Molecular Genetics and Metabolism Reports 27 (2021) 100734

Contents lists available at ScienceDirect

Molecular Genetics and Metabolism Reports

journal homepage: www.elsevier.com/locate/ymgmr

Carrier frequency and predicted genetic prevalence of Pompe disease based on a general population database

Kyung Sun Park

Department of Laboratory Medicine, Kyung Hee University School of Medicine and Kyung Hee University Medical Center, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Pompe disease
Population database
Carrier frequency
Genetic prevalence
Clinical severity
gnomAD

ABSTRACT

Background: The genetic prevalence of Pompe disease was estimated based on the proportion of individuals who have a causative genotype in a general population database. In addition, clinical severity for causative genotypes was assessed based on currently available locus-specific databases (LSDBs), which contain information on both genotype and clinical severity.

Methods: Genetic variants in the *GAA* gene in the Genome Aggregation Database (gnomAD) (v2.1.1) were analyzed in combination with LSDBs of ClinVar, ClinGen Evidence Repository, Pompe disease *GAA* variant database, and the Pompe Registry. Carrier frequency (CF) and predicted genetic prevalence (pGP) were estimated.

Results: Of 7 populations, East Asian and African showed higher proportions of pathogenic or likely pathogenic variants (PLPVs) associated with classic infantile-onset Pompe disease. Total CF and pGP in the overall population were 1.3% (1 in 77) and 1:23,232, respectively. The highest pGP was observed in the East Asian population at 1:12,125, followed by Non-Finnish European (1:13,756), Ashkenazi Jewish (1:22,851), African/African-American (1:26,560), Latino/Admixed American (1:57,620), South Asian (1:93,087), and Finnish (1:1,056,444).

Conclusions: Pompe disease has a higher pGP (1:23,232) than earlier accepted (1:40,000). The pGP for Pompe disease was expectedly wide by population and consistent with previous reports based on newborn screening programs (approximately 1:10,000–1:30,000).

1. Introduction

Pompe disease is a monogenic autosomal recessive disorder that has been exceptionally well-studied in terms of pathogenesis, clinical features, prognosis, screening/diagnostic methods, and treatment strategy. Pompe disease or glycogen storage disease type II (MIM #232300) is caused by deficiency of the lysosomal alpha-glucosidase (GAA), resulting in accumulation of glycogen within the lysosomes and subsequent progressive cardiac and skeletal muscle destruction [1]. The *GAA* gene encodes GAA protein. Different pathogenic variants in the *GAA* gene affect the level of GAA enzyme deficiency. Clinical severity and clinical manifestation of Pompe disease appear to be dependent on residual enzyme activity. Classic infantile-onset Pompe disease with almost complete deficiency of the GAA enzyme (<1% of normal) is typically fatal without treatment by enzyme replacement therapy (ERT) due to cardiorespiratory insufficiency within the first year of life. Late-onset Pompe disease is characterized by slowly progressive proximal muscle weakness and respiratory dysfunction. Patients with late-onset Pompe disease include individuals with onset before age 1 year (12 months) without cardiomyopathy and all individuals with onset after age 1 year. Because of the great success of ERT for early treatment [2,3], Pompe disease was included in the newborn screening program [4].

To date, *GAA* is reported to be the only gene associated with Pompe disease in the Online Mendelian Inheritance in Man database. More than 1000 genetic variants in *GAA* have been submitted to ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/). Among those, more than 300 variants were classified as pathogenic or likely pathogenic variants with different review statuses (accessed August 2020). Recently, the ClinGen lysosomal storage disorders expert panel released specifications to the 2015 American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) standards and guidelines [5] for the *GAA* gene (https://clinicalgenome.org/affiliation/50009/). Classification of *GAA* variants based on the severity rating system (6 classes) using in vitro study was introduced [6,7], and an extended *GAA* variant database has been released. This updated database (Pompe disease *GAA* variant database) provides information on *GAA* variants linking clinical severity based on in silico predictions, expression study results, and clinical information of the reported phenotype [8]. In

E-mail addresses: drkyungsun@gmail.com, drparkkyungsun@khu.ac.kr.

https://doi.org/10.1016/j.ymgmr.2021.100734
Received 19 September 2020; Received in revised form 15 February 2021; Accepted 15 February 2021

addition, data analysis of *GAA* variants were recently reported based on the Pompe Registry, which is an international multi-center observation program designed to track the clinical outcomes of patients with Pompe disease; that study was sponsored by Sanofi Genzyme (Cambridge, MA), and *GAA* variants were listed as 3 phenotypic subgroups (groups A, B, C) [9].

The introduction of next-generation sequencing (NGS) techniques made it possible to obtain massive amounts of genomic data with reduced cost and time. Recently, the Genome Aggregation Database (gnomAD), which contains a very large amount of genomic information associated with the general population, has been released [10,11]. Genomic information from the general population could provide important clues for predicting the prevalence of particular Mendelian disorders by analyzing the proportion of unaffected carriers.

These latest research results for Pompe disease have been aggregated in the present study. Specifically, the worldwide prevalence and clinical severity of Pompe disease according to population group were estimated based on gnomAD and currently available locus-specific databases (LSDBs), which contain information about both genotype and clinical severity.

2. Materials and methods

2.1. Analysis of GAA variant in gnomAD

The *GAA* gene (search by genomic region: chr17:78,075,355-78,093,679 (GRCh37/hg19)) was analyzed from gnomAD (v2.1.1) database (https://gnomad.broadinstitute.org/), which contains genetic variants from 125,748 exomes sequences and 15,708 genome sequences. The populations comprise 12,487 unrelated individuals in the African/African-American (AFR) group, 17,720 individuals in the Latino/Admixed American (AMR), 5185 individuals in the Ashkenazi Jewish (ASJ), 9977 individuals in the East Asian (EAS), 12,562 individuals in the Finnish (FIN), 64,603 individuals in the non-Finnish European (NFE), 15,308 individuals in the South Asian (SAS), and 3614 individuals in the other group.

The genetic variants in gnomAD database (v.2.1.1) were classified following the 2015 American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) standards and guidelines [5] and ClinGen Lysosomal Storage Disorders Expert Panel specifications (https://www.clinicalgenome.org/affiliation/50009). All variants were analyzed based on NM_000152.5 selected by matched annotation from NCBI and EMBL-EBI (MANE) (https://www.ncbi.nlm.nih.gov/refseq/MANE/). The GRCh37/hg19 genomic build was used for all position descriptions. For comparison of variants in all databases, sequence variants in nomenclature and genomic position were assessed using Mutalyzer (https://mutalyzer.nl/). Loss-of-function variants in *GAA* gene are responsible for development of Pompe disease. Therefore, PVS1 decision tree [12] by ClinGen sequence variant interpretation working group and ClinGen Lysosomal Storage Disorders Expert Panel specifications was applied for PVS1 ACMG/AMP variant criterion. For the PM2 code, minor allele frequency < 0.001 in all continental populations with >2000 alleles in gnomAD. For multiple lines of computational evidence (PP3), REVEL [13] (>0.75 for missense variants), Mutation Taster [14] and PROVEAN [15] (for in-frame insertion or deletion variants), MaxEntScan [16] and dbscSNV [17] (for predicted impact on splicing) were used to estimate the potential effect of variant pathogenicity. The active site and catalytic barrel of GAA for PVS1 decision tree were checked using Pfam (https://pfam.xfam.org/), InterPro (https://www.ebi.ac.uk/interpro/) and UniProt (https://www.uniprot.org/).

In addition, ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/, last accessed December 15, 2020), Pompe disease *GAA* variant database [8] (for simplicity, Pompe DB, http://www.pompevariantdatabase.nl/, last accessed December 15, 2020), and the Pompe Registry [9] were used as locus-specific databases (LSDBs). Variants classified by ClinGen were included as variants reviewed by expert panel (3 gold stars of review status) in ClinVar. A Venn diagram for 3 or 4 sets using interactiVenn [18] provided a comparison of genetic variants in gnomAD with those in ClinVar, ClinGen, Pompe DB, and Pompe Registry (Fig. 1). The Mutationmapper tool (https://www.cbioportal.org/mutation_mapper) was used for simultaneously visualizing genetic lesions [19]. Epitools was used (https://epitools.ausvet.com.au/) for statistical analysis.

2.2. Phenotype of clinical severity

In this study, clinical severity of specific variants was described by combining data from the Pompe DB and Pompe Registry. In the Pompe DB, patients who present symptoms with onset age younger than 12 months and hypertrophic cardiomyopathy were classified as classic infantile Pompe disease. Patients who present symptoms with onset age before 18 years and without cardiomyopathy were classified as childhood Pompe disease, and patients with onset age 18 years or later were classified as adult Pompe disease [8]. In the Pompe registry, patients with symptom onset at younger than 12 months of age with cardiomyopathy were classified as group A, patients with symptom onset at 12 years or younger without cardiomyopathy as group B, and patients with onset age 12 years or older as group C [9]. In this study, if the Pompe DB or Pompe registry reported that particular variants were associated with classic infantile type, those were regarded as possibly 'classic infantile' variants.

2.3. Carrier frequency and predicted genetic prevalence analysis

The carrier frequency (CF) and predicted genetic prevalence (pGP) for Pompe disease were analyzed based on the presumed pathogenic or likely pathogenic variants (PLPVs) in *GAA* in each population group. The allele frequency of PLPV (AF_V) and CF_V for a variant V were calculated considering only heterozygous PLPV [20,21] as follows:

$$AF_V = \frac{allele\ count - 2{*}homozygous\ count}{allele\ number}$$

$$CF_V = \frac{AF_V{*}allele\ number}{Number\ of\ individuals} = 2AF_V$$

where the allele count (number of variant alleles), allele number (number of genotyped alleles = 2 * number of individuals), and homozygous count (number of homozygous individuals) for a variant were provided by gnomAD.

The CF and pGP at the gene level (CF_G and pGP_G, respectively) were calculated as previously described [21]:

$$CF_G = 1 - \prod_{k=1}^{n}(1 - CF_V)$$

$$pGP_G = \frac{\sum_{k=1}^{n}(CF_V)_{ik}(CF_V)_{ik}}{4}$$

3. Results

3.1. Genetic variants in the GAA gene in gnomAD

GnomAD contained a total of 3270 genetic variants in *GAA* gene across populations. Among those, genetic variants within low complexity region or with flag of "uncertain", "likely not loss-of-function (LOF)", "not LOF" by loss-of-function manual curation in gnomAD were not included in this analysis. To date, 1716 variants are included in the ClinVar (including ClinGen), Pompe DB, or Pompe registry (Fig. 1A). Among those, 225 variants were included in all LSDBs.

The presumed PLPVs in the *GAA* gene in gnomAD are described in Table S1. In this study, a total of 154 different variants were classified

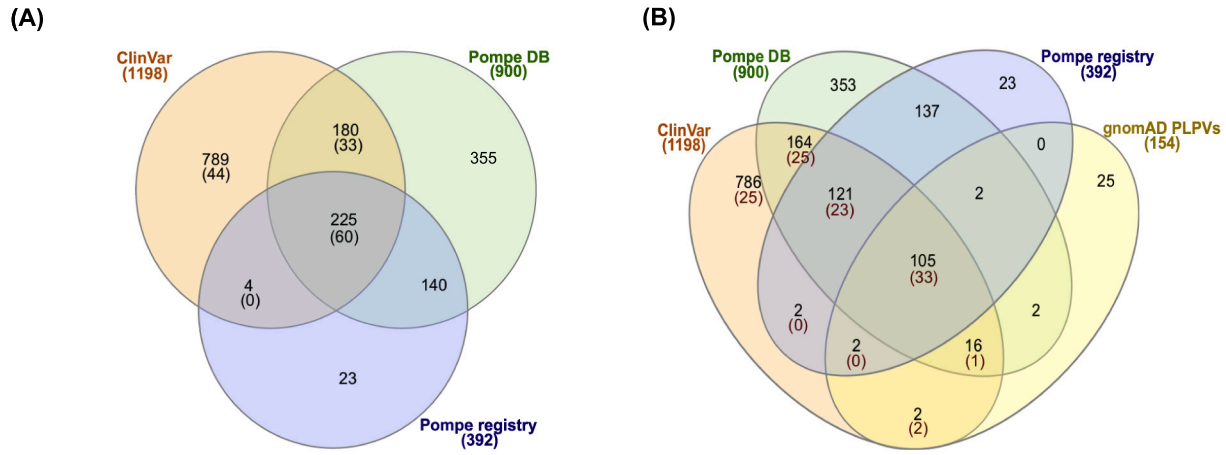

Fig. 1. Genetic variants in gnomAD, ClinVar (including ClinGen), Pompe DB, or Pompe registry. (A) All genetic variants in the ClinVar, Pompe DB, or Pompe registry. The number in parentheses gives the number of genetic variants found in ClinGen. (B) Comparison of PLPVs in gnomAD and genetic variants in ClinVar, Pompe DB, or Pompe registry. The red color number in parentheses gives the number of PLPVs classified by ClinGen. All LSDBs were accessed December 15, 2020.

into PLPVs: 75 variants were (likely) pathogenic variants with a review status of two or more gold stars (two or three) in ClinVar, and the other 79 were (likely) pathogenic variants with a review status of below 2 gold stars (zero or one), variants of uncertain significance, variants with conflicting interpretations of pathogenicity or absent in ClinVar. The ACMG evidence codes for the other 79 variants were described in Table S1. Of the 154 different PLPVs, 125 variants (81.2%) were in both gnomAD and ClinVar, 125 variants (81.2%) were in both gnomAD and Pompe DB, 109 variants (70.8%) were in both gnomAD and Pompe registry, and 105 variants (68.2%) were included in all LSDBs (Fig. 1B).

There were representative 9 PLPVs in global population: (i) which had an allele frequency (AF) greater than 0.0001 in global populations or (ii) an AF greater than 0.0001 in a particular population and also found in multiple populations with AF greater than 0.0001 (Fig. 2, Table 1). The AF of c.-32-13 T > G was highest (AF = 0.0034 in global population), followed by those of c.2238G > C (p.Trp746Cys, AF = 0.00031), c.2560C > T (p.Arg854Ter, AF = 0.000209), c.841C > T (p.Arg281Trp, AF = 0.000205), [c.752C > T; c.761C > T] ([p.Ser251Leu; p.Ser254Leu], AF = 0.00019), 1552-3C > G (AF = 0.00013), 1935C > A (p.Asp645Glu, AF = 0.00012), c.525del (p.Glu176ArgfsTer45, AF = 0.00010) and c.2237G > C (p.Trp746Ser, AF = 0.00007). Notably, a particular population group occupied a dominant proportion (more than 70%) in allele counts of each of representative PLPVs: NFE in allele counts of c.-32-13 T > G (71.7% of total, NFE AF = 0.0053), c.525del (85.2%, NFE AF = 0.00019), c.841C > T (89.5%, NFE AF = 0.00041), 1552-3C > G (92.1%, NFE AF = 0.00027), and c.2238G > C (83.9%, NFE AF = 0.00057); EAS in allele counts of [c.752C > T; c.761C > T] (100% of total, EAS AF = 0.0028), 1935C > A (100%, EAS AF = 0.0017); and AFR in allele counts of c.2560C > T (81.0% of total, AFR AF = 0.0019). Of the 7 populations, the EAS group showed the highest sum of AF of PLPVs with classic infantile type (0.00426), followed by AFR (0.003838), NFE (0.001347), ASJ (0.000975), AMR (0.000957), SAS (0.000939), and FIN (0.000314).

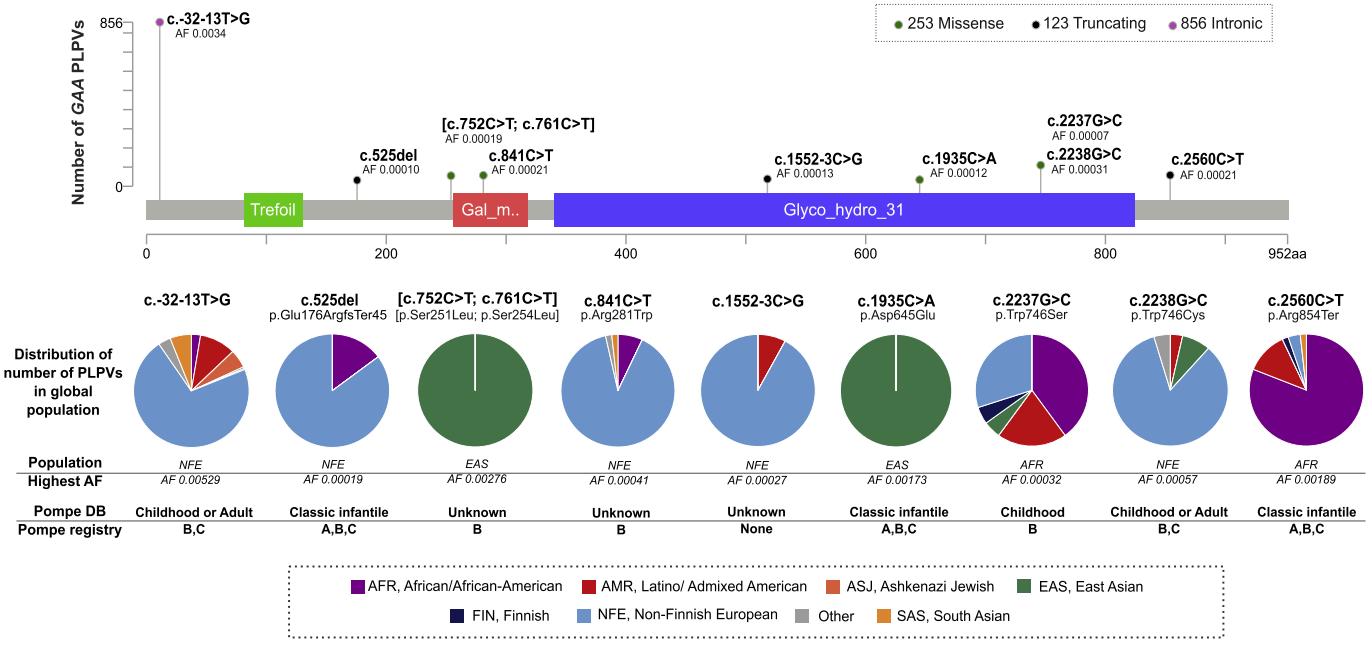

Fig. 2. Representative 9 pathogenic or likely pathogenic variants in the overall population. Trefoil, Trefoil (P type) domain (81–130 amino acids); Gal_m.., Gal_-mutarotas_2: galactose mutarotase-like (256–318 amino acids); Glyco_hydro_31, Glycosyl hydrolases family 31 (340–824 amino acids).

Table 1
Representative pathogenic or likely pathogenic variants with allele frequency greater than 0.0001 and at least 3 or more alleles by population group.

Variant	Pompe DB[a]		Pompe registry [b]		Major population in gnomAD [c]	Allele frequency in gnomAD (v2.1.1), #order of high frequency in population [d]							
	No. of patients	Phenotype with null allele	No. of patients	Group		Global (*N* = 141,456)	AFR (*N* = 12,487)	AMR (*N* = 17,720)	ASJ (*N* = 5185)	EAS (*N* = 9977)	FIN (N = 12,562)	NFE (*N* = 64,603)	SAS (*N* = 15,308)
c.-32-13 T > G	733	CH, AD	689	B,C	Global	0.00340, #1	0.00094, #2	0.00269, #1	0.00554, #1	0.00021, #9	0.00016, #2	0.00529, #1	0.00190, #1
c.525del, p.Glu176ArgfsTer45	154	CI	111	A,B,C	Global	0.00010, #8	0.00017, #5	0	0	0	0	0.00019, #5	0
c.546 + 5G > T	2	Unknown	0	–	EAS	0.00004	0	0	0	0.00050, #3	0	0	0
[c.752C > T; c.761C > T], [p. Ser251Leu; p.Ser254Leu]	26	Unknown	<5	B	Global	0.00019, #5	0	0	0	0.00276, #1	0	0	0
c.841C > T, p.Arg281Trp	0	–	<5	B	Global	0.00021, #4	0.00016, #6	0	0	0	0	0.00041, #3	(0.00003)
c.853C > T, p.Pro285Ser	3	CH, AD	<5	C	AFR	0.00001	0.00013, #7	0	0	0	0	0	0
c.1316 T > A, p.Met439Lys	14	CI	<5	C	EAS	0.00003	0	0	0	0.00038, #4	0	0	0
c.1411_1414del, p. Glu471ProfsTer5	18	CI	<5	A	EAS	0.00002	0	0	0	0.00025, #7	0	0	0
c.1552-3C > G	2	Unknown	0	–	Global	0.00013, #6	0	(0.00008)	0	0	0	0.00027, #4	0
c.1634C > T, p.Pro545Leu	7	CH, AD	7	B,C	FIN	0.00002	0	0	0	0	0.00012, #3	(0.000007)	0
c.1725C > A, p.Tyr575Ter	1	Unknown	0	–	FIN	0.00002	0	0	0	0	0.00020, #1	0	0
c.1843G > A, p.Gly615Arg	9	CI	9	A,B,C	EAS	0.00002	0	0	0	0.00022, #8	0	0	0
c.1856G > A, p.Ser619Asn	4	CH, AD	<5	B	SAS	0.00002	0	0	0	0	0	0	0.00013, #3
c.1935C > A, p.Asp645Glu	104	CI	42	A,B,C	Global	0.00012, #7	0	0	0	0.00173, #2	0	0	0
c.1942G > A, p.Gly648Ser	24	Unknown	7	A,C	SAS	0.00005	(0.00007)	0	0	0	0	(0.000009)	0.00034, #2
c.1979G > A, p.Arg660His	6	CH	9	A,B,C	AFR	0.00004	0.00025, #4	(0.00009)	0	0	0	(0.000008)	0
c.2173C > T, p.Arg725Trp	8	CH, AD	8	A,C	ASJ	0.00004	0	0	0.00077, #2	0	0	(0.00002)	0
c.2237G > C, p.Trp746Ser	1	CH	<5	B	Global	0.00007, #9	0.00032, #3	0.00011, #3	0	(0.00005)	(0.00003)	(0.00005)	0
c.2238G > C, p.Trp746Cys	37	CH, AD	38	B,C	Global	0.00031, #2	0	(0.00008)	0	0.00035, #5	0	0.00057, #2	0
c.2560C > T, p.Arg854Ter	77	CI	45	A,B,C	Global	0.00021, #3	0.00189, #1	0.00020, #2	0	0	(0.00004)	(0.00002)	(0.00003)
c.2662G > T, p.Glu888Ter	26	CI	14	A,B,C	EAS	0.00002	0	0	0	0.00027, #6	0	0	0
c.2815_2816del, p. Val939LeufsTer78	6	CI	5	A,B,C	EAS	0.00001	0	0	0	0.00016, #10	0	0	0

CI, classic infantile; CH, childhood; AD, adult; AFR, African/African-American; AMR, Latino/Admixed American; ASJ, Ashekenazi Jewish; EAS, East Asian; FIN, Finnish; NFE, Non-Finnish European; SAS, South Asian.

[a] In the Pompe DB [8], patients who present symptoms with onset age younger than 12 months and hypertrophic cardiomyopathy were classified as classic infantile Pompe disease, patients who present symptoms with onset age before 18 years and without cardiomyopathy as childhood Pompe disease, and patients with onset age 18 years or later as adult Pompe disease.

[b] In the Pompe registry [9], patients with symptom onset at younger than 12 months of age with cardiomyopathy were classified as group A, patients with symptom onset at 12 years or younger without cardiomyopathy as group B, and patients with onset age 12 years or older as group C.

[c] Global refers to pathogenic or likely pathogenic variants (PLPVs) with allele frequency greater than 0.0001 by global population or representative PLPVs (greater than allele frequency 0.0001) in particular population found also in multiple populations with AF greater than 0.0001.

[d] Allele frequency in parentheses addresses less than 0.0001. N refers to number of individuals included in gnomAD (2.1.1).

3.2. Distribution of carrier frequency (CF) and predicted genetic prevalence (pGP) in each population

Unaffected carriers for *GAA* (the total CF for *GAA)* are predicted to comprise 1.3% of the overall population (Fig. 3A). Of the 7 population groups, the EAS showed the highest CF (1.8%), followed by NFE (1.7%), ASJ (1.3%), AFR(1.2%), AMR (0.8%), SAS (0.7%), and FIN (0.2%).

Overall, the pGP for Pompe disease was 4.30 individuals per 100,000 births (1:23,232) (Fig. 3B). The pGP of the EAS was 8.25 per 100,000 births (1:12,125), followed by NFE (7.27 per 100,000 births, 1:13,756), ASJ (4.38 per 100,000 births, 1:22,851), AFR (3.76 per 100,000 births, 1:26,560), AMR (1.74 per 100,000 births, 1:57,620), SAS (1.07 per 100,000 births, 1:93,087), and FIN (0.09 per 100,000 births, 1:1,056,444).

4. Discussion

In this study, the major questions were how *GAA* variants detected in a generally healthy population reflect the probabilities of developing Pompe disease and how these probabilities change with respect to clinical severity between population groups. To answer these questions, the *GAA* variants found in gnomAD for each population group were analyzed in combination with data from other LSDBs that include the clinical and genotype information of Pompe patients.

Generally, the incidence of Pompe disease is estimated to be 1 in 40,000 live births [22–24] and varies depending on ethnicity and geographic region from 1:14,000 (African American), 1:40,000 (Netherlands and USA), 1:50,000 (Taiwan), 1:145,000 (Australia) to 1:600,000 (Portugal) [22–28]. However, the incidence of Pompe disease reported by newborn screening programs is much higher than those estimated: 1:16,919 in Taiwan [29], 1:8684 in Austria [30], 1:4400 in Hungary [31], and from 1:10,152 to 1:27,581 in USA (Table S2) [32–38]. There was a difference in incidence of Pompe disease between countries, although their major ethnic populations are the same. These differences might be caused by how distribution of ethnic subpopulations in the particular country or geographic region, choice of screening methods, how the cutoff is set for screening, or whether second tier tests such as genetic testing are included in newborn screening programs. Because of those factors, it was difficult to directly compare the incidence of Pompe disease according to population. In the current study, pGP showed wide variations due to population group, ranging from 8.24 per 100,000 births (1:12,125) to 0.09 per 100,000 births (1:1,056,444). Although this study estimated the genetic prevalence (not incidence) of Pompe disease based on the CF in gnomAD (the difference between prevalence and incidence is described in https://www.cureffi.org/2019/06/05/using-genetic-data-to-estimate-disease-prevalence/), these results in this study were somewhat consistent with previous reports on newborn screening programs. The pGP (highest pGP in this study was 1:12,125) of Pompe disease in the EAS group was slightly higher than the real incidence in Taiwan (1:16,919), but the difference was not significant ($P = 0.765$ by z-test). Previous reports [39–43] collectively show that there are differences in most frequent pathogenic variants even within the same Chinese group according to geographic region. Therefore, either prevalence or pGP might be different depending on which subpopulation was more prevalent in gnomAD: here, there were genomic data of a total of 9977 from the EAS group including 1909 Koreans, 76 Japanese, and 7992 other East Asian in gnomAD (V.2.1.1). Notably, the pGP of Pompe disease in FIN was the lowest in this study. The same findings were reported by previous studies in Pompe patients [44,45].

Although Pompe disease is well studied and LSDBs contains a lot of genetic information, only 13% (225 variants/1726 known variants) overlapped in all the 3 LSDBs. Of those 225 variants, 100 variants were classified as PLPVs with a review status of two and more gold stars in ClinVar (data not shown). Among 154 presumed PLPVs in gnomAD, 105 PLPVs (68%) overlapped in all 3 LSDBs, of those, 67 were PLPVs with a review status of two and more gold stars in ClinVar, and 33 were PLPVs with three gold stars (reviewed by ClinGen).

Among presumed PLPVs in gnomAD (v.2.1.1.), c.-32-13 T > G was most frequently found in all population except East Asian. As expected, the c.525del (AF = 0.00019) in NFE, [c.752C > T; c.761C > T] (AF = 0.00276) and c.1935C > A (AF = 0.00173) in EAS, c.2560C > T in AFR (AF = 0.00189) showed high allele frequency. However, there were several unexpected findings for linking population to high AF of PLPVs. The c.525del in AFR (AF = 0.00017) is as frequent as in NFE (AF = 0.00019). In Pompe DB (last accessed December 15, 2020), 154 patients

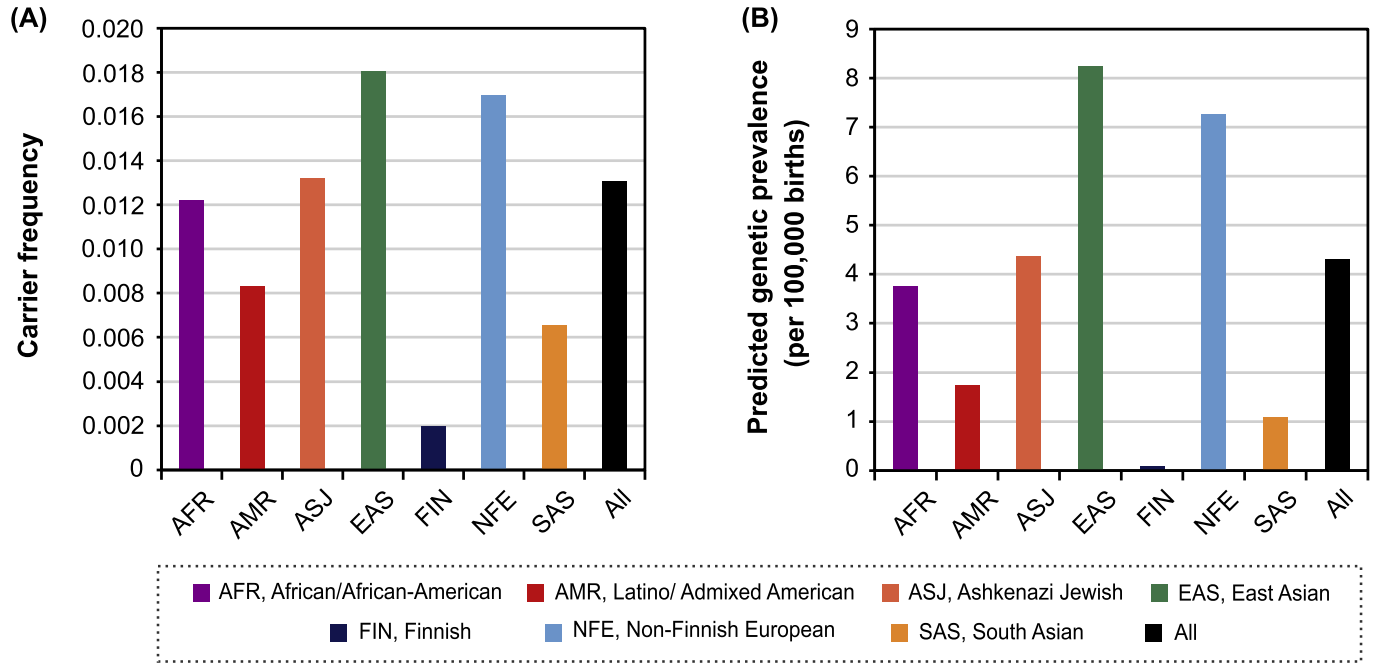

Fig. 3. Distribution of carrier frequency and predicted genetic prevalence in each population. (A) Distribution of carrier frequency in each population. (B) Distribution of predicted genetic prevalence in each population.

with c.525del has been reported, however, only 1 patient was African American. The AF (0.0002) of c.2560C > T in AMR is higher than expected. To date, total of 77 patients with c.2560C > T has been reported in Pompe DB, 18 Pompe patients with c.2560C > T were reported in Latin America. Although the c.841C > T and c.1552-3C > T showed relatively higher AF, clinical reports about these variants were rare. The c.2237G > C (global AF = 0.00007) and c.2238G > C (global AF = 0.00031) were found in multiple populations with high AF. However, only 1 clinical case (Chinese patient with c.2237G > C) has been reported in Pompe DB. Although the c.2238G > C (NFE AF = 0.00057) was the second most frequently found in NFE in gnomAD, there were only 2 European patients' reports among 37 patients (most of them were Asian) in Pompe DB. Because of these unexpected findings, the AF of selected PLPVs in gnomAD between version 2 and version 3 were compared. There was not much difference in some variants (c.2560C > T in AMR (AF = 0.0003926) and AFR (AF = 0.00198) in gnomAD (v3); c.2237G > C in global (AF = 0.000085)), while others differed between the database versions (c.525del in AFR (AF 0.00012) and NFE (AF 0.00032) in gnomAD (v3); c.2238G > C in NFE (AF = 0.00069) and EAS (AF = 0)). The differences between predicted and real development of Pompe disease might be caused by the difference of haplotypes between populations or ethnic subpopulatoins, therefore, resulting in difference of compound heterozygosity, phenotype (whether Pompe disease develops), and clinical severity. In addition, the difference between the two versions (v2 and v3) of gnomAD might be due to the difference in the number of samples between populations, distribution of subpopulations or test methods.

Genetic screening is the process of genetic testing designed to identify those at increased risk of having or developing a disease, with the goal of prevention or early treatment in a specified population [46]. The most common causative variants in Pompe disease reported in previous studies also were found in gnomAD [8,9,39,47]. In addition, a sum of PLPVs AF greater than 0.0001 in 5 population groups (except FIN and SAS groups) represented more than 70% of the total PLPVs AF. These results indicate that genetic screening for Pompe disease may be feasible according to population group.

Knowledge of the *GAA* PLPVs combined with clinical severity in a particular continental population could be valuable for setting up a screening program. When considering screening for Pompe disease (either measuring enzymatic activity or genetic testing), it is important that no classic infantile type be missed. This is particularly important because enzymatic activity of GAA might change depending on combination of PLPVs in each population group [48,49], and this has an effect on the chosen cutoff. Remarkably, the proportions of PLPVs associated with classic infantile type in AFR and EAS groups were higher than those of other populations. The most common PLPVs associated with classic infantile type were c.2560C > T (p.Arg854Ter) in AFR (AF = 0.00189), c.1935C > A (p.Asp645Glu) in EAS groups (AF = 0.00173) and c.525del (p.Glu176ArgfsTer45) in NFE (AF = 0.00019).

In this study, structural variations (SV) or copy number variations (CNV) were not included. A deletion variant of exon 18, c.2481 + 102_2646 + 31del (537 bp deletion, or c.2481 + 110_2646 + 39del, chr17:78091650–78,092,187 based on GRCh37/hg19) was reported as a frequent mutation in Dutch patients [50]. In gnomAD SVs (v2.1), one heterozygous CNV deletion (530 bp deletion, chr17:78091657–78,092,187) was reported in European population: global AF of this variant was 0.000046 (1/21,694) and European AF was 0.00013 (1/7,624). This variant was assumed to be the famous exon 18 deletion. However, it was unclear whether this CNV deletion estimated by CNV caller in gnomAD SVs (v2.1) was same as the variant of c.2481 + 102_2646 + 39del because of difference of genomic position and deletion size. A gnomAD v3 includes only genomic data, therefore, SV reference data is expected to be expanded. If large SV reference data is released through gnomAD v3, further research including SV/ CNV as well as single nucleotide variants, insertion/deletion variants in *GAA* is needed.

Unfortunately, the number of people in each population enrolled in gnomAD does not reflect the real-world population. In addition, the variants with 2 or more gold stars of review status accounts for less than 30% of the total variants in ClinVar (accessed 15 Dec 2020). Additionally, most genetic studies on Pompe disease have been analyzed based on specific populations. Since some real causative variants can be classified as variants of uncertain significance and not as PLPVs due to insufficient clinical, genetic, or functional information, the CF and pGP calculated in the present study might be underestimated.

5. Conclusions

To the best of the author's knowledge, this is the first study that analyzed Pompe disease based on genomic data of the general population and estimated unaffected carriers and genetic prevalence by population. With this novel and alternative approach, Pompe disease (1:23,232) is estimated to be more frequent than formerly accepted (1:40,000). The pGP in the present study support the latest outcome of newborn screening programs (approximately 1:10,000–1:30,000). In addition, the pGP for Pompe disease was wide based on population. The proportions of PLPVs associated with classic infantile type also varied by population and were notably higher in African and East Asian groups. This approach to analyze genomic information of Mendelian disorders in the general population suggests another helpful direction for predictive and preventive medicine.

Author statement

Carrier frequency and predicted genetic prevalence of Pompe disease based on a general population database (MGMREPORTS-D-20-00195).

Kyung Sun Park (Department of Laboratory Medicine, School of Medicine, Kyung Hee University, Seoul, Korea)**:**

Conceptualization, Methodology, Data analysis, Visualization, Original draft preparation, and Writing- Reviewing and Editing.

Funding

This study was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2019R1A2C2002177).

Declaration of Competing Interest

The author has no competing interests.

Acknowledgments

The author thanks Professor Jong-Won Kim at the Department of Laboratory Medicine and Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine, for his valuable comments on this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymgmr.2021.100734.

References

[1] R. Puertollano, N. Raben, Editorial for focused issue "Pompe disease: from basics to current and emerging therapies", Ann. Transl. Med. 7 (2019) 275.
[2] P.S. Kishnani, D. Corzo, N.D. Leslie, D. Gruskin, A. Van der Ploeg, J.P. Clancy, R. Parini, G. Morin, M. Beck, M.S. Bauer, M. Jokic, C.E. Tsai, B.W. Tsai, C. Morgan, T. O'Meara, S. Richards, E.C. Tsao, H. Mandel, Early treatment with alglucosidase alpha prolongs long-term survival of infants with Pompe disease, Pediatr. Res. 66 (2009) 329–335.
[3] C.F. Yang, C.C. Yang, H.C. Liao, L.Y. Huang, C.C. Chiang, H.C. Ho, C.J. Lai, T. H. Chu, T.F. Yang, T.R. Hsu, W.J. Soong, D.M. Niu, Very early treatment for

infantile-onset pompe disease contributes to better outcomes, J. Pediatr. 169 (2016) 174–180 (e171).
[4] O.A. Bodamer, C.R. Scott, R. Giugliani, G. Pompe, Disease newborn screening working, newborn screening for pompe disease, Pediatrics 140 (2017) S4–S13.
[5] S. Richards, N. Aziz, S. Bale, D. Bick, S. Das, J. Gastier-Foster, W.W. Grody, M. Hegde, E. Lyon, E. Spector, K. Voelkerding, H.L. Rehm, A.L.Q.A. Committee, Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology, Genet Med 17 (2015) 405–424.
[6] M. Kroos, R.J. Pomponio, L. van Vliet, R.E. Palmer, M. Phipps, R. Van der Helm, D. Halley, A. Reuser, G.A.A.D. Consortium, Update of the Pompe disease mutation database with 107 sequence variants and a format for severity rating, Hum. Mutat. 29 (2008). E13-26.
[7] M. Kroos, M. Hoogeveen-Westerveld, H. Michelakakis, R. Pomponio, A. Van der Ploeg, D. Halley, A. Reuser, G.A.A.D. 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, Hum. Mutat. 33 (2012) 1161–1165.
[8] M.Y. Nino, S.L.M. In't Groen, A.J. Bergsma, N. van der Beek, M. Kroos, M. Hoogeveen-Westerveld, A.T. van der Ploeg, W. Pijnappel, Extension of the Pompe mutation database by linking disease-associated variants to clinical severity, Hum. Mutat. 40 (2019) 1954–1967.
[9] A.J.J. Reuser, A.T. van der Ploeg, Y.H. Chien, J. Llerena Jr., M.A. Abbott, P. R. Clemens, V.E. Kimonis, N. Leslie, S.S. Maruti, B.J. Sanson, R. Araujo, M. Periquet, A. Toscano, P.S. Kishnani, S. On Behalf Of The Pompe Registry, GAA variants and phenotypes among 1,079 patients with Pompe disease: data from the Pompe Registry, Hum. Mutat. 40 (2019) 2146–2164.
[10] K.J. Karczewski, L.C. Francioli, G. Tiao, B.B. Cummings, J. Alfoldi, Q. Wang, R. L. Collins, K.M. Laricchia, A. Ganna, D.P. Birnbaum, L.D. Gauthier, H. Brand, M. Solomonson, N.A. Watts, D. Rhodes, M. Singer-Berk, E.M. England, E.G. Seaby, J.A. Kosmicki, R.K. Walters, K. Tashman, Y. Farjoun, E. Banks, T. Poterba, A. Wang, C. Seed, N. Whiffin, J.X. Chong, K.E. Samocha, E. Pierce-Hoffman, Z. Zappala, A. H. O'Donnell-Luria, E.V. Minikel, B. Weisburd, M. Lek, J.S. Ware, C. Vittal, I. M. Armean, L. Bergelson, K. Cibulskis, K.M. Connolly, M. Covarrubias, S. Donnelly, S. Ferriera, S. Gabriel, J. Gentry, N. Gupta, T. Jeandet, D. Kaplan, C. Llanwarne, R. Munshi, S. Novod, N. Petrillo, D. Roazen, V. Ruano-Rubio, A. Saltzman, M. Schleicher, J. Soto, K. Tibbetts, C. Tolonen, G. Wade, M.E. Talkowski, C. Genome Aggregation Database, B.M. Neale, M.J. Daly, D.G. MacArthur, The mutational constraint spectrum quantified from variation in 141,456 humans, Nature 581 (2020) 434–443.
[11] M. Lek, K.J. Karczewski, E.V. Minikel, K.E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J.S. Ware, A.J. Hill, B.B. Cummings, T. Tukiainen, D. P. Birnbaum, J.A. Kosmicki, L.E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D.N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M.I. Kurki, A.L. Moonshine, P. Natarajan, L. Orozco, G.M. Peloso, R. Poplin, M.A. Rivas, V. Ruano-Rubio, S.A. Rose, D.M. Ruderfer, K. Shakir, P.D. Stenson, C. Stevens, B. P. Thomas, G. Tiao, M.T. Tusie-Luna, B. Weisburd, H.H. Won, D. Yu, D. M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J. C. Florez, S.B. Gabriel, G. Getz, S.J. Glatt, C.M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M.I. McCarthy, D. McGovern, R. McPherson, B.M. Neale, A. Palotie, S. M. Purcell, D. Saleheen, J.M. Scharf, P. Sklar, P.F. Sullivan, J. Tuomilehto, M. T. Tsuang, H.C. Watkins, J.G. Wilson, M.J. Daly, D.G. MacArthur, C. Exome Aggregation, Analysis of protein-coding genetic variation in 60,706 humans, Nature 536 (2016) 285–291.
[12] A.N. Abou Tayoun, T. Pesaran, M.T. DiStefano, A. Oza, H.L. Rehm, L.G. Biesecker, S.M. Harrison, G., ClinGen sequence variant interpretation working, Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion, Hum. Mutat. 39 (2018) 1517–1524.
[13] N.M. Ioannidis, J.H. Rothstein, V. Pejaver, S. Middha, S.K. McDonnell, S. Baheti, A. Musolf, Q. Li, E. Holzinger, D. Karyadi, L.A. Cannon-Albright, C.C. Teerlink, J. L. Stanford, W.B. Isaacs, J. Xu, K.A. Cooney, E.M. Lange, J. Schleutker, J. D. Carpten, I.J. Powell, O. Cussenot, G. Cancel-Tassin, G.G. Giles, R.J. MacInnis, C. Maier, C.L. Hsieh, F. Wiklund, W.J. Catalona, W.D. Foulkes, D. Mandal, R. A. Eeles, Z. Kote-Jarai, C.D. Bustamante, D.J. Schaid, T. Hastie, E.A. Ostrander, J. E. Bailey-Wilson, P. Radivojac, S.N. Thibodeau, A.S. Whittemore, W. Sieh, REVEL: an ensemble method for predicting the pathogenicity of rare missense variants, Am. J. Hum. Genet. 99 (2016) 877–885.
[14] J.M. Schwarz, D.N. Cooper, M. Schuelke, D. Seelow, MutationTaster2: mutation prediction for the deep-sequencing age, Nat Methods 11 (2014) 361–362.
[15] Y. Choi, A.P. Chan, PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels, Bioinformatics 31 (2015) 2745–2747.
[16] G. Yeo, C.B. Burge, Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals, J. Comput. Biol. 11 (2004) 377–394.
[17] X. Jian, E. Boerwinkle, X. Liu, In silico prediction of splice-altering single nucleotide variants in the human genome, Nucleic Acids Res. 42 (2014) 13534–13544.
[18] H. Heberle, G.V. Meirelles, F.R. da Silva, G.P. Telles, R. Minghim, InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams, BMC Bioinformat. 16 (2015) 169.
[19] S. Vohra, P.C. Biggin, Mutationmapper: a tool to aid the mapping of protein mutation data, PLoS One 8 (2013) e71711.
[20] M. Hanany, G. Allon, A. Kimchi, A. Blumenfeld, H. Newman, E. Pras, O. Wormser, S.B. O, L. Gradstein, E. Banin, T. Ben-Yosef, D. Sharon, Carrier frequency analysis of mutations causing autosomal-recessive-inherited retinal diseases in the Israeli population, Eur. J. Hum. Genet. 26 (2018) 1159–1166.
[21] M. Hanany, C. Rivolta, D. Sharon, Worldwide carrier frequency and genetic prevalence of autosomal recessive inherited retinal diseases, Proc. Natl. Acad. Sci. U. S. A. 117 (2020) 2710–2716.
[22] M.G. Ausems, J. Verbiest, M.P. Hermans, M.A. Kroos, F.A. Beemer, J.H. Wokke, L. A. Sandkuijl, A.J. Reuser, A.T. van der Ploeg, Frequency of glycogen storage disease type II in The Netherlands: implications for diagnosis and genetic counselling, Eur. J. Hum. Genet. 7 (1999) 713–716.
[23] F. Martiniuk, A. Chen, A. Mack, E. Arvanitopoulos, Y. Chen, W.N. Rom, W.J. Codd, B. Hanna, P. Alcabes, N. Raben, P. Plotz, Carrier frequency for glycogen storage disease type II in New York and estimates of affected individuals born with the disease, Am. J. Med. Genet. 79 (1998) 69–72.
[24] B.J. Poorthuis, R.A. Wevers, W.J. Kleijer, J.E. Groener, J.G. de Jong, S. van Weely, K.E. Niezen-Koning, O.P. van Diggelen, The frequency of lysosomal storage diseases in The Netherlands, Hum. Genet. 105 (1999) 151–156.
[25] M. Martinez, M.G. Romero, L.G. Guereta, M. Cabrera, R.M. Regojo, L. Albajara, M. L. Couce, M.S. Pipaon, Infantile-onset Pompe disease with neonatal debut: A case report and literature review, Medicine (Baltimore) 96 (2017) e9186.
[26] R. Pinto, C. Caseiro, M. Lemos, L. Lopes, A. Fontes, H. Ribeiro, E. Pinto, E. Silva, S. Rocha, A. Marcao, I. Ribeiro, L. Lacerda, G. Ribeiro, O. Amaral, M.C. Sa Miranda, Prevalence of lysosomal storage diseases in Portugal, Eur. J. Hum. Genet. 12 (2004) 87–92.
[27] C.Y. Lin, B. Hwang, K.J. Hsiao, Y.R. Jin, Pompe's disease in Chinese and prenatal diagnosis by determination of alpha-glucosidase activity, J. Inherit. Metab. Dis. 10 (1987) 11–17.
[28] P.J. Meikle, J.J. Hopwood, A.E. Clague, W.F. Carey, Prevalence of lysosomal storage disorders, JAMA 281 (1999) 249–254.
[29] S.C. Chiang, W.L. Hwu, N.C. Lee, L.W. Hsu, Y.H. Chien, Algorithm for Pompe disease newborn screening: results from the Taiwan screening program, Mol. Genet. Metab. 106 (2012) 281–286.
[30] T.P. Mechtler, S. Stary, T.F. Metz, V.R. De Jesus, S. Greber-Platzer, A. Pollak, K. R. Herkner, B. Streubel, D.C. Kasper, Neonatal screening for lysosomal storage disorders: feasibility and incidence from a nationwide study in Austria, Lancet 379 (2012) 335–341.
[31] J. Wittmann, E. Karg, S. Turi, E. Legnini, G. Wittmann, A.K. Giese, J. Lukas, U. Golnitz, M. Klingenhager, O. Bodamer, A. Muhl, A. Rolfs, Newborn screening for lysosomal storage disorders in hungary, JIMD Rep. 6 (2012) 117–125.
[32] B.K. Burton, J. Charrow, G.E. Hoganson, J. Fleischer, D.K. Grange, S.R. Braddock, L. Hitchins, R. Hickey, K.M. Christensen, D. Groepper, H. Shryock, P. Smith, R. Shao, K. Basheeruddin, Newborn screening for pompe disease in illinois: experience with 684,290 infants, Int. J. Neonatal. Screen 6 (2020) 4.
[33] T.L. Klug, L.B. Swartz, J. Washburn, C. Brannen, J.L. Kiesling, Lessons learned from pompe disease newborn screening and follow-up, Int. J. Neonatal. Screen 6 (2020) 11.
[34] M.P. Wasserstein, M. Caggana, S.M. Bailey, R.J. Desnick, L. Edelmann, L. Estrella, I. Holzman, N.R. Kelly, R. Kornreich, S.G. Kupchik, M. Martin, S.M. Nafday, R. Wasserman, A. Yang, C. Yu, J.J. Orsini, The New York pilot newborn screening program for lysosomal storage diseases: Report of the First 65,000, Infants Genet. Med. 21 (2019) 631–640.
[35] P.L. Hall, R. Sanchez, A.F. Hagar, S.C. Jerris, A. Wittenauer, W.R. Wilcox, Two-tiered newborn screening with post-analytical tools for pompe disease and mucopolysaccharidosis type i results in performance improvement and future direction, Int. J. Neonatal. Screen 6 (2020).
[36] H. Tang, L. Feuchtbaum, S. Sciortino, J. Matteson, D. Mathur, T. Bishop, R. S. Olney, The first year experience of newborn screening for pompe disease in California, Int. J. Neonatal. Screen 6 (2020) 9.
[37] C. Ficicioglu, R.C. Ahrens-Nicklas, J. Barch, S.R. Cuddapah, B.S. DiBoscio, J. C. DiPerna, P.L. Gordon, N. Henderson, C. Menello, N. Luongo, D. Ortiz, R. Xiao, Newborn screening for pompe disease: Pennsylvania experience, Int. J. Neonatal. Screen 6 (2020).
[38] M.M. Minter Baerg, S.D. Stoway, J. Hart, L. Mott, D.S. Peck, S.L. Nett, J. S. Eckerman, J.M. Lacey, C.T. Turgeon, D. Gavrilov, D. Oglesbee, K. Raymond, S. Tortorelli, D. Matern, L. Morkrid, P. Rinaldo, Precision newborn screening for lysosomal disorders, Genet Med 20 (2018) 847–854.
[39] P. Peruzzo, E. Pavan, A. Dardis, Molecular genetics of Pompe disease: a comprehensive overview, Ann. Transl. Med. 7 (2019) 278.
[40] X. Liu, Z. Wang, W. Jin, H. Lv, W. Zhang, C. Que, Y. Huang, Y. Yuan, 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 (2014) 141.
[41] L. Fu, W. Qiu, Y. Yu, Y. Guo, P. Zhao, X. Zhang, C. Liu, F. Li, H. Huang, M. Huang, S. Chen, Clinical and molecular genetic study of infantile-onset Pompe disease in Chinese patients: identification of 6 novel mutations, Gene 535 (2014) 53–59.
[42] X. Chen, T. Liu, M. Huang, J. Wu, J. Zhu, Y. Guo, X. Xu, F. Li, J. Wang, L. Fu, Clinical and molecular characterization of infantile-onset pompe disease in Mainland Chinese patients: identification of two common mutations genet test, Mol Biomark. 21 (2017) 391–396.
[43] P. Labrousse, Y.H. Chien, R.J. Pomponio, J. Keutzer, N.C. Lee, V.R. Akmaev, T. Scholl, W.L. Hwu, Genetic heterozygosity and pseudodeficiency in the Pompe disease newborn screening pilot program, Mol. Genet. Metab. 99 (2010) 379–383.
[44] M.P. Korpela, A. Paetau, M.I. Lofberg, M.H. Timonen, A.E. Lamminen, S.M. Kiuru-Enari, A novel mutation of the GAA gene in a Finnish late-onset Pompe disease patient: clinical phenotype and follow-up with enzyme replacement therapy, Muscle Nerve 40 (2009) 143–148.

[45] J. Palmio, M. Auranen, S. Kiuru-Enari, M. Lofberg, O. Bodamer, B. Udd, Screening for late-onset Pompe disease in Finland, Neuromuscul. Disord. 24 (2014) 982–985.
[46] A. Andermann, I. Blancquaert, Genetic screening: A primer for primary care, Can. Fam. Physician 56 (2010) 333–339.
[47] K. Momosaki, J. Kido, S. Yoshida, K. Sugawara, T. Miyamoto, T. Inoue, T. Okumiya, S. Matsumoto, F. Endo, S. Hirose, K. Nakamura, 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. 64 (2019) 741–755.
[48] H.C. Liao, M.J. Chan, C.F. Yang, C.C. Chiang, D.M. Niu, C.K. Huang, M.H. Gelb, Mass spectrometry but not fluorimetry distinguishes affected and pseudodeficiency patients in newborn screening for pompe disease, Clin. Chem. 63 (2017) 1271–1277.
[49] S. Elliott, N. Buroker, J.J. Cournoyer, A.M. Potier, J.D. Trometer, C. Elbin, M. J. Schermer, J. Kantola, A. Boyce, F. Turecek, M.H. Gelb, C.R. Scott, Pilot study of newborn screening for six lysosomal storage diseases using Tandem Mass Spectrometry, Mol. Genet. Metab. 118 (2016) 304–309.
[50] M. Van der Kraan, M.A. Kroos, M. Joosse, A.G. Bijvoet, M.P. Verbeet, W.J. Kleijer, A.J. Reuser, Deletion of exon 18 is a frequent mutation in glycogen storage disease type II, Biochem. Biophys. Res. Commun. 203 (1994) 1535–1541.