



New genotype-phenotype correlations in a large European cohort of patients with sarcoglycanopathy patients.

Jorge Alonso-Pérez, Claudio Semplicini, Lidia González-Quereda, Michela Guglieri, Volker Straub, Pia Gallano Petit, Elena Pegoraro, Andrés Nascimento Osorio, Carlos Ignacio Orteza Gonzalez, Giacomo Comi, et al.

► To cite this version:

Jorge Alonso-Pérez, Claudio Semplicini, Lidia González-Quereda, Michela Guglieri, Volker Straub, et al.. New genotype-phenotype correlations in a large European cohort of patients with sarcoglycanopathy patients.. *Brain - A Journal of Neurology* , 2020, 143 (9), pp.2696-2708. 10.1093/brain/awaa228 . hal-03003854v2

HAL Id: hal-03003854

<https://hal.science/hal-03003854v2>

Submitted on 29 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

New genotype-phenotype correlations in a large international cohort of sarcoglycanopathy patients

Jorge Alonso-Pérez¹, Claudio Semplicini², Lidia González-Quereda^{3,4}, Michela Guglieri⁵, Volker Straub⁵, Pia Gallano Petit^{3,4}, Elena Pegoraro², Andrés Nascimiento Osorio⁶, Carlos Ignacio Ortez González⁶, Giacomo Comi⁷, Marianne Devisser⁸, AJ van der Kooi⁸, Cristina Garrido⁹, Manuela Santos⁹, Ulrike Schara¹⁰, Andrea Gangfuß¹⁰, Nicoline Løkken¹¹, John Vissing¹¹, Benedikt Schoser¹², Bjarne Udd¹³, Adele D'Amico¹⁴, Luisa Politano¹⁵, Claudio Bruno¹⁶, Anna Sarkozy¹⁷, Omar Abdel-Mannan¹⁷, Alicia Alonso-Jimenez¹⁸, Kristl G Claeys¹⁹, David Gomez-Andrés²⁰, Francina Munell²⁰, Jana Haberlová²¹, De Vos Elke²², Jan De Bleecker²², Cristina Dominguez-González^{4,23}, Giorgio Tasca²⁴, Claudia Weiss²⁵, Nicolas Deconinck²⁶, Roberto Fernández Torrón²⁷, Adolfo López de Munain Arregui²⁷, Ana Camacho-Salas²⁸, Melegh Belá²⁹, Hadzsiev Kinga²⁹, Lea Leonardis³⁰, Blaz Koritnik³⁰, Matteo Garibaldi³¹, Juan Carlos de Leon-Hernández³², Edoardo Malfatti³³, Arturo Fraga-Bau³⁴, Isabelle Richard³⁵, Isabel Illa^{1,4}, Jordi Díaz-Manera^{1,4,5}

¹ Neuromuscular Diseases Unit, Department of Neurology, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain.

² Department of Neuroscience, University of Padova, Italy.

³ U705 and U703 CIBERER, Genetics Department, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Spain.

⁴ Centro de Investigación Biomédica en Red en Enfermedades Raras (CIBERER), Spain.

⁵ John Walton Muscular Dystrophy Research Centre, University of Newcastle, Newcastle Upon Tyne, United Kingdom

⁶ Neuromuscular Disorder Unit. Hospital Sant Joan de Deu, Barcelona, Spain.

⁷ Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Dino Ferrari Centre, University of Milan, Milan, Italy.

⁸Department of Neurology, Amsterdam UMC, University of Amsterdam, Amsterdam Neuroscience, Amsterdam, The Netherlands.

⁹Neuropediatric Department, Centro Hospitalar do Porto, Porto, Portugal.

¹⁰Department of Paediatric Neurology, University Hospital Essen, Essen, Germany.

¹¹Copenhagen Neuromuscular Center, Department of Neurology, Rigshospitalet and University of Copenhagen, Copenhagen, Denmark.

¹²Friedrich-Baur-Institute, Department of Neurology Klinikum München Ludwig-Maximilians-University Munich, Munich, Germany.

¹³Neuromuscular Research Center, University of Tampere and Tampere University Hospital, Tampere, Finland.

¹⁴Unit of Neuromuscular and Neurodegenerative Diseases, Department of Neurosciences, Bambino Gesù Children's Hospital, Rome, Italy.

¹⁵Cardiology and Medical Genetics, Department of Experimental Medicine, University of Campania, Naples, Italy.

¹⁶Center of Translational and Experimental Myology, Istituto Giannina Gaslini, Genova, Italy.

¹⁷Dubowitz Neuromuscular Centre, MRC Centre for Neuromuscular Diseases, UCL Great Ormond Street Institute of Child Health, London, UK.

¹⁸Neurology Department, Antwerpen, Universitair Ziekenhuis Antwerpen.

¹⁹Department of Neurology, University Hospitals Leuven, Herestraat 49, 3000, Louvain, Belgium. Laboratory for Muscle Diseases and Neuropathies, Department of Neurosciences, KU Leuven, Louvain, Belgium.

²⁰Paediatric Neurology, Vall d'Hebron University Hospital and VHIR (Euro-NMD, ERN-RND), Barcelona, Spain

²¹Department of Child Neurology, Charles University, 2nd Medical School, University Hospital Motol, Prague, Czech Republic.

²²Department of Neurology, Ghent University and University Hospital Ghent, Ghent, Belgium.

²³Neuromuscular Unit, department of Neurology, Hospital Universitario 12 de Octubre, Instituto de Investigación i+12, Madrid, Spain.

²⁴UOC Neurologia, Fondazione Policlinico Universitario A. Gemelli IRCCS, Roma, Italy.

²⁵Department of Neuropediatrics, Charité - Universitätsmedizin Berlin, Berlin, Germany.

²⁶Department of Neurology, Queen Fabiola Children's University Hospital, Free University of Brussels, Brussels, Belgium.

²⁷Neurosciences, BioDonostia Health Research Institute, Hospital Donostia, San Sebastián, Spain.

²⁸Division of Child Neurology, Hospital Universitario 12 de Octubre, Avenida de Córdoba s/n, 28041 Madrid, Spain.

²⁹Department of Medical Genetics, and Szentagothai Research Center, University of Pécs, Pécs, Hungary.

³⁰Division of Neurology, Institute of Clinical Neurophysiology, University Medical Centre Ljubljana, Ljubljana, Slovenia.

³¹Department of Neurosciences, Mental Health and Sensory Organs NESMOS, Università di Roma "La Sapienza", Rome, Italy.

³²Neurology Department, Hospital Universitario Nuestra Señora de la Candelaria, Tenerife, Spain.

³³Neurology Department, Raymond-Poincaré teaching hospital, centre de référence des maladies neuromusculaires Nord/Est/Ile-de-France, AP-HP, Garches, France

³⁴Neurology Department, Álvaro Cunqueiro Hospital, Vigo, Spain.

³⁵Integrare (UMR_S951), Inserm, Généthron, Univ Evry, ³⁵Université Paris-Saclay, 91002, Evry, France.

Corresponding author: Dr. Jordi Díaz-Manera, The John Walton Muscular Dystrophy Research Centre, University of Newcastle, Newcastle Upon Tyne, United Kingdom Central Parkway

Statistical analysis was performed by Jorge Alonso-Pérez and Jordi Díaz-Manera.

Submission type: Article

Title Count:

Abstract Count:

Paper Count:

Number of Tables: 2

Number of Figures: 5

Study funding: This investigation was sponsored by a grant from the Spanish Ministry of Health, Fondos FEDER-ISCIII PI18/01525 to Dr. Jordi Díaz-Manera.

Financial disclosures: All authors report no relevant disclosures

Keywords: limb girdle muscular dystrophies, sarcoglycan

Abstract

Background: Sarcoglycanopathies comprise four subtypes of autosomal recessive limb-girdle muscular dystrophies (LGMDR3, LGMDR4, LGMDR5 and LGMDR6) that are caused, respectively, by mutation in the *SGCA*, *SGCB*, *SGCG* and *SGCD* genes. In 2016, several clinicians involved in the diagnosis, management and care of patients with LGMDR 3-6 created a European sarcoglycanopathy consortium. The aim of the present study reported by this consortium was to determine the clinical and genetic spectra of a large cohort of European patients with sarcoglycanopathies.

Methods: A total of 33 neuromuscular centers agreed to establish a retrospective observational study for sarcoglycanopathies in Europe. All sarcoglycanopathy patients followed in participating centers underwent an extensive clinical evaluation or review of medical records. Demographic, genetic and clinical data were collected for this study.

Results: Data from a total of 439 patients from 13 different countries were collected. An additional forty-three patients were not enrolled in the study because insufficient genetic and clinical information. A total of 159 patients had a confirmed diagnosis of LGMDR3, 73 of LGMDR4, 157 of LGMDR5 and 7 of LGMDR6. Patients with LGMDR3 had a later onset and slower progression of the disease. Cardiac involvement was most frequent in LGMDR4. Sixty per cent of LGMDR3 patients carried one of the following mutations, either in a homozygous or heterozygous state: c.229C>T (p.R77C), c.739G>A or c.850C>T. Similarly, the most common mutations in LMGDR5 patients were c.525delT or c.848G>A. In LGMDR4 patients, the most frequent mutation was c341C>T. In all sarcoglycanopathy groups, patients with disease onset after 10 years of age and residual protein expression of 30% or more had a better prognosis.

Conclusions: This study reports clinical, genetic and protein data of a large European cohort of sarcoglycanopathy patients. Improving our knowledge about these extremely rare autosomal recessive forms of LGMD was helped by a collaborative effort of neuromuscular centers across Europe. Our study provided important data on the genotype-phenotype correlation that is relevant for the design of natural history studies and upcoming interventional trials in sarcoglycanopathies.

INTRODUCTION

Limb girdle muscular dystrophies (LGMD) are a group of genetic disorders characterized by progressive muscle weakness and wasting of the shoulder and pelvic girdle muscles leading to increasing disability¹. Mutations in thirty-two genes have been reported to cause different types of LGMD, with the sarcoglycanopathies being one of the most frequent forms^{2–4}.

There are four different sarcoglycan genes associated with autosomal recessive LGMD (LGMDR): *SGCA*, *SGCB*, *SGCD* and *SGCG* encoding for the alpha-, beta-, delta- and gamma-sarcoglycan proteins, respectively. Two other sarcoglycan genes, *SGCE* and *SGCZ* are not expressed in muscle and have not been associated to any muscular condition so far. Sarcoglycans are transmembrane glycoproteins that form a tetrameric complex across the cell membrane of skeletal and cardiac muscle fibers^{5–7}. They play an important role in maintaining muscle membrane integrity during cycles of contraction and relaxation through association with the dystroglycan complex, which links the subsarcolemmal protein dystrophin to the basement membrane^{7,8}. Mutations in any of the 4 sarcoglycan genes disrupts the whole complex, leading to a loss of muscle membrane integrity^{9,10}.

The sarcoglycan genes were identified in the 90s, and, since then, several studies have reported the clinical features of small cohorts of patients^{2,11–16}. Most of the patients develop progressive proximal muscle weakness starting during the first decade of life, although other, rarer phenotypes, such as asymptomatic hyperCKemia or exercise intolerance, have also been reported^{17–19}. In general, sarcoglycanopathies are severe disorders characterized by loss of ambulation during adolescence or in early adulthood, however, milder cases, in which patients are still ambulant at 50 or 60 years, have been described^{20,21}. Cardiac and respiratory muscle involvement is frequently part of the clinical picture^{22–24}. Several studies have tried to establish a correlation between genetic findings and phenotypes. It has been suggested that mutations leading to the absence of protein expression cause a severe phenotype characterized by early

loss of ambulation, but it is not entirely clear if partial loss of expression is associated with milder phenotypes^{14,25}. Moreover, even if there are certain clinical similarities between the different types of sarcoglycanopathies, their relative frequency and the natural history of the individual subtypes is not fully understood.

To answer these questions, we decided to collect demographic, genetic, clinical and muscle protein data of a large cohort of patients with mutations in the sarcoglycan genes that were followed in neuromuscular centers across Europe. Our aim was to describe the main clinical and genetic features of every disease subtype, investigate potential correlations between genotype and phenotype, and identify factors influencing the progression of the disease.

MATERIALS AND METHODS

Study design

This study was designed as an observational retrospective study collecting clinical and genetic data from patients with confirmed mutations in the sarcoglycan genes (*SGCA*, *SGCB*, *SGCD* and *SGCG*). The study was approved by institutional review boards and written informed consent was obtained from patients or parents in the participating centers.

A total of 62 neuromuscular centers in Europe were contacted and invited to collaborate in the study, with 33 centers contributing data to our study.

Patient cohort

The inclusion criteria for the study were: 1) genetically confirmed diagnosis of sarcoglycanopathy by identification of two heterozygous or one homozygous pathogenic mutation in one of the sarcoglycan genes, 2) sufficient clinical data available in the records to answer questions about disease onset and progression, presence or absence of cardiac and respiratory involvements and the muscle function status at last visit. To analyze all the data collected, patients were categorized by the mutated gene in four groups, LGMDR3, R4, R5 and R6, and then further stratified by the presence or absence of the different symptoms.

Data sources

All participating centers completed a survey with the data of every patient followed at the center (A copy of the survey is provided as supplemental information 1). The survey collected demographic, clinical and genetic data and also information about muscle biopsy if performed, including morphological features and levels of protein expression measured by western-blot. Cardiac involvement was defined by an ejection fraction lower than 50%, the existence of morphological abnormalities in the ventricular walls evaluated by echocardiography or the

existence of cardiac conduction defects²⁶. The need of ventilatory support was also recorded as well as the age at which it was started.

Mutations in the *SCGA*, *SCGB*, *SCGG* or *SCGD* genes were collected and centrally analysed by experienced geneticists from Hospital Santa Creu I Sant Pau (L.G-Q and P.G.) to predict pathogenicity. We considered nonsense, frameshift and canonical splice site mutations as incompatible with full length protein production since these mutations will cause a disruption of the reading frame or a considerable shortening of the transcript resulting in aberrant protein degradation. Missense variants and mutations that maintained the reading frame were considered as compatible with protein production. Based on our experience, deep intronic mutations were also considered as compatible with protein production, since a mutation affecting splicing, but not located in a canonical splice site, usually produces a certain amount of wild type transcript.

In order to reduce the selection bias, patients with more than 30% of missing data were excluded from the study.

Statistical analysis

Quantitative variables were analyzed using the Kolmogorov-Smirnov test, to see if they showed normal distribution. Comparison between the different subgroups of patients was performed using the ANOVA test for quantitative variables and the Chi-Squared test for categorical variables. We used a Cox proportional hazard regression model for the analyses of time to wheelchair. In order to select which variables were associated with loss of ambulation before 18 years old, we performed a two-step analysis. First, we performed a univariate analysis using the Chi-Squared test for categorical variables and the Student's T Test for quantitative variables. Those variables that showed a significant different distribution among groups (considering in this case $p < 0.1$) were included in a multivariable logistic regression modelling and backward eliminated until reaching significance. A ROC curve was generated to test the predictive power

of the final model and an area under the curve (AUC) was provided. We also performed a ROC analysis to identify the most balanced cut-point of age. P was considered significant if lower than 0.05. Statistics analysis was performed by J.D-M. and J.A-P. using SPSS software version 20 from IBM.

RESULTS

Patient cohort

A total of 439 patients were collected (STARD diagram - Figure 1). Forty-three patients were not included in the study because they did not have sufficient genetic or clinical data to complete the survey. We divided the 396 patients from 329 different families in 4 different subgroups depending on the mutated gene (Figure 1): 159 patients were classified as LGMDR3 (from now defined as alpha patients), 73 as LGMDR4 (from now beta patients), 157 patients as LGMDR5 (from now gamma patients), and 7 as LGMDR6 (from now delta patients). We excluded the LMDR6 patients from the general comparisons between groups due to the low number of patients.

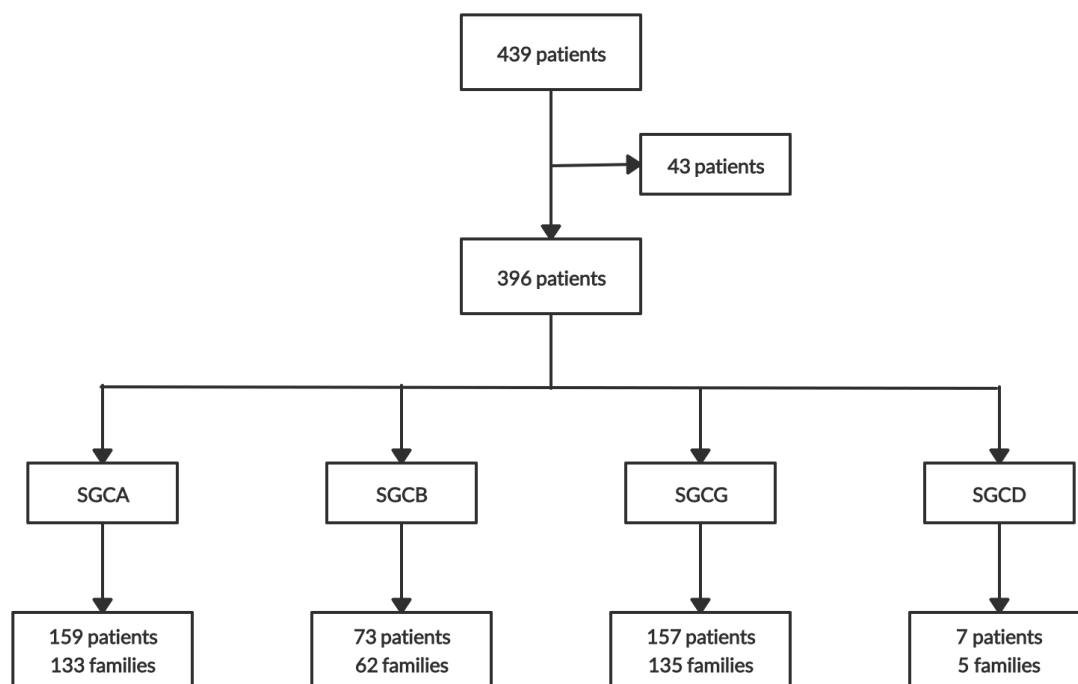
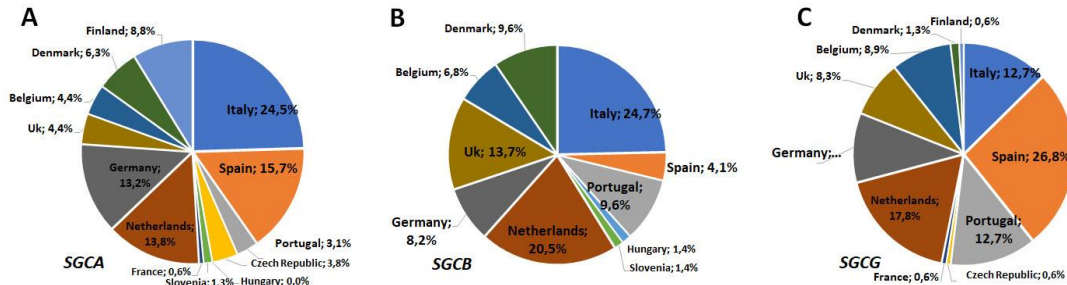


Figure 1 STARD diagram

Patient are from 33 hospitals in 13 European countries. We did not identify statistical differences in the geographic distribution of subgroups around Europe. However, beta patients were more prevalent in Northern countries, such as The Netherlands, United Kingdom and Denmark, while gamma patients were more prevalent in Southern countries, such as Spain, Italy and Portugal

(Supplemental Figure 1B-C). Most of the patients were of Caucasian origin (91.9% alpha, 84.5% beta, 45% gamma and 57.1% delta). In the case of gamma-patients, 27.7% had a Gypsy background and 17.8% a Maghreb background.



Supplemental Figure 1

There were not significant differences in gender in any of the subgroups studied (ANOVA, $p=0.69$) (Table 1). Thirty seven percent of alpha, 28.8% of beta, 43.3% of gamma and, 57.1% of delta patients had relatives affected by the disease. Consanguinity was reported in 15.5%, 25%, 49.2% and 85.7% of cases respectively.

	SGCA	SGCB	SGCG
Nº of patients	159	73	157
Sex, M/F (uk)	70/75 (14 uk)	36/37	77/80
Families, n	133	62	135
Related Family, n (%)	60 (37,7)	21 (28,8)	68 (43,3)
Symptomatic, n (%)	148 (93,1)	73 (100)	152 (96,8)
Age onset, average $\pm$ SD (range)	10,2 $\pm$ 9 (1 - 45)	5,8 $\pm$ 4,1 (1 - 22)	5,8 $\pm$ 3,4 (1 - 20)
Diagnostic delay, average $\pm$ SD(range)	14,1 $\pm$ 13,9 (-2 - 53)	14,3 $\pm$ 13,4 (1 - 50)	9,2 $\pm$ 11,6 (-5 - 50)
Age at last evaluation, average $\pm$ SD (range)	28,9 $\pm$ 17,8 (2 - 74)	28,3 $\pm$ 15,9 (8 - 66)	19,9 $\pm$ 13 (2 - 61)
Disease evolution (y), average $\pm$ SD (range)	19,1 $\pm$ 13,6 (0 - 54)	21,5 $\pm$ 14,2 (4 - 60)	14,7 $\pm$ 11,6 (0 - 50)
Symptom onset, n (%)	102 (46 uk)	46 (27 uk)	107 (45 uk)
Proximal LL weakness	45 (44,1)	25 (54,3)	53 (49,5)
Proximal UL weakness	30 (20,3)	9 (19,6)	15 (14,0)
HyperCKemia	27 (18,2)	13 (28,3)	28 (26,2)
Muscle pain	26 (25,5)	8 (17,4)	23 (21,5)
Symptom, n (%); average $\pm$ SD (range)	87 (61 uk)	35 (28 uk)	103 (49 uk)
Tendon contraction	48 (55,2)	25 (71,4)	79 (76,7)
Tiptoe gait	20 (22,9); 7,4 $\pm$ 5,4 (3 - 25)	23 (65,7); 8,9 $\pm$ 5,9 (2 - 25)	38 (36,9); 5,7 $\pm$ 3 (1 - 11)
Calf hypertrophic	50 (57,5); 12,4 $\pm$ 11,5 (3 - 51)	21 (60); 6,1 $\pm$ 2,6 (2 - 12)	66 (64,1); 6,4 $\pm$ 3,7 (2 - 16)
Muscle pain	32 (36,8); 9,4 $\pm$ 8,3 (3 - 40)	10 (28,6); 6 $\pm$ 4,4 (2 - 15)	35 (34); 7,5 $\pm$ 5,3 (1 - 25)
Scoliosis	29 (33,3); 16 $\pm$ 7,5 (6 - 35)	17 (48,6); 17,6 $\pm$ 13,6 (7 - 37)	43 (39,8); 14,7 $\pm$ 8 (7 - 31)
Scapular winging	39 (44,8);	18 (51,4); 14,9 $\pm$ 13 (4 - 43)	41 (39,8); 12,5 $\pm$ 7,3 (2 - 30)
Motor function n (%); average age $\pm$ SD (range)	129 (19 uk)	65 (8 uk)	135 (17 uk)
Running	15 (11,7); 13,1 $\pm$ 4,6 (5 - 19)	2 (3,2); 16 $\pm$ 2,8 (14 - 18)	10 (7,7); 8,4 $\pm$ 8,7 (2 - 9)
Walk without aids	30 (23,4); 29,7 $\pm$ 18,5 (4 - 63)	7 (10,9); 19,3 $\pm$ 11,6 (8 - 34)	25 (19,2); 12,2 $\pm$ 8,7 (4 - 48)
WCB	72 (56,3); 18,6 $\pm$ 11,9 (6 - 62)	50 (78,1); 16,7 $\pm$ 9,5 (5 - 45)	87 (66,9); 14,6 $\pm$ 7 (8 - 42)
Cardiac involment n (%); average age $\pm$ SD (range)	14 (10,4); 34,1 $\pm$ 14,4 (11 - 55)	30 (41,1); 25,8 $\pm$ 13,1 (12 - 55)	32 (22,5); 21,1 $\pm$ 11,2 (5 - 48)
Heart rhythm abnormalities, n (%)	10 (71,4)	4 (13,3)	6 (18,8)
Dilated cardiomyopathy, n (%)	1 (7,1)	19 (63,3)	19 (59,4)
Unknown type of cardiac involment, n (%)	3 (21,4)	7 (23,3)	7 (21,8)
Respiratory support n (%); average age $\pm$ SD (range)			
Noninvasive ventilation	40 (27); 29 $\pm$ 11 (12 - 58)	21 (28,8); 29,8 $\pm$ 9,2 (12 - 44)	35 (23,6); 28,7 $\pm$ 11,4 (8 - 58)

Table 1 Demographic***Symptoms and muscle function status.***

Ninety-six percent of the patients were symptomatic at the time of inclusion in the study (Table 1). Non-symptomatic patients were diagnosed either because they were relatives of patients with the disease or because they had asymptomatic hyperCKemia. Age of onset of symptoms varied between the different subtypes: alpha patients experienced first symptoms later than beta and gamma patients (ANOVA test, $p<0.001$), while there were not differences in the age of onset between beta and gamma patients (Figure 2A). Delay in diagnosis, defined as the time from onset of symptoms to the genetic confirmation of the disease varied from one subtype to the other, but it was significantly shorter in gamma patients compared to alpha and beta patients (ANOVA test, $p=0.004$) (Figure 2B). The most frequent symptom at onset in all subgroups was proximal lower limb weakness. However, about 25% of all patients were diagnosed due to the presence of asymptomatic hyperCKemia, muscle pain or because they were relatives of already diagnosed patients (Table 1).

At the time of the survey, 23.4% of all alpha patients were able to walk without aids while 56.3% were wheelchair bound (WCB). Only 10.9% of beta patients were able to walk without aids while 78.1% of them were WCB. In the case of gamma patients, 19.2% were able to walk without aids and 66.9% were WCB (Table 1). There were not significant differences in the age at loss of ambulation between subgroups (ANOVA, $p=0.053$), but we identified a non-significant trend for an earlier loss of ambulation in beta and gamma patients compared to alpha patients (Figure 2C). We identified a significant correlation between the age of onset of symptoms and the age at loss of ambulation (Pearson correlation, $p<0.001$, $r=0.61$) (Figure 2D). Interestingly, there were significant differences in the age at loss of ambulation when patients were classified as age of onset of symptoms before and after 10 years old (Mantel-Cox test, $p<0.001$) both when the analysis was performed with all sarcoglycan patients and when it was performed by every

subgroup (Figure 3). Furthermore, we observed that those patients who started with muscle weakness after 18 years old (n:20) did not lose ambulation during the follow-up assessments (Supplemental figure 2).

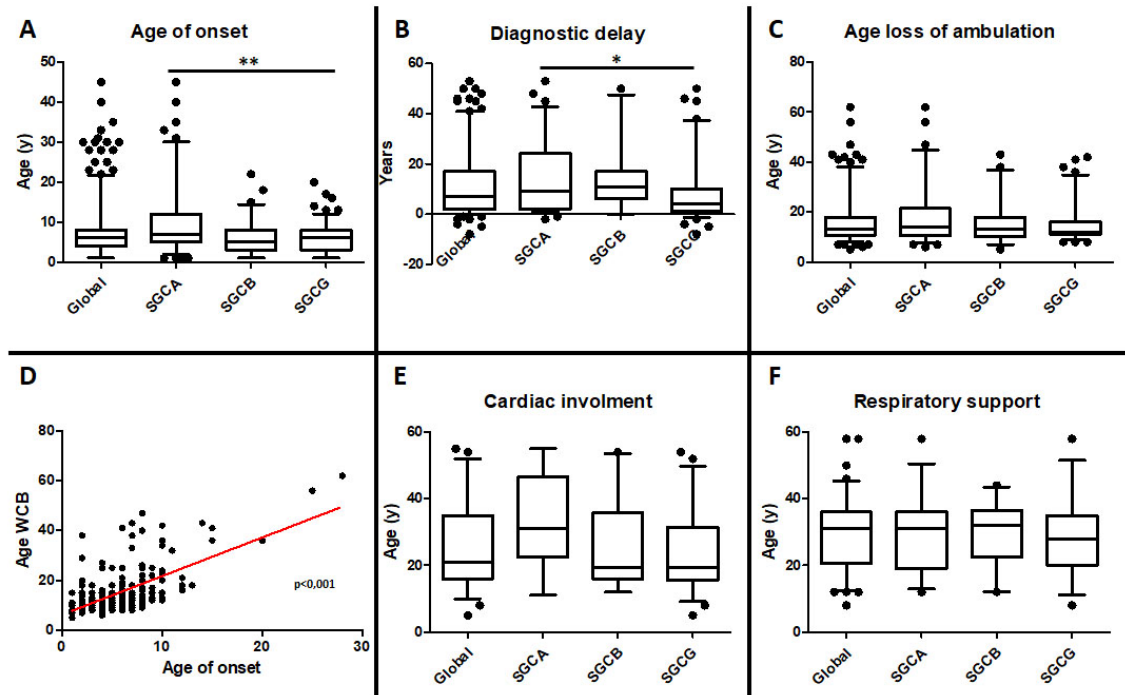


Figure 2 Group comparison

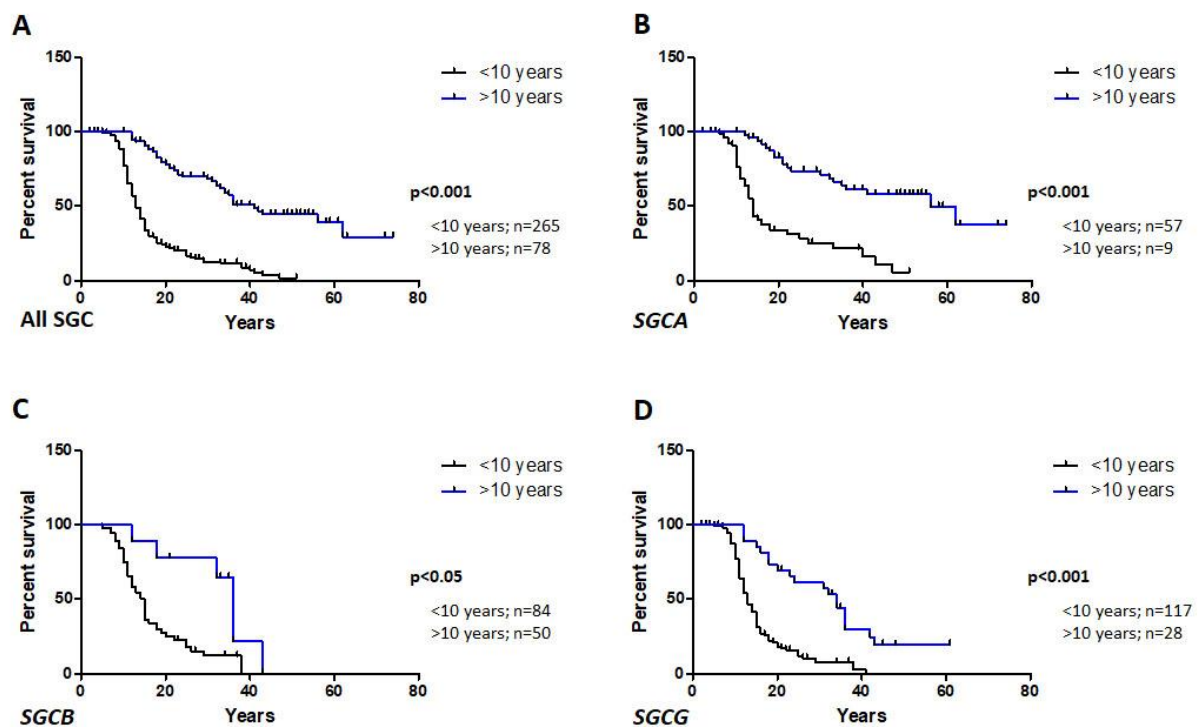


Figure 3 Age onset and WCB***Cardiac and respiratory involvement***

The frequency and type of cardiac involvement varied between subgroups as described in Table 1. Nineteen percent of all patients had cardiac involvement defined as a left ventricle ejection fraction lower than 50% or cardiac conduction defects. There were significant differences in the frequency of cardiac involvement (ANOVA test, $p=0.001$) between the subgroups, with beta patients showing more frequent cardiac involvement than alpha or gamma patients, but there was no difference in the age at which cardiac involvement was identified (ANOVA test, $p=0.34$) (Figure 2E). However, we identified a statistically significant correlation between the duration of the disease, measured as the time from the onset of symptoms, and the presence or cardiac involvement both in beta and gamma patients as described in supplemental table 2.

Up to twenty-five percent of all patients required ventilatory support, with the treatment recommended at a mean age of 29.1 years (range 8 – 58). Between the different subgroups, we neither identified differences in the frequency of patients requiring ventilatory support (ANOVA test, $p=0.64$) nor in the age at which it was recommended (ANOVA test, $p=0.93$) (Figure 2F). However, patients requiring ventilatory support had a significant longer duration of the disease (Table 2).

We did not identify differences in the age at onset, age at loss of ambulation or the remaining sarcoglycan expression between those patients with and without cardiomyopathy and those patients with and without ventilatory support (Supplemental table 2).

	SGCA	SGCB	SGCG
	Cardiac involment		
Age of onset	p=0.44	p=0.24	p=0.61
Age at last visit	p=0.85	p=0.005	p<0.001
Duration disease	p=0.73	p<0.001	p<0.001
Age WCB	p=0.68	p=0.45	p=0.61
Time to WCB	p=0.79	p=0.24	p=0.28
Protein expresión	p=0.34	p=0.47	p=0.17
	Respiratoty support		
Age of onset	p=0.13	p=0.28	p=0.61
Age at last visit	p=0.11	p<0.001	p<0.001
Duration disease	p<0.001	p<0.001	p<0.001
Age WCB	p=0.52	p=0,68	p=0.76
Time to WCB	p=0.47	p=0,91	p=0.47
Protein expresión	p=0.10	p=0.08	p=0.23

Supplemental Table 2 Cardiac and respiratory involvement

Genetic studies

A total of 131 different pathogenic variants were identified in the four sarcoglycan genes in patients included in this study (Figure 4). Interestingly, in every gene there were 2 or 3 mutations that were more prevalent than the others. Moreover, mutations were mainly concentrated in 2 or 3 exons of the genes.

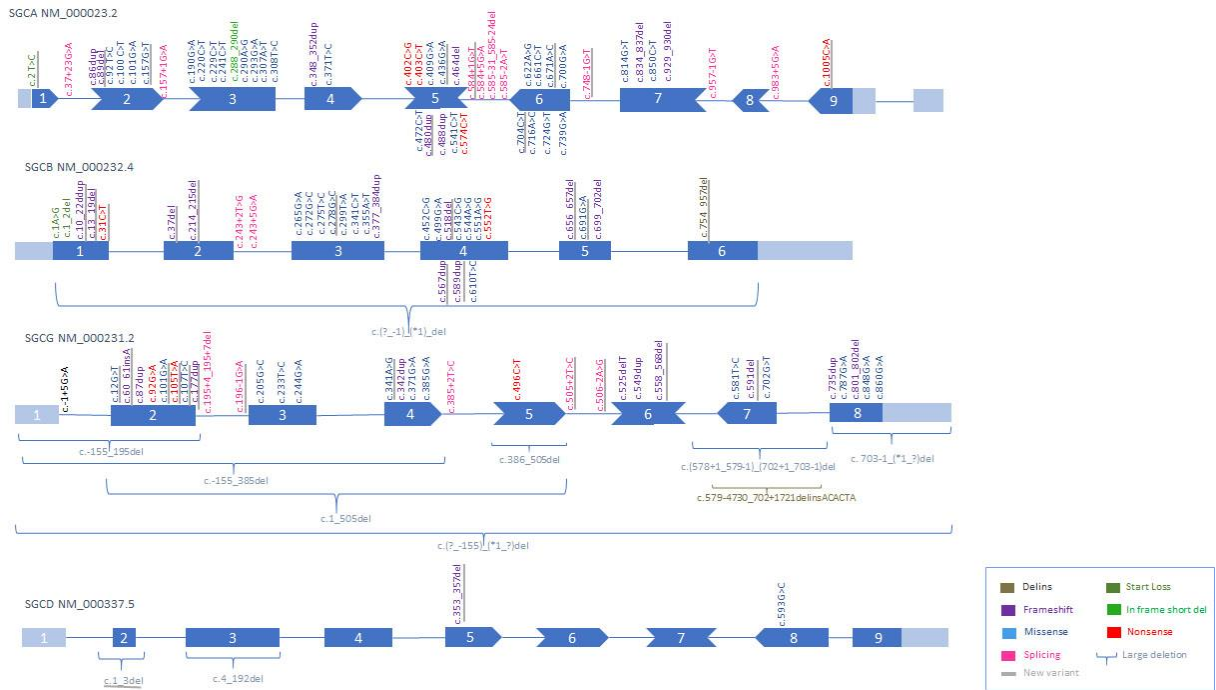


Figure 4 Genetic

Alpha patients:

Eighty-three alpha patients (52.2%) were homozygous for a single variant. In total, 52 different pathogenic variants were found in the *SGCA* gene and three of them were more prevalent than the others: c.229C>T (p.Arg77Cys) was present in at least one allele in 75 patients (47.2% of all alpha patients), c.850C>T (p.Arg284Cys) was present in 30 patients (18.9%) and, c.739G>A (p.Val247Met) in 26 patients (16.4%). Nine of the 53 variants had not been previously described (Figure 4). Most of the variants were located in the exons 3, 6 and 7, with 103 out of 159 patients having their two variants in these exons. Most of the patients had two missense mutations (74.8%), while only 3.1% carried two frameshift or nonsense mutations.

Beta patients:

Fifty-five beta patients (75.2%) were homozygous for a single variant. In total, 34 different pathogenic variants were found in the *SGCB* gene, but one of them, c.341C>T (p.Ser114Phe), was the most prevalent, being found in 30 patients (41.1% of all beta patients). Twelve of the 34 variants had not been previously described (Figure 4). Most of the variants were located in exons

3 and 4, with 55 out of 73 patients having their two variants in these exons (Figure 3). Fifty three percent of the patients had two missense mutations while 28.7% carried two frameshift or nonsense mutations.

Gamma patients:

One hundred and thirty-five gamma patients (86%) were homozygous for a single variant. In total, 41 different pathogenic variants were found in the *SGCG* gene and two of them were more prevalent than the others: c.525delT (p.Phe175Leufs*20) was present in at least one allele in 71 patients (45.2% of all gamma patients) and c.850C>T (p.Arg284Cys), which was present in 38 patients (24.2% of all gamma patients). Thirteen of the 41 variants had not been previously described (Figure 4). Most of the variants were located in exons 6 and 8, with 112 out of 157 patients having their two variants in these exons. Most of the patients had two frameshift or nonsense mutations (47.7%) while 33.8% carried two missense mutations. Interestingly, 43.9% of gamma patients carried the mutation c.525delT (known as a Maghrebien founder mutation), while 43.9% carried the c.848G>A mutation (known as a Gypsy founder mutation).

Delta patients:

The seven delta patients were homozygous for a single variant. In total, 4 different pathogenic variants were found in the *SGCD* gene. Five patients (71.4%) were homozygous for a deletion mutation. Two of the four variants had not been previously described (Figure 4).

Genotype-protein expression-phenotype correlation

To analyze the potential genotype-phenotype correlation, we decided to study the impact that the mutation has on the residual protein expression. Based on data obtained from western-blot of muscle biopsies, we were able to classify patients as having no protein expression, less than 30% protein expression and more than 30% protein expression.

We observed significant differences in the age of onset of the disease (ANOVA, $p < 0.001$), the age of loss of ambulation (ANOVA, $p < 0.001$) and the time from onset of symptoms to being wheelchair bound (ANOVA, $p < 0.001$) between patients with no expression or less than 30% expression and those patients with expression higher than 30% (Figure 5). When we performed the analysis by gene subgroups, we observed that age at being wheelchair bound in alpha patients was significantly influenced by protein expression (Mantel-Cox study, $p < 0.001$) (Figure 5B). In the case of beta patients, there was a non-significant trend in the age at being wheelchair bound depending on protein expression levels (Mantel-Cox, $p > 0.05$), although we observed that those expressing more than 30% of protein had a longer time from onset of symptoms to the loss of ambulation (Figure 5C). In gamma patients, significant differences were only found between patients having no expression or less than 30% and patients having more than 30% expression (Mantel-Cox, $p < 0.05$) (Figure 5D).

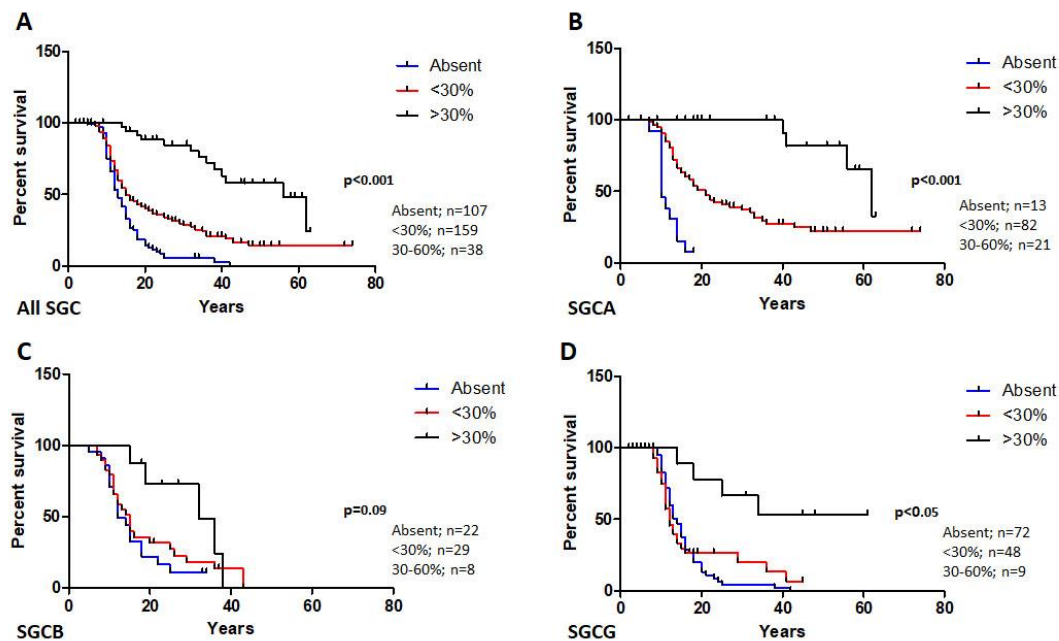


Figure 5 WCB and protein expression

Risk factors associated with rapid disease progression

To determine which risk factors could be associated with losing ambulation before the age of 18 years old, we performed a multivariable logistic regression analysis. First, we identified which factors could be associated using a univariate analysis (Sup. Table 2). The factors that showed statistical significance were age of onset, protein expression less than 30% and having mutations in the *SGCG* gene. These variables were further included in the multivariable logistic regression model that confirmed that age of onset (OR=1.37 (IC:1.18-1.59), $p=0.001$) and protein expression <30% (OR= 10.7 (IC:2.04-56.22), $p=0.020$) were independent risk factors for losing ambulation before 18 years old. The ROC curve for the final model, showed an AUC of 0.77 (Figure 6). In addition, we performed a ROC analysis to test which cut-point of age presented the most balanced sensitivity/specificity and the results are shown in Supplemental Figure 2. We found that the most balanced cut-point was age lower or equal to 6 years old, which showed a sensitivity of 69.33%, a specificity of 68.63% and an AUC of 0.71. The ROC curve for this cut-point of age is shown in Supplemental Figure 3.

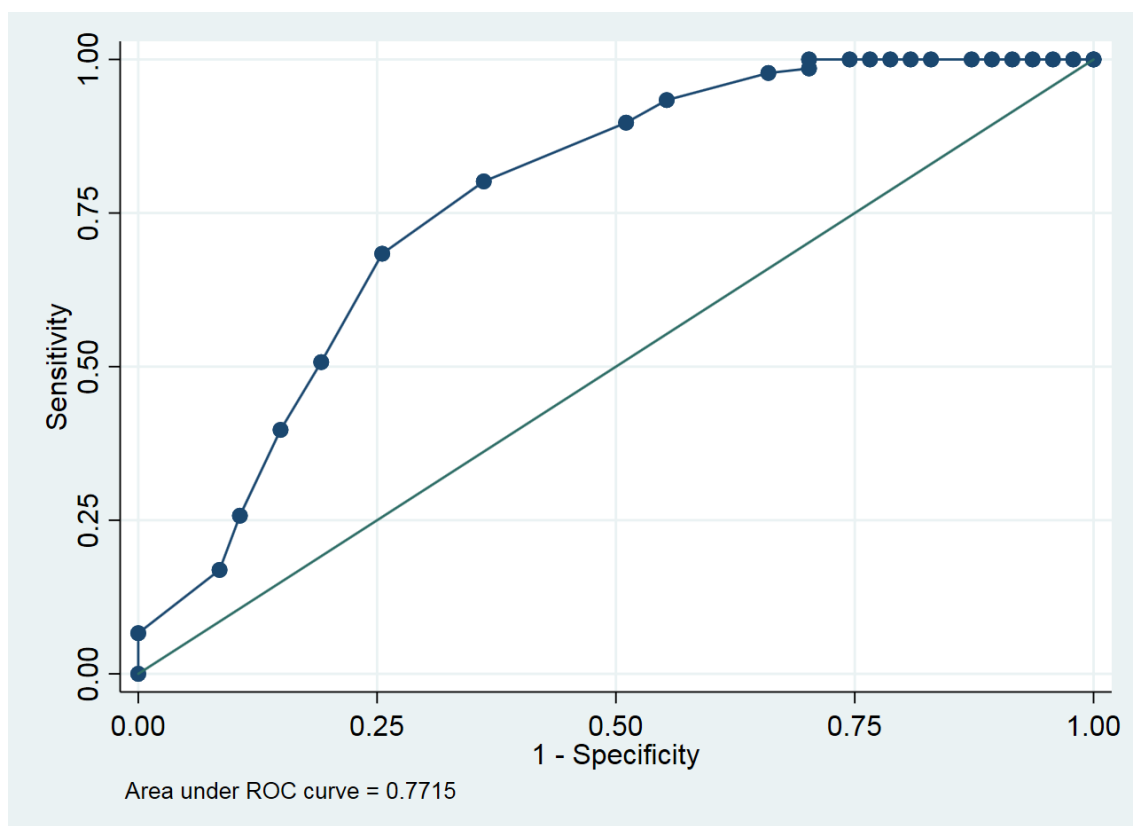


Figure 6 or Supplemental Figure 3 ??? ROC Age onset and protein expression

Discussion

We report the clinical, genetic and protein data of the largest series of patients with mutations in the sarcoglycan genes described so far. The analysis of this cohort revealed crucial data to establish a prognosis for the diseases as we observed that onset of symptoms during the first decade of life and protein expression in muscle biopsy lower than 30% were risk factors associated with loss of ambulation before 18 years of age. In contrast, onset of symptoms later than 10 years old and/or protein expression higher than 30% was related to a slower progression with a lower risk of losing ambulation before 18 years old. This data is especially relevant at present, as genetic therapies are under development and interventional clinical trials in patients with sarcoglycanopathies are being designed ^{27,28}.

Sarcoglycanopathies, as a group, are a common cause of LGMDR ^{3,29}. Mutations in these genes have classically been associated with a rapid progression of muscle weakness leading to severe disability, loss of ambulation in the second decade of life and frequent association with cardiac and respiratory problems ¹⁴. However, several reports suggested that disease progression was not homogeneous, as there were patients who never lose ambulation or patients in which the onset of symptoms was in the third or fourth decade of life. Our data showed that a series of factors may influence clinical progression including the mutated gene, age of onset of symptoms and residual protein expression. For example, progression of the disease was more rapid and severe for beta patients as a group than for alpha patients, as it had previously been suggested ¹⁴. The onset of symptoms for beta patients was during the first decade of life in 87% of the cases and 78.4% of them were wheelchair bound before 18 years old. However, we also identified beta patients in which clinical progression was milder, and in all these cases, remaining protein expression was higher than 30%. Previous series had suggested that gamma patients may have a milder progression than beta patients ^{13,25}. However, our data suggested that gamma patients have a more severe clinical phenotype than previously reported: 87.2% of them started with symptoms before 10 years of age and 84.4% of them were wheelchair bound before 18 years

old. In contrast, only 32% of alpha patients started the symptoms before 10 years of age and just 56.3% of them were wheelchair bound by the age of 18 years old. In addition, our study confirmed that remaining protein muscle expression was an independent risk factor for disease's progression as was previously suggested ^{14,25}. However, it was especially interesting that alpha patients, as a group, needed less remaining protein expression than beta and gamma to maintain ambulation beyond 18 years old. However, we think that these results should be interpreted carefully, as we have identified cases in which protein expression was higher than 30% and who lost ambulation before the age of 18 years old.

Analysis of mutations can be useful to predict protein expression in many disorders. Harboring two truncating mutations is generally associated with a loss of protein expression, however, the effect of having one or two missense mutations is not always easy to predict ². It is well known that after protein translation, sarcoglycans are glycosylated in the endoplasmic reticulum where some aberrant proteins could be identified by the ubiquitin-proteasome system and degraded. Consequently, some missense mutations can induce very low or no protein expression ^{30,31}. Therefore, in our opinion, caution is needed to predict protein expression based on genetic results only. In our study, we had information about protein expression studied by skeletal muscle WB in a considerable number of patients, which, based on the fact that many patients shared the same mutations, allowed us to predict protein expression in most of our cohort. Some of the most common missense mutations, such as c.229C>T (p.Arg77Cys) in the *SGCA* gene, c.341C>T (p.Ser114Phe) in the *SGCB* gene and c.848G>A (p.Cys283Tyr) in the *SGCG* gene were associated with a very reduced protein expression, always less than 30%, and consequently with a worse prognosis characterized by early onset of symptoms, more rapid progression and early loss of ambulation. In contrast, c.850C>T (p.Arg284Cys) and c.739G>A (p.Val247Met) in the *SGCA* gene, were associated with a higher protein expression and therefore a better prognosis. In our opinion, the data reported here can help in the design of clinical trials in sarcoglycanopathies. These trials should distinguish two groups of patients: those with 2

truncating mutations and/or none or less than 30% protein expression were commonly associated with an earlier onset of the disease, quicker progression and loss of ambulation before the age of 18 years old. In order to know the effectivity of gene therapy in these cases, children younger than 10 years should be included in the clinical trials. In contrast, patients with protein expression higher than 30% were often associated with milder progression, with most of them still ambulant in the fourth decade of life. If these patients were also included in clinical trials, analysis of the effectivity of the drug should be performed separately to the other group, as results could be biased by the milder progression of the disease.

Our results can also be interpreted in a different way. We have shown that protein expression higher than 30% was associated with a better prognosis. Therefore, gene replacement therapies able to reach almost this range of expression should be effective enough to show a change in the natural history of the disease. This data, although just hypothetical, is extremely important, as the efficacy of gene replacement therapies will depend on the amount of protein expression levels reached, which will also depend on the amount of viral genomes infused.

Another interesting finding of our analysis was the low frequency of cardiac and respiratory involvement. Previous cohort descriptions pointed out that up to 50% of patients developed cardiac and respiratory problems ^{2,22}. Our study had some limitations because our data were collected retrospectively and may not have been up to date for all patients and also because in the case of respiratory involvement, we were just collecting data regarding age at which ventilatory support was started. However, our results showed that only 21.6% of the patients had cardiac involvement and 26.2% required ventilatory support. We did not identify any correlation between protein expression, age of onset or age of losing ambulation and the development of cardiac and respiratory involvement. Therefore, it was not possible to predict the development of these symptoms based on protein expression data and repeated assessments of cardiac and respiratory involvement are needed in all cases.

This study provided also some interesting conclusions about the distribution of patients in different countries in Europe. We included patients from 13 different countries thanks to the collaboration of 33 different neuromuscular centers, which has been a huge effort of communication between the central data collecting center and every other contributing site. We confirmed that the two most prevalent subtypes were alpha- and gamma-sarcoglycanopathy, which represented 79.8% patients of our cohort, while beta patients were just 18.4% and delta patients were very rare with less than 1.8% of the cohort ^{32,33}. The distribution of patients between one country and the other was variable. Alpha patients were equally distributed in all countries. In contrast, most beta patients were from Northern European countries such as the Netherlands, the United Kingdom, Denmark and Germany, while most gamma patients were from Spain, Portugal and Italy. Forty percent of gamma patients harbor the Maghrebian founder mutation (c.525delT) and another 40% harbor the gipsy founder mutation (c.848G>A). These differences in the prevalence of subforms can be explained by historic migratory movements and are potentially susceptible to change in the next years.

The analysis of all mutations described in this cohort also revealed interesting data. We identified a total of 131 variants in the three genes, but most of these variants were concentrated in 2-3 exons per gene. For instance, 64.7% of alpha patients have their mutations in exons 3, 6 and 7, 75.4% of beta patients have their mutations in exons 3 and 4 while 73.7% of gamma patients have their mutations in exons 6 and 8, confirming that there are hot spots in these genes as has been previously described.

Our study has some obvious limitations. First, these data were retrospectively collected and there were missing data for many patients that could have enriched the data analyzed. We did not have data on protein expression in all cases, which would have allowed us to establish a clearer correlation between mutations identified and levels of remaining protein. Obviously, not all European patients were included in this analysis and although we contacted 62 centers to invite them to participate, not all of them answered and we missed data from some

countries such as Turkey, Russia, Romania and Bulgaria. However, we think that the data presented here is still relevant, because it describes the distribution of patients among many European countries, their clinical symptoms and the correlation between protein expression, age of onset and age of losing ambulation.

In conclusion, our study provides important information about the clinical and genetic features of a large cohort of limb girdle patients with mutations in the sarcoglycan genes. We have identified the age of onset of symptoms before 10 years old and remaining protein expression lower than 30% as independent risk factors for losing ambulation before 18 years old. This data will be useful for the design of clinical trials with gene replacement therapy that are currently under development.

References

1. Bushby K. Diagnosis and management of the limb girdle muscular dystrophies. *Pract Neurol*. 2009;9:314–323.
2. Winckler PB, da Silva AMS, Coimbra-Neto AR, et al. Clinicogenetic lessons from 370 patients with autosomal recessive limb-girdle muscular dystrophy. *Clin Genet*. 2019;96:341–353.
3. Vainzof M, Passos-Bueno MR, Pavanello RCM, Marie SK, Oliveira ASB, Zatz M. Sarcoglycanopathies are responsible for 68% of severe autosomal recessive limb-girdle muscular dystrophy in the Brazilian population. *J Neurol Sci*. 1999;164:44–49.
4. Straub V, Murphy A, Udd B. 229th ENMC international workshop: Limb girdle muscular dystrophies - Nomenclature and reformed classification Naarden, the Netherlands, 17-19 March 2017. *Neuromuscul Disord*. 2018;28:702–710.
5. Ervasti JM, Ohlendieck K, Kahl SD, Gaver MG, Campbell KP. Deficiency of a glycoprotein component of the dystrophin complex in dystrophic muscle. *Nature*. 1990;345:315–319.
6. Chan YM, Bönnemann CG, Lidov HGW, Kunkel LM. Molecular organization of sarcoglycan complex in mouse myotubes in culture. *J Cell Biol*. 1998;143:2033–2044.
7. Tarakci H, Berger J. The sarcoglycan complex in skeletal muscle. *Front Biosci - Landmark*. 2016;21:744–756.
8. Ozawa E, Mizuno Y, Hagiwara Y, Sasaoka T, Yoshida M. Molecular and cell biology of the sarcoglycan complex. *Muscle and Nerve*. 2005;32:563–576.
9. Lim LE, Campbell KP. The sarcoglycan complex in limb-girdle muscular dystrophy. *Curr Opin Neurol*. 1998;11:443–452.
10. Duclos F, Straub V, Moore SA, et al. Progressive muscular dystrophy in α -sarcoglycan-deficient mice. *J Cell Biol*. 1998;142:1461–1471.
11. Eymard B, Romero NB, Leturcq F, et al. Primary adhalinopathy (alpha-sarcoglycanopathy): clinical, pathologic, and genetic correlation in 20 patients with autosomal recessive muscular dystrophy. *Neurology*. 1997;48:1227–1234.
12. Xie Z, Hou Y, Yu M, et al. Clinical and genetic spectrum of sarcoglycanopathies in a large cohort of Chinese patients. *Orphanet J Rare Dis*. 2019;14:1–13.
13. Ferreira AFB, Carvalho MS, Resende MBD, Wakamatsu A, Reed UC, Marie SKN. Phenotypic and immunohistochemical characterization of sarcoglycanopathies. *Clinics*. 2011;66:1713–1719.
14. Semplicini C, Vissing J, Dahlqvist JR, et al. Clinical and genetic spectrum in limb-girdle muscular dystrophy type 2E. *Neurology*. 2015;84:1772–1781.
15. Angelini C, Fanin M, Freda MP, Duggan DJ, Siciliano G, Hoffman EP. The clinical spectrum of sarcoglycanopathies. 52. 1999;1.
16. Nalini A, Gayathri N, Thaha F, Das S, Shylashree S. Sarcoglycanopathy: Clinical and histochemical characteristics in 66 patients. *Neurol India*. 2010;58:691–696.
17. Tarnopolsky M, Hoffman E, Giri M, Shoffner J, Brady L. Alpha-sarcoglycanopathy

- presenting as exercise intolerance and rhabdomyolysis in two adults. *Neuromuscul Disord* [online serial]. Elsevier B.V.; 2015;25:952–954. Accessed at: <http://dx.doi.org/10.1016/j.nmd.2015.09.010>.
18. Kyriakides T, Angelini C, Vilchez J, Hilton-Jones D. European Federation of the Neurological Societies guidelines on the diagnostic approach to paucisymptomatic or asymptomatic hyperCKemia. *Muscle and Nerve*. 2020;61:E14–E15.
 19. Trabelsi M, Kavian N, Daoud F, et al. Revised spectrum of mutations in sarcoglycanopathies. *Eur J Hum Genet*. 2008;16:793–803.
 20. Gonzalez-Quereda L, Gallardo E, Töpf A, et al. A new mutation of the SCGA gene is the cause of a late onset mild phenotype limb girdle muscular dystrophy type 2D with axial involvement. *Neuromuscul Disord*. 2018;28:633–638.
 21. Oliveira Santos M, Coelho P, Roque R, Conceição I. Very late-onset limb-girdle muscular dystrophy type 2D: A milder form with a normal muscle biopsy. *J Clin Neurosci* [online serial]. Elsevier Ltd; Epub 2019.:10–12. Accessed at: <https://doi.org/10.1016/j.jocn.2019.12.003>.
 22. Politano L, Nigro V, Passamano L, et al. Evaluation of cardiac and respiratory involvement in sarcoglycanopathies. *Neuromuscul Disord*. 2001;11:178–185.
 23. Melacini P, Fanin M, Duggan DJ, et al. Heart involvement in muscular dystrophies due to sarcoglycan gene mutations. *Muscle Nerve* [online serial]. 1999;22:473–479. Accessed at: <http://doi.wiley.com/10.1002/%28SICI%291097-4598%28199904%2922%3A4%3C473%3A%3AAID-MUS8%3E3.0.CO%3B2-5>.
 24. Sveen ML, Thune JJ, Køber L, Vissing J. Cardiac involvement in patients with limb-girdle muscular dystrophy type 2 and becker muscular dystrophy. *Arch Neurol*. 2008;65:1196–1201.
 25. Guglieri M, Magri F, Angelo MGD, et al. Clinical , Molecular , and Protein Correlations in a Large Sample of Genetically Diagnosed Italian Limb Girdle Muscular Dystrophy Patients. *Hum Mutat*. 2008;29:258–266.
 26. Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J*. 2016;37:2129-2200m.
 27. Israeli D, Cosette J, Corre G, et al. An AAV-SGCG Dose-Response Study in a γ -Sarcoglycanopathy Mouse Model in the Context of Mechanical Stress. *Mol Ther - Methods Clin Dev* [online serial]. Elsevier Ltd.; 2019;13:494–502. Accessed at: <https://doi.org/10.1016/j.omtm.2019.04.007>.
 28. Vita G, Vita GL, Stancanelli C, Gentile L, Russo M, Mazzeo A. Genetic neuromuscular disorders: living the era of a therapeutic revolution. Part 1: peripheral neuropathies. *Neurol Sci. Neurological Sciences*; 2019;40:661–669.
 29. Fanin M, Nascimbeni AC, Aurino S, et al. Frequency of LGMD gene mutations in Italian patients with distinct clinical phenotypes. *Neurology*. 2009;72:1432–1435.
 30. Bianchini E, Fanin M, Mamchaoui K, Betto R, Sandonà D. Unveiling the degradative route of the V247M α -sarcoglycan mutant responsible for LGMD-2D. *Hum Mol Genet*. 2014;23:3746–3758.
 31. Sandonà D, Betto R. Sarcoglycanopathies: Molecular pathogenesis and therapeutic prospects. *Expert Rev Mol Med*. 2009;11:1–27.

32. Liu W, Pajusalu S, Lake NJ, et al. Estimating prevalence for limb-girdle muscular dystrophy based on public sequencing databases. *Genet Med* [online serial]. Springer US; 2019;21:2512–2520. Accessed at: <http://dx.doi.org/10.1038/s41436-019-0544-8>.
33. Ghaoui R, Cooper ST, Lek M, et al. Use of whole-exome sequencing for diagnosis of limb-girdle muscular dystrophy: Outcomes and lessons learned. *JAMA Neurol*. 2015;72:1424–1432.