A decade of molecular diagnosis of Mucolipidosis II and III in Brazil: a pooled analysis of 32 patients

Journal of Inborn Errors of Metabolism & Screening
2021, Volume 9: e20200029
DOI: https://doi.org/10.1590/2326-4594-JIEMS-2020-0029

Nataniel F Ludwig[1,2], Fernanda Sperb-Ludwig[1,2*], Dévora N Randon[1,2], Pricila Bernardi[4], Liane R Giuliani[5], Carolina A Moreno[6], Denise P Cavalcanti[6], Luiz CS da Silva[7] and Ida V D Schwartz[1,2,3]

Abstract

GlcNAc-1-phosphotransferase is a hexameric complex formed by subunits α, β, and γ, where the first two are encoded by the *GNPTAB* gene and the third by the *GNPTG* gene. Pathogenic variants identified in the *GNPTAB* gene cause the diseases Mucolipidosis II and III alpha/beta, which are severe and characterized by an overflow of lysosomal hydrolases into the extracellular environment, and their absence in lysosomal compartments causes an accumulation of non-degraded macromolecules.

Methodology: a retrospective study that included 32 unrelated Brazilian patients with a clinical and genetic diagnosis of Mucolipidosis II/III alpha/beta. The regional frequency of the altered alleles was determined.

Results: The patients were from all regions of Brazil. The most prevalent variants were c.3503_3504del, associated with the severe form of the disease, and c.1208T>C, associated with the milder form. Variant c.3503_3504del is the most frequently found in the Midwest, Northeast, and Southeast regions of Brazil. In the South, 42.8% of the alleles present the c.1196C>T variant.

Conclusions: From the perspective of all patients diagnosed with Mucolipidosis II/III in Brazil, it is possible to conclude that different regions present allelic frequencies of specific pathogenic variants, which can be explained by the occurrence of a founding effect or high inbreeding rates.

Keywords

Genetic diagnosis, mucolipidosis II/III, pooled analysis.

Introduction

The Mucolipidosis (ML) type II, III alpha/beta, and III gamma are autosomal recessive disorders caused by a defect in the pathway of targeting lysosomal enzymes to lysosome through the mannose-6-phosphate (M6P) signal. Lysosomal enzymes synthesized in the Endoplasmic Reticulum are transported to the Golgi complex, where an enzymatical two-step process generates the M6P signal that guarantees the recognition of lysosomal enzymes by M6P receptors in the trans-Golgi network and subsequent delivery to lysosomal compartments [1].

The GlcNAc-1-phosphotransferase, responsible for the first step in the M6P pathway, is a cis-Golgi transmembrane resident and hexameric complex formed by α-, β- and γ-subunits (α2β2γ2), codified by two genes. The *GNPTG* gene is located on chromosome 16p13.3, has 11 exons, and encodes the soluble γ-subunit. The *GNPTAB* gene is located on chromosome 12q23.3, expands to 21 exons that encode a transmembrane precursor

[1] BRAIN Laboratory, Hospital de Clínicas de Porto Alegre, Porto Alegre, RS, Brazil.
[2] Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil, Programa de Pós-Graduação em Genética e Biologia Molecular, Porto Alegre, RS, Brazil.
[3] Serviço de Genética Médica, Hospital de Clínicas de Porto Alegre, Porto Alegre, RS, Brazil.
[4] Universidade Federal de Santa Catarina, Serviço de Genética Médica do Hospital Universitário, Florianópolis, Brazil.
[5] Universidade Federal de Mato Grosso do Sul, Faculdade de Medicina, Campo Grande, MS, Brazil.
[6] Universidade Estadual de Campinas, Faculdade de Ciências Médicas, Departamento de Genética Médica, Campinas, SP, Brazil.
[7] Universidade Federal do Pará, Instituto de Ciências Biológicas, Laboratório de Erros Inatos do Metabolismo, Belém, PA, Brazil.

Received December 7, 2020, and in revised form February 25, 2021. Accepted for publication March 5, 2021.

Corresponding Author:
Fernanda Sperb-Ludwig, Hospital de Clínicas de Porto Alegre, Porto Alegre, RS, Brazil.

Email: fsperb@hcpa.edu.br

protein of 1256 amino acids activated by the Golgi-resident site 1-protease (S1P) proteolytic cleavage into the mature α- and β-subunits [2,3].

Pathogenic variants in *GNPTAB* can be related to severe ML II disease (MIM#252500), in which patients may present symptoms at birth; this disease progresses fast and death occurs in the first years of life. However, pathogenic variants in this gene can also be associated with ML III alpha/beta (MIM#252600), which shows mild symptoms, normally not present at birth. Compared to MLII, the disease progression is slower and life expectancy is longer. ML III gamma (MIM #252605) is caused only by pathogenic variants in the *GNPTG* gene and presents as the milder disease, with joints and bone symptoms and average life expectancy [1,4,5].

The functional characterization of pathogenic variants in the *GNPTAB* gene over the last years demonstrated that levels of 10% of GlcNAc-1-phosphotransferase residual activity appear to protect against the severe phenotype ML II [6]. The spectrum of pathogenic variants described so far highlights the wide distribution of the c.3503_3504del in different patients' populations, mainly to the possibility of a single event in the Mediterranean region of Europe and then spread to European colonized countries, but not Japan and South Korea [7–10]. A recent review of *GNPTAB* genetic data described 258 pathogenic variants in 459 patients. In contrast to c.3503_3504del, 72% of pathogenic variants relate to individual families, and just 10% are found in more than two families, demonstrating the high heterogeneity of this gene [6].

Few studies describe data from the frequency of ML II/III; nevertheless, the current estimated incidence is 1:123.500 in Portugal, 1:650.000 in the Netherlands, and an estimated global incidence of 2.5 to 10 cases per 1.000.000 live births. However, in Quebec, Canada, the high incidence of 1:6.184 for ML II was further explained by a founder effect connected to six founders of the local population. It is the only population with high ML II rates described so far [11–13]. This paper describes new clinical and genetic information of four ML II/III alpha/beta patients and reviews the geographic distribution and frequencies of pathogenic variants identified in the last ten years in Brazil.

Material and Methods

Patients

Twenty-six patients included in this study have already been described in the literature [5,14–18]. Four patients are described herein for the first time (P29 to 32). The consanguinity rate was 18.7% (n= 6/32), totalizing 58 alleles in this study.

Biochemical and molecular diagnoses were performed at the Medical Genetics Service of the Hospital de Clínicas de Porto Alegre (MGS-HCPA), Brazil. The MGS-HCPA is the reference center for the diagnosis of lysosomal diseases in the country and the only research laboratory that performs genotyping of the *GNPTAB* and *GNPTG* genes (this analysis is not available in the public health system). Since 2010, the MGS-HCPA performs research with ML II and III in clinical, biochemical, and genetic data of patients diagnosed with the disease in the country (approved by the local IRB, 2011-0477 and 2019-0374).

Molecular diagnosis of patients P29 to P32

The molecular diagnosis was performed as described elsewhere [14]. In brief, genomic DNA was extracted from blood samples using the Easy-DNA purification kit (Thermo Fisher Scientific). *GNPTAB* gene amplification, performed by standard PCR, was submitted to automated DNA sequencing on an ABI Prism 3500 Genetic Analyzer (Applied Biosystems). Sanger sequencing results analysis was performed using the reference sequence NM_024312.4 and pathogenic nomenclature follows the Human Genome Variation Society's recommendations. Positive samples were twice independently sequenced with forward and reverse primers. The new variants identified were evaluated to determine the pathogenicity following the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG)[19].

Development and validation of molecular diagnosis protocol

To perform molecular diagnosis faster and more cost-effectively, we use retrospective *GNPTAB* genetic data of 26 patients (Table 1 – P1 to P26) to define a rational, sequential order of exon sequencing, where the most frequent exons with pathogenic variants were first sequenced. These 26 patients have the entire coding region of the *GNPTAB* gene sequenced. Next, we use a prospective cohort (Table 1, P27 to P32) to validate the protocol.

Results

Non-published patients

P29 and P30 were clinically diagnosed with ML III alpha/beta; the first patient presents the missense variant c.1196C>T (p.Ser399Phe) in homozygosity. The second is compound heterozygous for the small deletion c.3503_3504del and the missense variant c.1208T>C (p.Ile403Thr). P31 is a ML II patient homozygous to c.3503_3504del (p.Leu1168Glnfs*5). P32 is a ML II patient diagnosed with two variants identified for the first time, c.2034dupT (p.Phe678Phefs*1) and c.2720_2721del (p.Glu907Glyfs*11) (Table 1). The variants identified were not found in genetic databases (gnomAD and 1000genomes); however, frameshift variants cause the disease's severe phenotype, which agrees with the clinical diagnosis. Using ACMG recommendations for the classification of variants, both were considered pathogenic (following criteria: PVS1, PM2, and PP4).

Table 1. The Clinical, geographical and genetic information of the 32 Brazilian patients included in this work.

Patient ID	Phenotype	Consanguinity	Region of Origin	Genotype Allele 1/ Allelle2		Exon	Reference
				cDNA	Protein		
1	II	No	Northeast	c.[3503_3504del];[3503_3504del]	p.[Leu1168Glnfs*5];[Leu1168Glnfs*5]	19/19	Cury *et al.*, 2013
2	II	No	Midwest	c.[3503_3504del];[3503_3504del]	p.[Leu1168Glnfs*5];[Leu1168Glnfs*5]	19/19	Cury *et al.*, 2013
3	II	No	Midwest	c.[2808A>G];[3503_3504del]	p.[Tyr937_Met972del];[p.Leu1168Glnfs*5]	14/19	Cury *et al.*, 2013
4	II	Yes	Northeast	c.[3503_3504del];[3503_3504del]	p.[Leu1168Glnfs*5];[Leu1168Glnfs*5]	19/19	Cury *et al.*, 2013
5	III	No	Southeast	c.[1514G>A];[1759C>T]	p.[Cys505Tyr];[(Arg587*)]	12/13.1	Cury *et al.*, 2013
6	II	No	Southeast	c.[2269_2273del];[2269_2273del]	p.[(Glu757Lysfs*2)];[(Glu757Lysfs*2)]	13.2/13.2	Cury *et al.*, 2013
7	III	No	South	c.[1196C>T];[3503_3504del]	p.[Ser399Phe];[Leu1168Glnfs*5]	10/19	Cury *et al.*, 2013
8	III	No	Northeast	c.[1208T>C];[3503_3504del]	p.[Ile403Thr];[Leu1168Glnfs*5]	19/10	Cury *et al.*, 2013
9	III	No	Northeast	c.[3503_3504del];[?]	p.[Leu1168Glnfs*5];[(?)]	19/ ?	Cury *et al.*, 2013
10	II	No	Southeast	c.[1123C>T];[?]	p.[Arg375*];[(?)]	10/ ?	Cury *et al.*, 2013
11	III	No	Southeast	c.[1931C>T];[3668_3670del]	p.[Thr644Met];[(Thr1223del)]	13.1/20	Velho *et al.*, 2015
12	III	No	Southeast	c.[1208T>C];[1723G>A]	p.[Ile403Thr];[Gly575Arg]	10/13.1	Sperb-Ludwig *et al.*, 2015
13	II	Yes	Southeast	c.[3503_3504del];[3503_3504del]	p.[Leu1168Glnfs*5];[Leu1168Glnfs*5]	19/19	Ludwig *et al.*, 2017
14	II	No	Southeast	c.242G>T(;)2249dup	p.Trp81Leu(;)Asn750Lysfs*8	3/13.2	Ludwig *et al.*, 2017
15	II	No	Southeast	c.[1154C>T];[3503_3504del]	p.[Ser385Leu];[Leu1168Glnfs*5]	19/10	Ludwig *et al.*, 2017
16	II	No	Southeast	c.2249dup(;)3503_3504del	p.Asn642Leufs*10(;)Leu1168Glnfs*5	19/13.2	Ludwig *et al.*, 2017
17	II	No	South	c.[831del];[3503_3504del]	p.[(Phe277Phefs*3)];[Leu1168Glnfs*5]	19/8	Ludwig *et al.*, 2017
18	II	No	Southeast	c.[242G>T];[1924_1927del]	p.[Trp81Leu];[Asn642Leufs*10]	3/13.1	Ludwig *et al.*, 2017
19	III	No	Southeast	c.[1208T>C];[832C>T]	p.[Ile403Thr];[(Gln278*)]	10/8	Ludwig *et al.*, 2017
20	II	Yes	Northeast	c.[3503_3504del];[3503_3504del]	p.[Leu1168Glnfs*5];[Leu1168Glnfs*5]	19/19	Ludwig *et al.*, 2017
21	II	No	Northeast	c.1763dup(;)3333T>G	p.(His588Glnfs*26)(;)(Try1111*)	13.1/17	Ludwig *et al.*, 2017
22	III	No	Northeast	c.1514G>A(;)3503_3504del	p.Cys505Tyr(;)Leu1168Glnfs*5	19/12	Ludwig *et al.*, 2017
23	III	No	Northeast	c.[1208T>C];[3503_3504del]	p.[Ile403Thr];[Leu1168Glnfs*5]	19/10	Ludwig *et al.*, 2017
24	III	No	Northeast	c.[1208T>C];[3503_3504del]	p.[Ile403Thr];[Leu1168Glnfs*5]	19/10	Ludwig *et al.*, 2017
25	II	Yes	Northeast	c.227A>G(;)227A>G	p.Asp76Gly(;)Asp76Gly	3/3	Ludwig *et al.*, 2017
26	III	No	Southeast	c.[242G>T];[?]	p.[Trp81Leu];[(?)]	3/ ?	Alegra *et al.*, 2019
27	II	No	Southeast	c.[2249dup];[2249dup]	p.[Asn750Lysfs*8];[Asn750Lysfs*8]	13/13	Ludwig, 2020
28	III	No	South	c.[1196C>T];[2757_2758del]	p.[Ser399Phe];[(Asp919Glufs*21)]	10/14	Ludwig, 2020
29	III	Yes	South	c.1196C>T(;)1196C>T	p.Ser399Phe(;)Ser399Phe	10/10	This study
30	III	No	Midwest	c.1208T>C(;)3503_3504del	p.Ile403Thr(;)Leu1168Glnfs*5	19/10	This study
31	II	Yes	Southeast	c.[3503_3504del];[3503_3504del]	p.[Leu1168Glnfs*5];[Leu1168Glnfs*5]	19/19	This study
32	II	No	North	**c.2034dup(;)2720_2721del**	p.(Phe678Phefs*1)(;)(Glu907Glyfs*11)	13/14	This study

Pathogenic variants being described for the first time are presented in bold.

Characterization of the total cohort of Brazilian patients

Of the 32 patients included in this report, eighteen (56.2%) were clinically diagnosed with ML II; the identification of both variants using Sanger sequencing was 90.6% (n=29/32 patients). In the total cohort, we identified 23 unique pathogenic variants that vary from frameshift (n= 10/23; 43.4%), missense (n= 8/23; 34.7%), nonsense (n= 4/23; 17.3%), and splicing variant (n= 1/23; 4.3%) (Figure 1 and Table 1). The pathogenic variants that account for the higher number of affected alleles is the small deletion in exon 19, c.3503_3504del (p.Leu1168Glnfs*5) (n= 19/58; 32.7%), followed by the missense c.1208T>C (p.Ile403Thr) (n= 6/58; 10.3%), the frameshift c.2249dup (p.Asn750Lysfs*8) (n=4/58; 6.8%), the missense c.242G>T (p.Trp81Leu) and c.1196C>T (p.Ser399Phe) (n=3/58; 5.1%) and the missense c.1514G>A (p.Cys505Tyr) (n=2/58; 3.4%). The other pathogenic variants were identified in only one allele.

Pathogenic variants were mainly associated with the severe form of the disease ML II (n=14/23; 60.8%), including all frameshift, the nonsense c.3333T>G (p.Try1111*), and the splicing c.2808A>G (p.Tyr937_Met972del) variants. The nonsense variants c.832C>T (p.Gln278*) and c.1759C>T (p.Arg587*) are expected to generate a premature stop codon in the amino acids 278 and 587, respectively, and to be associated with ML II. However, P19 and P5 are compound heterozygous to the missense c.1208T>C (p.Ile403Thr) and c.1514G>A (p.Cys505Tyr), respectively, which can explain the mild ML III alpha/beta phenotype observed. Of the eight missense variants found in Brazil, the data indicate that genotypes containing these variant types can be associated with both phenotypes, the mild ML III alpha/beta (n=5/8 variants, 62.5%; patients 5,7,11,12,18,19, 22-24,26, 28 and 30), and the severe ML II (n=3/8 variants, 37.5%; patients 14, 15 and 25). The missense variant c.242G>T (p.Trp81Leu) was found in the ML II P14 and P18 in compound heterozygous with c.2249dup (p.Asn750Lysfs*8) and c.1924_1927del (p.Asn642Leufs*10), respectively. We also identified it in the ML III alpha/beta P26, but no other variant was identified.

The geographical distribution of the pathogenic variants' shows that some alleles were specific to some regions in the country (Figure 2). For instance, the missense variant c.1196C>T (p.Ser399Phe) accounts for three alleles found just in the Southern region of the country, specifically in the state of SC (P29 is homozygous with consanguineous parents). The four alleles of the small duplication c.2249dup (p.Asn750Lysfs*8) were identified in the state of SP, localized in the Southeast region, where P27 was born from non-consanguineous parents and diagnosed as homozygous for this variant. In this same region, the missense c.242G>T(p.Trp81Leu) was identified in three alleles from three different patients, although in two distinct states (SP and MG).

The two pathogenic variants that account for the higher number of affected alleles presented a broader distribution in the country. The worldwide identified c.3503_3504del (p.Leu1168Glnfs*5) variant was found spread-out in the country, corresponding to 52.9% (n= 9/17) and 66.7% (n= 4/6) of alleles in the Northeast and Midwest regions, respectively, but not found in the Northern region of the country. We found the c.1208T>C (p.Ile403Thr) variant in the Midwest, Southeast and Northwest; in the last, we identify a high variant frequency (n=3/6). We identify the compound heterozygous genotype c.3503_3504del/c.1208T>C in four patients, three in the Northeast region (P8, P23 and P24).

Next, we use the gnomAD (version 2.1.1) database to search for allelic frequencies of the six recurrent pathogenic variants identified in the Brazilian ML II/III population. We found that all populations, except in the Ashkenazi Jewish and East Asian, present the c.3503_3504del (p.Leu1168Glnfs*5) variant, with the highest frequencies in South Asian and Non-Finish Europeans, followed by Latinos and Others (Table S1).

Mainly in the non-Finnish European population, we identify the c.1196C>T (p.Ser399Phe) in five alleles (asides, one allele from South Asia), the c.1514G>A (p.Cys505Tyr) with three alleles, the c.2249dup (p.Asn750Lysfs*8) with two alleles (North-western Europeans, Table S2), and the c.1208T>C (p.Ile403Thr) identified once (Table S1). We have not identified the variant c.242G>T (p.Trp81Leu) in the gnomAD database. We found two alleles of the c.3503_3504del (0.000854) variant in the ABraOM database (a cohort of 1171 healthy elderly Brazilians) [20].

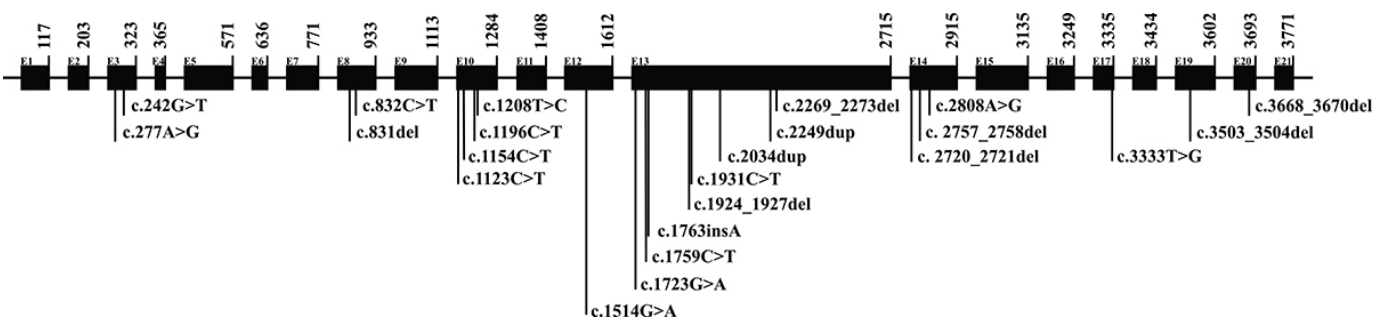

Figure 1. Pathogenic variants in a cohort of Brazilian patients with ML II and III alpha/beta. Black boxes represent exons of the *GNPTAB* gene in proportional size, and the numbers on the top of the boxes represent the number of that exon (horizontal) and the last nucleotide of that exon (vertical). **E**: exon.

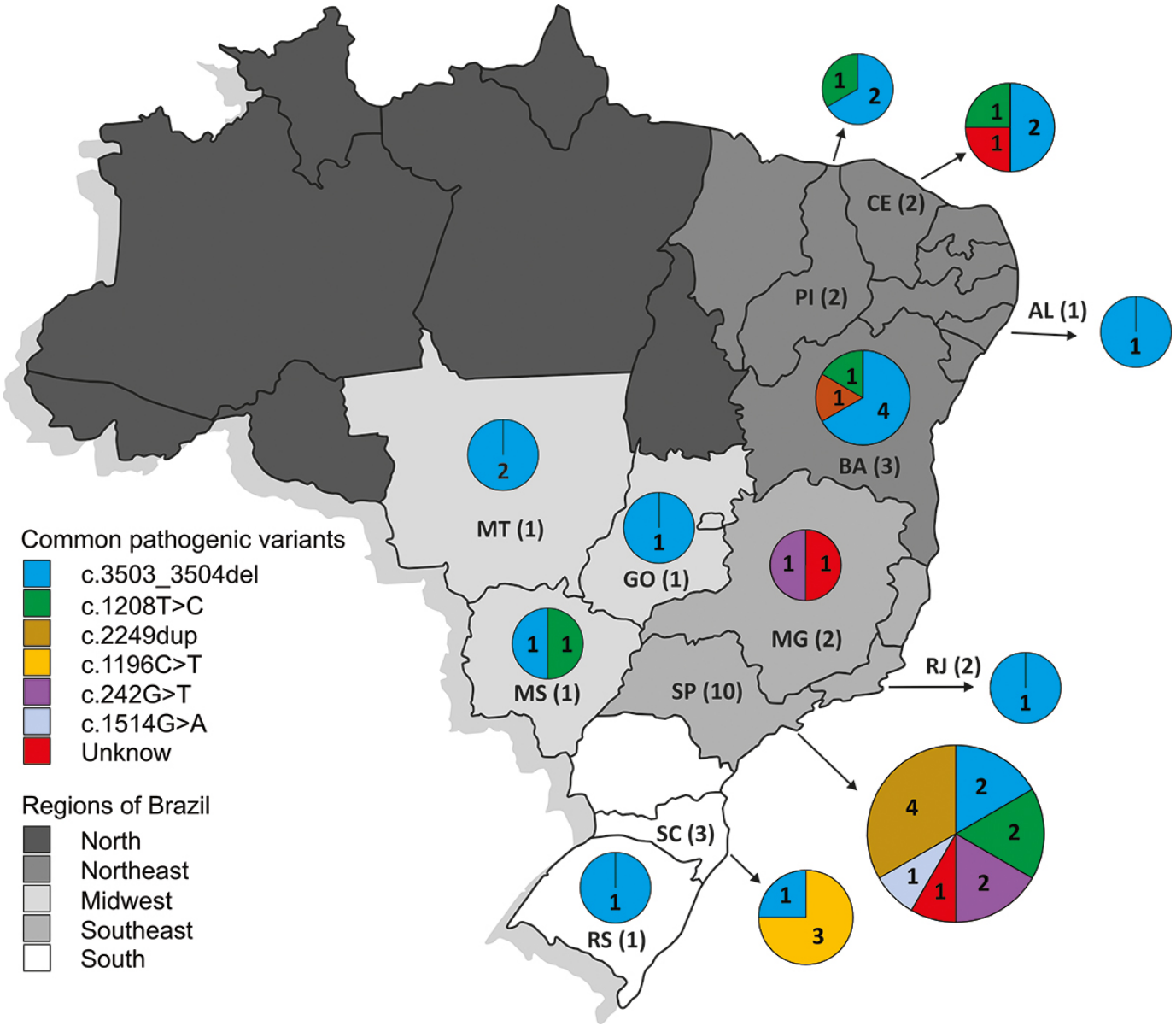

Figure 2. The geographical distribution of recurrent pathogenic variants identified in the 32 Brazilian ML II and III alpha/beta patients. The five regions of the country are shown with different background colors. The two-letter code identified the state, and the number of patients from that state are represented between brackets. The Figure displays 39/58 of the total number of alleles since the private variants were not shown. CE, Ceará; PI, Piauí; AL, Alagoas; MT, Mato Grosso; BA, Bahia; MS, Mato Grosso do Sul; MG, Minas Gerais; SP, São Paulo; RJ, Rio de Janeiro; SC, Santa Catarina; RS, Rio Grande do Sul.

A protocol for GNPTAB genetic diagnosis

Since the *GNPTAB* gene coding sequence presented as 85.461 base pairs long divided through 21 exons, we used retrospective *GNPTAB* data from 26 patients, from whom the whole gene Sanger sequencing data were available, to evaluate the exons with higher frequencies of pathogenic variants. We identified pathogenic variants in 10 exons (Table 1 – P1 to P26) with distinct frequencies (exon 19, 17/48 or 35.4%; exon 13, 9/48 or 18.7%; exon 10, 8/48 or 16.6%; exon 3, 4/48 or 8.3%; exons 8 and 12, 2/48 or 4.1% each; exons 14, 17 and 20, 1/48 or 2% each). Using these data, we listed the most frequent exons in descending order and grouped those with similar frequencies, and proposed protocol 1 (Figure 3). There are three alleles with no pathogenic variant identified. Following the protocol, we performed the molecular diagnosis of patients P27 to P32. We concluded the diagnosis of all six patients in the first four steps, with the mean number of exons analyzed of 4.6 (which vary from one exon in P31 - homozygous for variant in exon 19 - to nine exons in P28 and P32, sequenced until step 4).

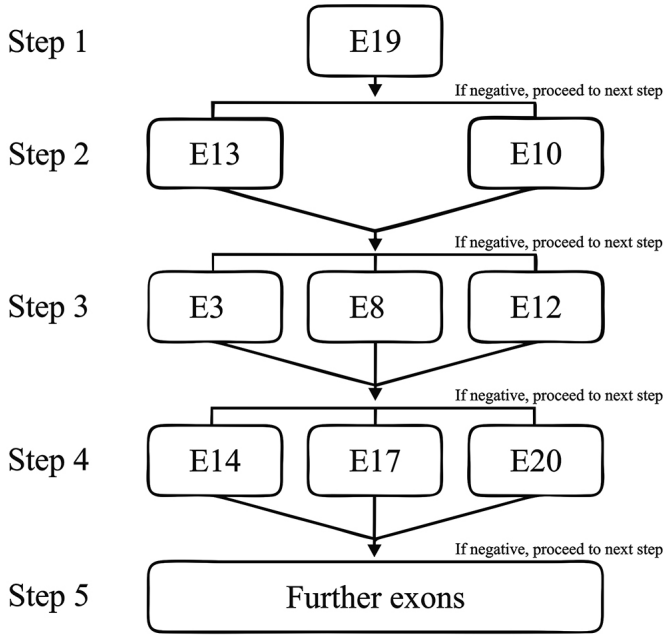

Figure 3. The sequential protocol for molecular diagnosis of Brazilian ML II and III alpha/beta patients. The allelic frequencies of pathogenic variants identified in 26 patients were used to generate the protocol by steps, where the exons with higher number of pathogenic alleles were analyzed first. The numbers in the boxes represent the exons of *GNPTAB* gene and those grouped in the same step were analyzed simultaneous.

Discussion

Here we described genetic and clinical information of four new patients diagnosed with ML II and III alpha/beta in Brazil, with the protocol newly established. This novel information, pooled with data of 26 patients' previously published by our group, enables the analysis of frequencies and distribution of pathogenic variants in Brazil. The grouped analysis allows us to design and validate the protocol to conclude the molecular diagnosis efficiently, which is a powerful tool to deliver the information for genetic counseling of affected families in a country with limited resources.

The analyses confirm the expected high frequency of the worldwide spread c.3503_3054del (p.Leu1168Glnfs*5) variant in the Brazilian population, which accounts for 32.7% of affected alleles found in all regions of the country. Which is confirmed by healthy individual databases that present high frequencies of this variant in different genetic backgrounds, although absent in East Asian populations as previously described [8,9]. The high proportion of this variant in the Northeast region of the country agrees with recent data from haplotypes analysis, demonstrating six different haplotypes containing the variant in this region of Brazil [10]. The aforementioned sustains the hypothesis that migration events of European populations', responsible for the majority of Brazilian genetic background [21], introduced the variant, which diversified through mutational and recombinant events [10].

Most pathogenic variants identified in the *GNPTAB* gene are rare [6]; however, we found high frequencies of a subset of specific variants in Brazil. Distributed in the Southeast, Midwest, and Northeast of the country, the missense c.1208T>C (p.Ile403Thr), the second most identified variant, counts 10.3% of affected alleles (n=6/58). Despite the low frequency in the gnomAD database (Table S1 and S2), the variant was identified in three patients from European populations, from Portugal [22], Italy [23], and Spain/Portugal [24]. Even though these countries were significant contributors to Brazilian genetic background, the frequency of this variant in patient populations from Europe and the healthy databases are insufficient to explain the high proportion in Brazilian patients.

Further, the c.2249dup (p.Asn750Lysfs*8) variant described in a patient from Finland [22], a well-known country with a distinct genetic background [25], was identified in four alleles in Brazil (6.8%). In the gnomAD, however, just the North-western European population presents the variant. The detection of the variant c.1196C>T (p.Ser399Phe) in three alleles of Brazilian patients and five alleles of the gnomAD database (South Asia and non-Finnish Europeans) indicates a broad distribution. In previously published patients, this variant was detected in Portugal (n=3)[22,26], France (n=1)[27], Belgium/Romania (n=1) [28] and in the USA (different origins, n=3) [29].

Interesting to note that the four alleles of the c.2249dup (p.Asn750Lysfs*8) and the three alleles of c.1196C>T (p.Ser399Phe) were identified, in each case, specifically in patients from the same state; the first variant in three non-consanguineous patients

from SP, and the second in three patients, one consanguineous, from SC. Clusters of isolated populations with a high prevalence of autosomal recessive disorders, most likely attributed to the elevated rates of consanguinities, have been identified in the Northeast and Southeast regions of Brazil [30,31]. Further studies to address the question whether the high rates of affected alleles identified in these areas could be related to or constitute one of these clusters are needed.

ML II/III alpha/beta diseases molecular diagnosis is a time and resource-consuming task, which involves sequencing two highly heterogeneous genes regarding pathogenic variants, therefore presenting a considerable amount of private variants (in Brazil, n=18/24, 75%) [6]. Besides, the rate of successful molecular diagnosis with Sanger sequencing achieves ~95% worldwide [1], but recent works demonstrate the necessity of utilizing a complementary approach to conclude exceptional cases such as allelic drop [26], large delins, and splicing variants [6]. To improve the efficiency of the molecular diagnosis in Brazil, we implemented a simple protocol to perform the *GNPTAB* gene investigation that prioritizes exons with more variants. Our results with six new samples show that the patient molecular diagnosis' conclusion requires fewer exons to be sequenced.

In conclusion, our effort to study the genetic characteristics of the Brazilian ML II/III alpha/beta patients in the last decade provide data to achieve an overview of *GNPTAB* pathogenic variant frequencies and specific distributions. These efforts highlight distinct and private frequencies in different states of the country, which support the hypothesis of founder effect or high rates of consanguinity, or a combination of both. Future studies will address this question, highlighting the importance of molecular studies to provide cost-effective and accurate information for patient genetic counseling.

Acknowledgments

The present study was financed by the Brazilian National Council for Scientific and Technological Development (CNPq), the Brazilian Federal Agency for the Support and Evaluation of Graduate Education (CAPES) and the Fund for the Incentive of Research of Hospital de Clínicas de Porto Alegre (FIPE-HCPA).

Authors' Contributions

NFL, FSL and IVDS participated in the design of the project. NFL, FSL, DNR and IVDS generated and analyzed the data. PB, LRG, CAM, DPC, LCSS and IVDS provided clinical assistance. NFL, FSL and IVDS write the manuscript. All authors read and approved the final manuscript.

Declaration of Conflict of Interests

None Declared.

Supplementary Material

The following online material is available for this article:

Table S1 – Allelic frequencies in healthy individuals' database of pathogenic variants identified in more than one allele in Brazil.
Table S2 – Frequencies of pathogenic variants in sub-populations of Non-Finnish Europeans.

References

1. Braulke T, Raas-Rothschild A, Kornfeld S. I-Cell Disease and Pseudo-Hurler Polydystrophy: Disorders of Lysosomal Enzyme Phosphorylation and Localization. In: Valle DL, Antonarakis S, Ballabio A, Beaudet AL, Mitchell GA, eds. *The Online Metabolic & Molecular Bases of Inherited Disease*. McGraw-Hill; 2015: chap 138. doi: 10.1036/ommbid.167
2. Tiede S, Storch S, Lübke T, *et al*. Mucolipidosis II is caused by mutations in GNPTA encoding the alpha/beta GlcNAc-1-phosphotransferase. *Nat Med*. 2005;11(10):1109-1112. doi: 10.1038/nm1305
3. Raas-Rothschild A, Cormier-Daire V, Bao M, *et al*. Molecular basis of variant pseudo-hurler polydystrophy (mucolipidosis IIIC). *J Clin Invest*. 2000;105(5):673-681. doi: 10.1172/JCI5826
4. Cathey S, Kudo M, Tiede S, *et al*. Molecular Order in Mucolipidosis II and III Nomenclature. *Am J Med Genet A*. 2008;143A(18):2106-2112. doi: 10.1002/ajmg.a
5. Alegra T, Sperb-Ludwig F, Guarany NR, *et al*. Clinical Characterization of Mucolipidoses II and III: A Multicenter Study. *J Pediatr Genet*. 2019;8(4):198-204. doi: 10.1055/s-0039-1697605
6. Velho RV, Harms FL, Danyukova T, *et al*. The lysosomal storage disorders mucolipidosis type II, type III alpha/beta and type III gamma: Update on GNPTAB and GNPTG mutations. *Hum Mutat*. 2019;40:842-864. doi: 10.1002/humu.23748
7. Coutinho M, Encarnação M, Gomes R, *et al*. Origin and spread of a common deletion causing mucolipidosis type II: insights from patterns of haplotypic diversity. *Clin Genet*. 2011;80(3):273-280. doi: 10.1111/j.1399-0004.2010.01539.x
8. Otomo T, Muramatsu T, Yorifuji T, *et al*. Mucolipidosis II and III alpha/beta: mutation analysis of 40 Japanese patients showed genotype-phenotype correlation. *J Hum Genet*. 2009;54(3):145-151. doi: 10.1038/jhg.2009.3
9. Paik KH, Song SM, Ki CS, *et al*. Identification of mutations in the GNPTA (MGC4170) gene coding for GlcNAc-phosphotransferase α/β subunits in Korean patients with mucolipidosis type II or type IIIA. *Hum Mutat*. 2005;26(4):308-314. doi: 10.1002/humu.20205
10. Soares MB, Turchetto-Zolet AC, Schwartz IVD, Sperb-Ludwig F. Haplotype analysis and origin of the most

common pathogenic mutation causing Mucolipidosis II and III alpha/beta in Brazilian patients. *Gene Reports*. 2020;19:100645. doi: 10.1016/j.genrep.2020.100645

11. Pinto R, Caseiro C, Lemos M, *et al*. Prevalence of lysosomal storage diseases in Portugal. *Eur J Hum Genet*. 2004;12(2):87-92. doi: 10.1038/sj.ejhg.5201044

12. Plante M, Claveau S, Lepage P, *et al*. Mucolipidosis II: a single causal mutation in the N-acetylglucosamine-1-phosphotransferase gene (GNPTAB) in a French Canadian founder population. *Clin Genet*. 2008;73(3):236-244. doi: 10.1111/j.1399-0004.2007.00954.x

13. Raas-Rothschild A, Pohl S, Braulke T. Multiple Enzyme Deficiencies: Defects in transport: Mucolipidosis II alpha/beta; mucolipidosis III alpha/beta and mucolipidosis III gamma. In: Mehta, AB, Winchester B (eds). L*ysosomal Storage Diseases: A Practical Guide*. Wiley-Blackwell; 2012: 121-126

14. Cury GK, Matte U, Artigalás O, *et al*. Mucolipidosis II and III alpha/beta in Brazil: analysis of the GNPTAB gene. *Gene*. 2013;524(1):59-64. doi: 10.1016/j.gene.2013.03.105

15. Sperb-Ludwig F, Alegra T, Velho RV, *et al*. Exome sequencing for mucolipidosis III: Detection of a novel GNPTAB gene mutation in a patient with a very mild phenotype. *Mol Genet Metab Reports*. 2015;2:34-37. doi: 10.1016/j.ymgmr.2014.12.001

16. Velho RV, De Pace R, Sarah Klünder, *et al*. Analyses of disease-related GNPTAB mutations define a novel GlcNAc-1- phosphotransferase interaction domain and an alternative site-1 protease cleavage site. *Hum Mol Genet*. 2015;24(12):3497-3505. doi: 10.1093/hmg/ddv100

17. Ludwig NF, Voltolini Velho R, Sperb-ludwig F, *et al*. GNPTAB missense mutations cause loss of GlcNAc-phosphotransferase activity in mucolipidosis tipe II through distinct mechanisms. *Int J Biochem Cell Biol*. 2017;92:90-94.

18. Ludwig NF. *Análise abrangente de variantes de GNPTAB associadas à atividade deficiente da GlcNAc-1-fosfotransferase*. Doctoral Dissertation. Universidade Federal do Rio Grande do Sul; 2020.

19. Richards S, Aziz N, Bale S, *et al*. 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*. 2015;17(5):405-424. doi: 10.1038/gim.2015.30

20. Naslavsky MS, Scliar MO, Yamamoto GL, *et al*. Whole-genome sequencing of 1,171 elderly admixed individuals from the largest Latin American metropolis (São Paulo, Brazil). *bioRxiv*. 2020.09.15.298026. Published online January 1, 2020. doi: 10.1101/2020.09.15.298026

21. Rodrigues de Moura R, Coelho AVC, de Queiroz Balbino V, Crovella S, Brandão LAC. Meta-analysis of Brazilian genetic admixture and comparison with other Latin America countries. *Am J Hum Biol*. 2015;27(5):674-680. doi: 10.1002/ajhb.22714

22. Encarnação M, Lacerda L, Costa R, *et al*. Molecular analysis of the GNPTAB and GNPTG genes in 13 patients with mucolipidosis type II or type III - identification of eight novel mutations. *Clin Genet*. 2009;76(1):76-84. doi: 10.1111/j.1399-0004.2009.01185.x

23. Tappino B, Chuzhanova N, Regis S, *et al*. Molecular characterization of 22 novel UDP-N-acetylglucosamine-1-phosphate transferase - α and β-subunit (GNPTAB) gene mutations causing mucolipidosis types IIα/β and IIIα/β in 46 patients. *Hum Mutat*. 2009;30(11):956-973. doi: 10.1002/humu.21099

24. Fernández-Marmiesse A, Morey M, Pineda M, *et al*. Assessment of a targeted resequencing assay as a support tool in the diagnosis of lysosomal storage disorders. *Orphanet J Rare Dis*. 2014;9:59. doi: 10.1186/1750-1172-9-59

25. Lim ET, Würtz P, Havulinna AS, *et al*. Distribution and Medical Impact of Loss-of-Function Variants in the Finnish Founder Population. *PLoS Genet*. 2014;10(7):e1004494. doi: 10.1371/journal.pgen.1004494

26. Coutinho MF, Encarnação M, Laranjeira F, Lacerda L, Prata MJ, Alves S. Solving a case of allelic dropout in the GNPTAB gene: implications in the molecular diagnosis of mucolipidosis type III alpha/beta. *J Pediatr Endocrinol Metab*. 2016;29(10):1225-1228. doi: 10.1515/jpem-2016-0173

27. Bargal R, Zeigler M, Abu-Libdeh B, *et al*. When Mucolipidosis III meets Mucolipidosis II: GNPTA gene mutations in 24 patients. *Mol Genet Metab*. 2006;88(4):359-363. doi: 10.1016/j.ymgme.2006.03.003

28. Gheldof A, Seneca S, Stouffs K, *et al*. Clinical implementation of gene panel testing for lysosomal storage diseases. *Mol Genet Genomic Med*. 2019;7(2):e00527. doi: 10.1002/mgg3.527

29. Cathey S, Leroy JG, Wood T, *et al*. Phenotype and Genotype in Mucolipidoses II and III alpha/beta: A Study of 61 Probands. *J Med Genet*. 2010;47(1):38-48. doi: 10.1136/jmg.2009.067736.

30. Cardoso GC, de Oliveira MZ, Paixão-Côrtes VR, Castilla EE, Schuler-Faccini L. Clusters of genetic diseases in Brazil. *J Community Genet*. 2019;10(1):121-128. doi: 10.1007/s12687-018-0369-1

31. Castilla EE, Schuler-Faccini L. From rumors to genetic isolates. *Genet Mol Biol*. 2014;37(1):186-193. doi: 10.1590/S1415-47572014000200005