disease. August 2021 – *Human Mutation* Online ahead of print

Ni-Chung Lee, Kai-Ling Chang, Stijn L. M. in 't Groen, **Douglas O. S. de Faria**, Hsiang-Ju Huang, W.W.M. Pijnappel, Wuh-Liang Hwu, Yin-Hsiu Chien. True incidence through clinical curation and functional analysis and outcome of late-onset Pompe disease via newborn screening. *Submitted*

Monica Y. Niño, **Douglas O. S. de Faria**, Marian Kroos, Marianne Hoogeveen-Westerveld,

Stijn L.M. in 't Groen, J.M.P. [Hannerieke] van Den Hout, Wilfred Van IJcken, Ans T. van der Ploeg, W.W.M. Pijnappel. Clinical diversity of Pompe patients with c.-32-13T>G genotypes investigated by gene expression profiling. *In preparation*

Priscila Nicolicht, **Douglas O.S. Faria**, Leonardo Martins-Silva, Luana S .M. Maia, Adriana S. Moreno, L.K. Arruda, Antonio A. Motta, Anete S . Grumach, João B. Pesquero. Gene mapping strategy for Alu elements rearrangements: Detection of new large deletions in the SERPING1 gene causing hereditary angioedema in Brazilian families. February 2019 - *Gene* 15;685:179-185

Lauro Thiago Turaça, **Douglas Oliveira Soares de Faria**, Sandra Obikawa Kyosen, Valber Dias Teixeira, Fabiana Louise Motta, Juliana Gilbert Pessoa, Marina Rodrigues E Silva, Sandro Soares de Almeida, Vânia D'Almeida, Maria Verônica Munoz Rojas, Ana Maria Martins, João Bosco Pesquero. Novel *GAA* mutations in patients with Pompe Disease. April 2015 - *Gene* 25;561(1):124-31

CURRICULUM VITAE

Personal Information

Name: Douglas Oliveira Soares de Faria
Date of Birth: 30-01-1987
Place of Birth: Uberaba, Minas Gerais, Brazil
Nationality: Brazilian
Email: douglasfariaxp@gmail.com

Education

Nov 2015- Jul 2021: Erasmus University Medical Center, Rotterdam, the Netherlands

- **Ph.D. training** in the Dept. of Clinical Genetics and Dept. of Pediatrics
  - Thesis - Genetics of Pompe Disease: The secrets within the coding region of the *GAA* gene and beyond

Mar 2012-Jul 2015: Federal University of São Paulo (UNIFESP), São Paulo, Brazil

- **Master's degree** in Molecular Biology
  - Thesis - Development of molecular diagnostic techniques for detection of large deletions and insertions in Pompe Disease and identification of Alu element in the *GAA* gene

Jul 2006 -Dec 2010: Federal University of Triângulo Mineiro (UFTM), Uberaba, Brazil

- **Bachelor's degree** in Biomedical Sciences
  - Thesis - Gene expression analysis of Major Surface Proteases in *Trypanosoma rangeli*.

Professional Experience

- Nov 2015 – Jul 2021 **Ph.D. training** in the Dept. of Clinical Genetics and Dept. of Pediatrics. Erasmus University Medical Center, Rotterdam, the Netherlands.
- Nov 2011 – Jul 2015 **MSc / Molecular diagnostic laboratory** - Federal University of São Paulo (UNIFESP), São Paulo, Brazil.
- Jan 2010 – Dec 2010 **Clinical laboratory Internship** - UFTM Clinical Hospital, Uberaba, Brazil
- Feb 2010 – Dec 2010 **Teaching Assistant** – Basic and applied Molecular Biology.

Federal University of Triângulo Mineiro (UFTM), Uberaba, Brazil

- Mar 2009 – Feb 2010 **Researcher** - Federal University of Triângulo Mineiro (UFTM), Uberaba, Brazil.
- Aug 2008 – Dec 2008 **Teaching Assistant** – Entomology. Federal University of Triângulo Mineiro (UFTM), Uberaba, Brazil
- Mar 2008 – Jun 2008 **Teaching Assistant** – Ecology and Evolution. Federal University of Triângulo Mineiro (UFTM), Uberaba, Brazil
- Feb 2007 - Nov 2007 **Teaching Assistant** – Bioinformatics. Federal University of Triângulo Mineiro (UFTM), Uberaba, Brazil

Certificate

- Article 9 - the Dutch Law on Animal Experiments (Wet op de Dierproeven), EMC, Rotterdam, the Netherlands - 2016

Awards

- 2015 – CAPES scholarship – Ph.D. abroad
- 2015 – CNPq scholarship – Ph.D. abroad
- 2013 - International GSD Conference – Travel grant
- 2012 - CNPq scholarship – Master's program
- 2009 – FAPEMIG scholarship- Scientific initiation

PhD PORTFOLIO

Courses:	Year	ECTS
Special topics – Chromatin	2015	0.3
Special topics – Signalling	2016	0.3
Laboratory animal science	2016	3
The SNP course XIII: SNPs and Human Disease	2016	2
Research integrity	2017	0.3
Special topics – CRISPR-Cas	2017	2
Course on R	2017	1.8
Genetics course	2017	3
Practical Linux	2017	0.4
ESP74 Genome-wide Association Studies	2017	0.7
Safely working in the Laboratory	2017	0.5
Code and Data Management with Git	2017	0.5
Biostatistical Methods I: Basic Principles part A	2018	2
Scripting for Life Science Researchers	2018	0.8
Next Generation Sequencing Data Analysis	2018	1.4
Biomedical English Writing and Communication	2018	3
Workshops:	**Year**	**ECTS**
23rd MGC PhD Workshop, Dortmund, Germany	2016	1
24th MGC PhD Workshop, Leuven, Belgium	2017	1
UCSC Gene Browsing workshop (Basic/Adv.)	2018	0.7
25th MGC PhD Workshop, Texel, the Netherlands	2018	1
26th MGC PhD Workshop, Maastricht, the Netherlands	2019	1
(Inter) National Meetings:	**Year**	**ECTS**
Muscles2Meet, Neuromuscular Young Talent Symposium, Zeist, NL	2016	1
MGC Symposium, Leiden, NL	2016	0.5
8th European Symposium on Steps Forward in Pompe Disease, Amsterdam, NL	2016	1
Muscles2Meet, Neuromuscular Young Talent Symposium, Zeist, NL	2017	1
Sophia Research Day, Rotterdam, NL	2019	1
Muscles2Meet, Neuromuscular Young Talent Symposium, Zeist, NL	2019	1
SSIEM Annual Symposium 2019, Rotterdam, NL	2019	1
MGC Symposium, Rotterdam, NL	2019	0.5
EMC Research meetings	2015-19	1

Presentations:	Year	ECTS
Poster presentation 24th MGC PhD Workshop, Leuven, BE	2017	0.25
Oral presentation 25th MGC PhD Workshop, Texel, NL	2018	0.5
Oral presentation 26th MGC PhD Workshop, Maastricht, NL	2019	0.5
Poster presentation SSIEM Annual Symposium	2019	0.25
Academic League of Bioinformatics and Molecular Biology, Uberaba, Brazil	2021	0.5
Teaching activity:	**Year**	**ECTS**
Supervisor of Bachelor student	2018-19	2
	Total ECTS	38.7

ACKNOWLEDGEMENTS / Agradecimentos

It is with great joy and a feeling of gratitude that I begin to write the acknowledgements.

Since I was a child, studying a lot, always enchanted by science, I would have never imagined that one day I would reap the fruits of many efforts at one of the best institutions in the world in education and research, **the Erasmus University**. I am very grateful for having the opportunity to grow professionally. I had the chance to immerse myself in the Dutch culture, which is so different from Brazilian culture. However, today I'm a much more resilient person, and I couldn't do my PhD if it weren't for the help of all of you.

First of all, I would like to thank my promoter and co-promoter, **Prof. dr Ans van der Ploeg** and **Dr Pim Pijnappel**, for the opportunity to work in their reference group for Pompe disease and for always supporting me when necessary. **Ans**, thank you for always seeing the human side of people. Your work and effort bring dignity to many people. Your leadership at the **Center for Lysosomal and Metabolic Diseases** is admirable. I would like to thank all people from the center, thank you for the research and social moments. **Pim**, I will always be grateful for the opportunity that changed my life and everyone connected with me. Thank you for being an excellent mentor, professionally and personally. You are a tough yet motivating person encouraging us to continuously pursue to grow as scientists, ensuring that anything can be achieved with hard work and diligence. I admire your work and wish you the best in your future endeavours.

Dear **Prof. dr. Rene.M.H. Wijnen**, dear **Prof. dr. Pieter A. van Doorn** and dear **Prof. dr. Clara D.M. van Karnebeek** thank you for your comments and the time that you have spent on my thesis. I am honoured to have you on my small committee.

A great thank you to all the patients and family members involved in our projects; you motivate our work. Moreover, I thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (**CNPq**) that collaborated in funding my doctorate. **João Bosco Pesquero**, thank you very much for giving me several opportunities in Brazil; if it weren't for everything I learned from you and in your laboratory, I wouldn't have earned my doctorate abroad.

Arnold Reuser, thank you so much for allowing me to meet you at the GSD Conference 2013 in Heidelberg. It's incredible to imagine that a 5-minute conversation would've changed my life in such a significant way. I admire your career, and for me, it's an honour to say that we're friends today. Thank you for all the help and advice. **Oscar Lao** it is an honour to have you on my committee, and it was a pleasure to collaborate with you. Your work in population genetics deserves to be admired. I will be pleased if one day I have the chance to work directly with you.

I would like to thank all the **Department of Clinical Genetics** members for creating a friendly working environment. The leadership of **Prof. Robert Hofstra** and **Prof. Rob Willemsen** was memorable, giving scientific feedback and making the working

environment much more pleasant. I wish all success for the new head of the Department, **Prof Ype Elgersma**. I would like to thank all the people who helped me in any way, whether it was with scientific comments, suggestions, teaching, or having a good talk, especially **Dr Tjakko van Ham**, **Prof. Vincenzo Bonifati**, **Erwin Brosens**, **Bianca de Graaf**, **Guido**, **Shami**, **Laura Kuil**, **Adriana Iglesias, Santoesha, Elena, Domenico**, **Marianne Hoogeveen** and **Fernanda Jehee** (*muito obrigado!*) Thank you for all the help when I found it was necessary. A special thanks to **Herma van der Linde**, a master in Zebrafish. I learned a lot from you, and your passion for your work and your patience in teaching made working with this animal model more pleasant.

Maria Alves *muito obrigado por todo o suporte, principalmente no começo do meu doutorado.* I wish all the best for you, Arya and Kaushal.

Many thanks to **Prof dr J.B.J van Meurs** and **Dr Cindy Boer**. Thank you for helping me on one of the topics that fascinated me the most: **eQTLs**. I would like to thank my neighbours from the **Department of Developmental Biology**, especially some people such as **Andrea, Teresa** and **Joachim**. But mainly, **Joana Moreira**, who became an excellent friend. *Muito obrigado Joana, obrigado pelas conversas, conselhos e sempre por me motivar. O seu conhecimento em bioinfórmatica é admiravel, além disso você é uma cientista brilhante! Todo sucesso para você e o Ruggero.*

I could not have made it this far without the help of the members of my group. In addition to colleagues, they became my friends and my time with you; I carry great memories with me.

Monica Nino, I don't know how to thank you, your friendship was essential during the trajectory of my studies. I am delighted that your friendship will last a lifetime. You saw my joys, sorrows and achievements up close. Thanks for everything, my friend.

Stijn, thank you so much for all your help. Your scientific knowledge is admirable, and discussions were always fruitful. I appreciate your friendship, which was hard to win. I am sure you will have a brilliant career, and I wish you all the success in the world –best of luck with the rest of your PhD.

Mike, my favourite Dutch, I had a lot of fun with you. Thanks for the conversations and advice and for always showing the positive side of things. I admire your scientific work and your artwork ability. My goal is to be a bit like you. Thank you so much for everything, and I wish you and Helene all the best.

Gerben, it was fun to sit next to you for a few years, and you showed me a little bit of Dutch culture. **Atze**, thanks for being a nice guy, all the scientific feedback, conversations, and readiness to interact and participate in all social activities. Thanks a lot! I wish you both all the best in your career because you deserve it.

My favourite Spaniards, **Pablo** and **Rodrigo**. Definitely, starting my doctorate with you as colleagues made this moment much more pleasant. You have an admirable nature, always ready to help when needed. Every conversation with you was fun and informative. I

had a lot of fun with you guys, and I have great memories. **Rodrigo**, thank you so much for being my paranymph; your company on the "big day" will be essential. I wish you all the best in your career!

My favourite Italians, **Alessandro** and **Fábio**. It's been enjoyable to be with you. It's always a lot of fun. Thanks for all the conversations and scientific feedback. Because of you, I like **Italy** even more. All the success for you and your doctorate. **Eva**, thank you so much for always being helpful and kind. You are an excellent, intelligent person with a unique personality. I wish you all the success in your career, and keep in mind that you can achieve everything you aspire to.

Thank you very much to my former colleagues, who I had the opportunity to meet and work with. **Merel**, your generosity is admirable. Thank you so much for helping me at the beginning of my PhD; you were always ready to help. You were one of the people who made me admire the Dutch people. **Erik van der Wal**, thank you for your friendship, conversations and inspirations. Always in some discussion, it was full of insights. I admire you a lot and wish you the best for you and your family. **Qiushi** and **Wenhao**, definitely the most wonderful Chinese people I had the opportunity to meet. Thank you so much for everything. **Tom** and **Joon**, it was a pleasure working with you and always a lot of fun. To former students **Max** and **Rehannah**, it was always fun working with you, and I hold great memories with me from our time together. My former student **Yasemin**. It was great working with you, and I learned a lot. Thank you for giving me this opportunity. I foresee a brilliant career for you. Keep it up!

Alta, **Bella** and **Banjo**, thanks for bringing joy to Vanessa and me.

Of course, my doctorate wouldn't be the same if I didn't have amazing people in my life, such as my wife, family and friends.

*Aos meus amigos **Paulo** e **Manu**, muito obrigado por serem prestativos e sempre ajudar no que for possível. **Paulo** muito obrigado por ser meu paraninfo e me apoiar no meu doutorado e ouvir minhas lamúrias hahah. Sua amizade traz a segurança que todo expatriado precisa! Muito obrigado.*

*A todos os meus amigos que sempre me apoiaram, muito obrigado. A minha grande família, muito obrigado por me apoiarem e acreditarem em mim. Espero ter representado bem a **família Oliveira e Soares de Faria**! Amo todos vocês!*

*A família que peguei emprestado da Vanessa, **Batista e Das Dores**, muito obrigado por confiarem em mim para entrar na família, e confiar em cuidar da sua filha, que está tão longe de todos que ela ama. Obrigado pelas conversas nos momentos mais difíceis.*

*Aos meus pais que amo tanto, que sempre me apoiaram em todas as minhas decisões e acreditarem no meu potencial, na qual eu possa sempre conquistar aquilo que eu almejo. Obrigado minha **mãe Ligia**, por acreditar em mim e tentar proporcionar as melhores condições nos estudos e sempre me mimando na medida do possível. Ao*

*meu **pai Riva**, meu eterno agradecimento por me apoiar nos estudos. E também sempre me incentivando a continuar lutando pelos meus objetivos com fé em **Deus**, porque tudo tem um significado maior. Obrigado pelas palavras de apoio nos momentos mais difíceis estando tão longe de "casa". **Amos vocês**.*

*A minha linda **esposa Vanessa**, que desde o meu mestrado sempre está me apoiando, sempre esteve lá nos momentos mais difíceis. Agradeço a **Deus** todos os dias por ter colocado você em minha vida. Não imagino ter chegado tão longe e ter conquistado isso sem você. Admiro sua fibra em largar tudo no Brasil e mudar de país para esta aventura. Só você sabe realmente por tudo que eu passei e serei eternamente grato por toda a sua ajuda. Obrigado por estar na minha vida. **Te amo hoje e sempre**.*

And last but not least. Thanks to **the Netherlands** for providing me with one of the best feelings that every human being can have: the feeling of freedom. **Dank je wel**!